



**НАУКА**

**и глобальные вызовы  
XXI века**

**ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ  
В РЕШЕНИИ АКТУАЛЬНЫХ СОЦИАЛЬНЫХ  
И ЭКОНОМИЧЕСКИХ ПРОБЛЕМ XXI ВЕКА**

Пермь, 2022

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ  
РОССИЙСКОЙ ФЕДЕРАЦИИ

Пермское отделение Научного совета при президиуме РАН  
по методологии искусственного интеллекта  
Российская ассоциация искусственного интеллекта  
Пермский государственный национальный исследовательский университет  
Пермский национальный исследовательский политехнический университет  
Национальный исследовательский университет «Высшая школа экономики»  
Пермский государственный гуманитарно-педагогический университет  
Пермский государственный медицинский университет им. академика Е.А.Вагнера  
Институт народнохозяйственного прогнозирования РАН  
Финансовый университет при Правительстве РФ

# ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ В РЕШЕНИИ АКТУАЛЬНЫХ СОЦИАЛЬНЫХ И ЭКОНОМИЧЕСКИХ ПРОБЛЕМ XXI ВЕКА

*Сборник статей по материалам Восьмой всероссийской  
научно-практической конференции с международным участием*

*(г. Пермь, 25–26 октября 2022 г.)*



Пермь 2022

УДК 004.8: 3  
ББК 32.813 + 6/8  
И868

**Искусственный интеллект** в решении актуальных социальных и экономических проблем XXI века : сборник статей по материалам Восьмой всероссийской научно-практической конференции с международным участием (г. Пермь, 25–26 октября 2022 г.) / под редакцией Л. Н. Ясницкого ; Пермский государственный национальный исследовательский университет. – Пермь, 2022. – 592 с.

ISBN 978-5-7944-3887-1

В сборнике представлены материалы Восьмой всероссийской научно-практической конференции с международным участием «Искусственный интеллект в решении актуальных социальных и экономических проблем XXI века», которая проводилась 25–26 октября 2022 г. в г. Перми в рамках Пермского естественнонаучного форума «Математика и глобальные вызовы XXI века».

Сборник предназначен для научных и педагогических работников, преподавателей, аспирантов, магистрантов, студентов и всех, кто интересуется проблемами и перспективами развития и применения методов искусственного интеллекта.

**УДК 004.8: 3**  
**ББК 32.813 + 6/8**

*Издается по решению кафедры прикладной математики и информатики  
Пермского государственного национального исследовательского университета*

*Сборник подготовлен при финансовой поддержке  
гранта в форме субсидии из бюджета Пермского края от 07.02.2022*

Статьи публикуются в авторской редакции, авторы несут ответственность за содержание статей, за достоверность приведенных в статье фактов, цитат, статистических и иных данных, имен, названий и прочих сведений

*Рецензенты:* президент Российской ассоциации искусственного интеллекта, профессор кафедры вычислительной техники Смоленского филиала НИУ «МЭИ», д-р техн. наук, **В. В. Борисов**;  
доцент кафедры информационных технологий в бизнесе НИУ ВШЭ – Пермь, канд. физ.-мат. наук **Е. Б. Замятина**

ISBN 978-5-7944-3887-1

© ПГНИУ, 2022

## ОГЛАВЛЕНИЕ

<b>Пленарные доклады .....</b>	<b>8</b>
<i>Ясницкий Л.Н., Кузнецов А.Г.</i>	
Как получить эффект от внедрения технологий искусственного интеллекта .....	9
<i>Сергеев С.Ф.</i>	
Искусственный и естественный интеллект в цифровых и биологических системах .....	11
<b>Искусственный интеллект в философии .....</b>	<b>22</b>
<i>Алексеев А.Ю.</i>	
Компьютерные средства комплексного теста тьюринга .....	23
<b>Искусственный интеллект в медицине и здравоохранении .....</b>	<b>29</b>
<i>Бутусов А.В., Сафронов Р.И., Филлист С.А.</i>	
Нейросетевой коррелятор для разделения аддитивной смеси кардиореспираторных сигналов двух и более пациентов.....	30
<i>Субботина Е.В.</i>	
Нейросетевая система прогнозирования уровня вакцинации от Covid-19.....	36
<i>Копнин М.В.</i>	
Нейросетевая система прогнозирования диабета.....	42
<i>Лавров К.М., Никитина А.Л.</i>	
Нейросетевая система прогнозирования обнаружения сахарного диабета на основе данных анамнеза.....	48
<i>Мокеева А.С.</i>	
Нейросетевая система прогнозирования злокачественной опухоли.....	55
<i>Кондрашов Д.С., Белозеров В.А., Горбачев И.Н., Аль-Дарраджи Часиб Хасан</i>	
Многомасштабные спектральные преобразования в классификаторах снимков видеоряда .....	62
<i>Крикунова Е.В., Кадырова С., Песок В.В., Сафронов Р.И.</i>	
Математические модели количественной оценки уровня защиты организма от внешних факторов .....	68
<i>Попов И.Е., Сайгин П.А.</i>	
Модель диагностики и обоснования коронавирусной инфекции по данным микроволновой радиотермометрии.....	74
<i>Завгородних Е.Д.</i>	
Нейросетевая система прогнозирования сердечного приступа .....	79
<i>Куликова Е.А.</i>	
Прогнозирование диабета методом нейросетевого моделирования .....	85
<i>Дементьев М.А.</i>	
Нейросетевая система прогнозирования продолжительности жизни человека.....	93
<i>Русаков В.В.</i>	
Нейросетевая система прогнозирования рака лёгких .....	100
<i>Соболева Д.А.</i>	
Нейросетевая система оценки риска обнаружения опухоли молочной железы .....	105
<i>Филатова П.Д., Гусев А.Л.</i>	
Нейросетевая система прогнозирования показателя общего коэффициента смертности в Пермском крае .....	111



<i>Стадниченко Н.С., Новоселов А.Ю., Пшеничный А.Е., Шаталова О.В.</i>	
Нейронные сети для классификации вирусных инфекций на модифицированных моделях Войта .....	119
<i>Коротышева А.А., Милов В.Р., Егоров Ю.С., Кербенева А.Ю.</i>	
Интеллектуальная система идентификации немаркированных элементов питания .....	125
<b>Искусственный интеллект в психологии и социологии .....</b>	<b>129</b>
<i>Болотная Т.Е.</i>	
Нейросетевая система прогнозирования уровня стресса после физической активности.....	130
<i>Красных Р.С.</i>	
Нейросетевая система прогнозирования склонности человека к депрессии .....	135
<i>Фаизова Л.Т.</i>	
Предсказание симптомов депрессии на основе социодемографических и поведенческих показателей с использованием нейронных сетей.....	142
<i>Шляев В.В.</i>	
Нейросетевая система определения зависимости от сериалов.....	149
<i>Трефилов Д.А.</i>	
Нейросетевая система прогнозирования склонности к суициду.....	157
<i>Зейтунян А.А.</i>	
Нейросетевое прогнозирование наличия нервно-психической неустойчивости .....	164
<i>Шарипова А.О.</i>	
Нейросетевая система оценки зависимости цветов фотографии профиля в социальной сети и количества отметок «нравится» .....	173
<i>Мельников П.А.</i>	
Нейросетевая система прогнозирования индекса счастья страны.....	180
<i>Густокашина В.М.</i>	
Нейросетевая система прогнозирования употребления алкоголя студентами.....	185
<i>Кузьминых У.И.</i>	
Нейросетевая система прогнозирования уровня успеха в IT-сфере .....	190
<i>Цыплякова Е.А.</i>	
Нейросетевая система прогнозирования ухода сотрудника из компании в течение двух лет .....	196
<b>Искусственный интеллект в экономике, производстве и бизнесе .....</b>	<b>205</b>
<i>Черемных А.А.</i>	
Нейросетевая система прогнозирования стоимости авиабилета .....	206
<i>Кононов Е.А., Клюев А.В.</i>	
Прогнозирование технического состояния авиационных двигателей при помощи нейронных сетей .....	214
<i>Гордеев М.В.</i>	
Нейросетевая система прогнозирования класса вязкости моторного автомобильного масла .....	218
<i>Бирюков А.Н.</i>	
Применение многофакторной нейросетевой модели для рационализации бюджетных расходов деятельности лечебных учреждений .....	224
<i>Осмоловский В.А.</i>	
Прогнозирование количества арендованных велосипедов.....	231

<i>Епишина Н.В., Семёнов С.П.</i>	
Нейросетевая система прогнозирования качества вина на основе данных о его химическом составе.....	238
<i>Тарасова В.Н.</i>	
Нейросетевая система прогнозирования качества красного вина .....	244
<i>Багин М.Н.</i>	
Нейросетевое прогнозирование стоимости квартир в городе Пермь.....	252
<i>Макшаков А.А.</i>	
Нейросетевая система оценки стоимости квадратного метра жилой недвижимости города Перми .....	257
<i>Тимофеев А.В., Ушакова О.А.</i>	
Нейросетевая система прогнозирования стоимости домов.....	264
<i>Толочко А.</i>	
Нейросетевая система прогнозирования цены жилья за квадратный метр .....	269
<i>Каменских Л.А.</i>	
Нейросетевая система прогнозирования средней заработной платы.....	273
<i>Неверова Е.А.</i>	
Нейросетевая система прогнозирования рейтинга кинофильма в прокате .....	279
<i>Кондратенков В.Ю., Плахина Т.С., Орлов А.А.</i>	
Нейросетевая система прогнозирования оценки компьютерных игр .....	285
<i>Овчинникова А.А.</i>	
Нейросетевая система автоматизации управления сушильной камерой для просушивания семян шишек .....	295
<i>Шориков А.Ф.</i>	
Интеллектуальный программный комплекс моделирования прогнозирования и минимаксного оценивания фазовых состояний дискретной управляемой динамической системы .....	300
<i>Брагин К.А.</i>	
Нейросетевая система прогнозирования одобрения кредита .....	306
<i>Южаков А.А., Сторожев С.А.</i>	
Адаптивное групповое управление переопределенным объектом на примере камеры сгорания газотурбинного двигателя .....	313
<i>Соколов Д.А., Нацанов М.А., Загвозкин В.Д., Хижняков Ю.Н.</i>	
Выбор и оптимизация параметров регулятора элемента Пельтье блока БУЦ, разработанного на отечественной элементной базе.....	320
<i>Агаев А.Р.</i>	
Нейросетевая система определения оптимальной агрокультуры для возделывания на участке.....	327
<i>Степанов В.А., Култышева С.Н.</i>	
Результаты разработки прототипа прогностической системы по моделированию распределения температуры по глубине ствола скважины в процессе пароциклических обработок .....	333
<i>Ковалева С.С.</i>	
Нейросетевая система прогнозирования стоимости домов.....	340
<i>Сырвачева Е.Р.</i>	
Нейросетевая система прогнозирования уровня стресса человека во сне.....	347

<i>Аухадиев М.Р.</i>	
Нейросетевая система прогнозирования заполненности зала на спектакле в Театре-Театре.....	355
<i>Силина А.А., Постных Д.В., Мезенцев А.С., Черепанов Ф.М., Морозов А.А., Голдобин М.А., Ясницкий В.Л., Ясницкий Л.Н.</i>	
Нейронные сети и управление качеством производственных процессов .....	362
<i>Поселенцева Д.Ю.</i>	
Нейросетевая система прогнозирования уровня удовлетворенности посетителей гостиниц .....	365
<i>Шилов И.А.</i>	
Нейросетевая система прогнозирования индекса производительности графических процессоров .....	373
<b>Искусственный интеллект в педагогике.....</b>	<b>381</b>
<i>Кобылно М.Е.</i>	
Нейросетевая система прогнозирования вероятности поступления в магистратуру.....	382
<i>Рычков А.В.</i>	
Нейросетевая система прогнозирования поступления абитуриента в вуз «Уральский федеральный университет» .....	389
<i>Шилова Е.А.</i>	
Нейросетевая система прогнозирования лучшего университета .....	394
<b>Искусственный интеллект в демографии и экологии.....</b>	<b>403</b>
<i>Данелян В.С., Клюев А.В.</i>	
Эффективный мониторинг несанкционированных мест размещения отходов по спутниковым снимкам с использованием методов искусственного интеллекта .....	404
<i>Шавкунов П.А.</i>	
Нейросетевая система прогнозирования ожидаемой продолжительности жизни людей для стран .....	410
<b>Искусственный интеллект в лингвистике .....</b>	<b>416</b>
<i>Гайнетдинова В.А., Ланин В.В., Стринюк С.А.</i>	
Разработка системы для анализа синтаксиса научных текстов на английском языке.....	417
<b>Искусственный интеллект в кинематографии .....</b>	<b>423</b>
<i>Чепокоев Е.С.</i>	
Нейросетевая система прогнозирования рентабельности кинобизнеса .....	424
<i>Веверица К.Е.</i>	
Нейросетевая система прогнозирования получения Оскара .....	429
<b>Искусственный интеллект в астрономии, метрологии, катастрофах.....</b>	<b>438</b>
<i>Соскин А.И.</i>	
Нейросетевая система прогнозирования необходимости тушения лесных пожаров .....	439
<i>Карибова А.С.</i>	
Нейросетевая система прогнозирования показателей солнечной освещенности .....	444
<i>Жуйкова С.К.</i>	
Нейросетевая система прогнозирования типа звезды .....	451
<i>Калинина М.О.</i>	
Нейросетевая система прогнозирования выживания человека в катастрофе на примере титаника .....	463

<b>Искусственный интеллект в проблемах безопасности .....</b>	<b>469</b>
<i>Понькин Н.А., Гильмутдинов Р.Р.</i>	
Нейросетевая система детектирования медицинских масок на лицах людей.....	470
<i>Гладкий С.Л., Жуланов В.Н.</i>	
Обучение архитектуры YOLOv3 для детектирования оружия на видео с камер видеонаблюдения.....	477
<i>Липин Ю.Н., Сторожев С.А.</i>	
Разработка программы моделирования алгоритмов систем распознавания лиц .....	482
<i>Аникина И.В., Бурылова А.А.</i>	
Нейросетевая система определения склонности человека к серийным убийствам.....	489
<b>Искусственный интеллект в спорте .....</b>	<b>496</b>
<i>Батаев Б.В.</i>	
Нейросетевая система оценки эффективности баскетбольного игрока.....	497
<i>Соболев Д.А.</i>	
Нейросетевая система прогнозирования получения награды самого ценного игрока сезона.....	504
<i>Белоусов А.А.</i>	
Нейросетевая система прогнозирования победителя баскетбольного матча на основе результативности игрока .....	512
<i>Коробов Н.А.</i>	
Нейросетевая модель прогнозирования занятого киберспортсменом места в голосовании за лучшего игрока года.....	516
<i>Стрекаловская Е.В.</i>	
Нейросетевая система прогнозирования победителя на скачках .....	523
<i>Верхоланцев Ф.А.</i>	
Прогнозирование исхода боев смешанных единоборств методом нейросетевого моделирования .....	531
<i>Мицкевич А.Д.</i>	
Нейросетевая система прогнозирования результатов в женском одиночном фигурном катании на Олимпиаде-2026 .....	539
<i>Павлов Д.А.</i>	
Нейросетевая система прогнозирования шансов победы сборных на Чемпионате мира по футболу .....	549
<b>Инструменты искусственного интеллекта и математика.....</b>	<b>558</b>
<i>Золотарева Т.А., Иванов А.И.</i>	
Прогноз необходимого числа бинарных нейронов при проверке гипотезы независимости данных .....	559
<i>Рабчевский А.Н.</i>	
Роль синтетических данных в развитии искусственного интеллекта .....	563
<i>Воробьев Л.О.</i>	
Использование DBLATEX в сочетании с ASCIIDOCTOR для оформления научно-исследовательских работ студентов программистских специальностей в соответствии с действующими стандартами .....	572
<i>Калиниченко Б.А.</i>	
Термодинамические следствия уравнения Ван дер Ваальса в области критического состояния.....	580
<i>Хижняков Ю.Н., Южаков А.А., Никулин В.С., Сторожев С.А.</i>	
Методы адаптации нечетких систем управления .....	586

«Думаю, она [Конференция] хорошо отображает тот факт, что ваша Пермская школа и ваша системная деятельность – лучшее, что есть в нашей стране в области многоплановой разработки проблем ИИ.»

*Д.И.Дубровский, Сопредседатель Научного совета при президиуме РАН по методологии искусственного интеллекта, д.ф.н., профессор.*

## **ПЛЕНАРНЫЕ ДОКЛАДЫ**

## КАК ПОЛУЧИТЬ ЭФФЕКТ ОТ ВНЕДРЕНИЯ ТЕХНОЛОГИЙ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

*Ясницкий Леонид Нахимович*

*Кузнецов Андрей Геннадьевич*

Пермский государственный национальный исследовательский университет,  
614990, Россия, г. Пермь, ул. Букирева, 15. yasn@psu.ru

Цифровая трансформация производства, а в более широком смысле, промышленности и экономики, является актуальнейшим направлением развития научно-технического прогресса, основой нового технологического уклада. В числе наиболее перспективных технологий современного информационного общества – технологии искусственного интеллекта. Развитие технологий искусственного интеллекта (ИИ) обеспечило возможность создания систем автоматизации принципиально нового уровня, обеспечивающих высокий уровень эффективности производства и повышение качества продукции.

Используемые в настоящий момент автоматизированные системы управления опираются на логику и точные расчеты, применяя упрощенную «модель производственных процессов», выбранную заранее. ИИ позволяет анализировать обстановку в реальном времени и сохранять работоспособность при смене целей управления, непредвиденных изменениях свойств управляемого объекта или параметров окружающей среды. Такая система способна менять алгоритм управления и искать оптимальные и эффективные решения. Можно забежать вперед и обозначить ИИ как область знания, занимающуюся автоматизацией разумного поведения технических систем.

Вместе с тем, как и любые технологии, технологии ИИ, наряду с очевидными достоинствами, объективно имеют недостатки, которые необходимо учитывать, принимая решение о применении решений, основанных на технологиях ИИ. И тем более необходимо минимизировать субъективные ошибки в организации процесса разработки и внедрения систем, основанных на технологиях ИИ.

Ученые механико-математического факультета Пермского классического университета (ПГНИУ) имеют многолетний положительный опыт развития и применения методов искусственного интеллекта (ИИ) в самых различных предметных областях [1, 2], включая промышленность и экономику. Обобщая этот опыт, мы сформулировали несколько правил, которые могут помочь избежать типичных ошибок, связанных с реализацией и внедрением технологий ИИ, и получить ожидаемый эффект.

1. Не следует ждать чуда от ИИ. Методы машинного обучения действительно позволяют многое. Они способны извлекать знания из статистической информации и на их основе делать прогнозы, помогать принимать

правильные решения. Но, если в предоставленных для обучения датасетах знаний нет, то методы машинного обучения бессильны.

2. Предприятия должны иметь возможность предоставить разработчикам ИИ достаточно полные датасеты, желательно, с минимумом ошибочной информации. На предприятиях должны быть определены сотрудники, заинтересованные в результатах проекта, способные собирать и предоставлять данные, активно участвовать в формулировке задачи и ходе выполнения проекта, иметь достаточную квалификацию для интерпретации промежуточных и финальных результатов.

3. Необходимо привлечение специалистов предприятия для подготовки необходимых датасетов с квалификацией, достаточной не только для формального переформатирования исходных данных, но и выявления недостоверной информации.

4. Успех машинного обучения напрямую зависит от объема обучающей выборки. Поэтому для применения методов машинного обучения хорошо подходят серийные предприятия, выпускающие существенные партии изделий, например, когда требуется снизить процент брака, улучшить качество продукции и др.

5. Методы машинного обучения позволяют создавать модели, учитывающие неограниченное количество факторов, влияющих на результат. Этим свойством полезно пользоваться, включая в датасеты сведения о как можно большем количестве параметров. Опыт показывает, что иногда факторы, казалось бы, не имеющие отношения к моделируемому явлению, оказываются важными для получения положительного результата.

6. Нейронные сети унаследовали от своего прототипа мозга множество положительных качеств. Но наряду с положительными, унаследовали и отрицательные. Например, способность обманывать человека, в частности, разработчика ИИ. Успешно пройдя все этапы тестирования, уже после передачи систем искусственного интеллекта заказчику, они могут выдать такой прогноз, который введет в ступор не только заказчика, но и разработчика. Поэтому руководителям предприятий не следует выбирать себе в исполнители недостаточно опытных разработчиков ИИ.

В последнее время, в связи необычайно возросшей популярностью ИИ, появилось много так называемых «продавцов ИИ». Освоив зарубежные программные инструменты на пользовательском уровне, очарованные рекламой, поверив в безграничные возможности ИИ, они смело берутся за выполнение грандиозных проектов. А руководители предприятий под воздействием все той же рекламы не скупятся на их финансирование. Участились случаи, когда такие проекты заканчиваются ничем, а иногда и требованиями вернуть аванс. Во избежание подобных ситуаций бывает полезно привлекать квалифицированных экспертов, в коем качестве, берем на себя смелость, можем предложить услуги ученых и специалистов нашего университета.

В настоящее время в Пермском крае сложилась уникальная ситуация, когда одновременно несколько коллективов ученых и разработчиков занимаются фундаментальными и прикладными исследованиями в области ис-

кусственного интеллекта применительно к различным областям науки и производства, Пермь заслуженно является одним из ведущих центров по разработке и внедрению систем ИИ. Примеры на слуху: это и всемирно раскрученный ПРОМОБОТ, и успешная реализация систем видеонаблюдения от компании Макроскоп, и многие другие.

### **Библиографический список**

1. Ясницкий Л.Н. Развитие научной школы искусственного интеллекта в пермских университетах: история и научный приоритет (обзорная статья) // Прикладная математика и вопросы управления. 2018. № 4. С. 99-130. <https://www.elibrary.ru/item.asp?id=36727550>

2. Ясницкий Л.Н. О научном приоритете пермских ученых в области искусственного интеллекта // В сборнике: Искусственный интеллект в решении актуальных социальных и экономических проблем XXI века. Часть I. Сборник статей по материалам Четвертой всероссийской научно-практической конференции, проводимой в рамках Пермского естественнонаучного форума «Математика и глобальные вызовы XXI века». 2019. С. 7-25. <https://cloud.mail.ru/public/qzQE/3CTntPDdo>

## **HOW TO GET THE EFFECT FROM THE IMPLEMENTATION OF ARTIFICIAL INTELLIGENCE TECHNOLOGIES**

*Yasnitsky Leonid N., Kuznetsov Andrey G.*

Perm State University,  
614990, Russia, Perm, st. Bukireva, 15, yasn@psu.ru

УДК 004.8

## **ИСКУССТВЕННЫЙ И ЕСТЕСТВЕННЫЙ ИНТЕЛЛЕКТ В ЦИФРОВЫХ И БИОЛОГИЧЕСКИХ СИСТЕМАХ**

*Сергеев Сергей Федорович*

Санкт-Петербургский государственный университет,  
199034, Россия, г. Санкт-Петербург, Университетская наб., 7-9,  
ssfpost@mail.ru

В статье дан краткий анализ истории появления и развития проблематики искусственного интеллекта и связанных с ней актуальных направлений в области инженерной психологии. При создании систем и техноло-



гий искусственного интеллекта, порождаемых эволюцией техногенной среды человеческой цивилизации необходимо учитывать вопросы обеспечения взаимодействия человека с организованной эволюционирующей средой. Рассматриваются перспективные направления психологических и инженерных исследований и возможные проблемы при использовании в этих областях концептуально-понятийного аппарата постнеклассической психологии, методологии самоорганизующихся сложных систем.

**Ключевые слова:** интеллект, искусственный интеллект, постнеклассическая психология, самоорганизация, организованная сложность.

**Введение.** Популярное в современном научно-технологическом дискурсе понятие «интеллект» имеет психологическое происхождение и как психологическая категория впервые появилось в работах Френсиса Гальтона, при обосновании определяющей роли наследственности в различиях в обучаемости между людьми.

В психологическом смысле интеллект определяет способности человеческого мышления к рациональному познанию мира, логическому мышлению, отражает ум во всех его проявлениях и формах. В силу столь широкого толкования это понятие допускает множественные варианты интерпретаций и определений что мешает общему пониманию в различных научных и инженерных дисциплинах.

Можно сказать, что «интеллект» и его технический аналог – «искусственный интеллект» – это самые популярные психологические и технические понятия, используемые в практике создания сложных объектов современной технологической цивилизации.

В рамках возникающего техно-гуманитарного дискурса речь идет об «умных алгоритмах», «умном городе», «разумной пыли», «программируемой материи», «роевых системах», сенсорных сетях, киберфизических системах, самообучающихся на базе своего опыта технических устройствах и комплексах. В философском плане искусственный интеллект – это разум кибернетической технической системы.

Понятие «интеллект» давно вышло за границы психологии, приобрело трансдисциплинарный характер. Отметим, что в западной литературе под интеллектом иногда понимается более широкая категория, ассоциирующаяся в отечественном научном дискурсе с понятием «разум». Особое положение категории интеллект в психологии связано с ожиданиями со стороны техники и технологии новых идей и понятий. Конвергентный характер категории интеллект особо ярко проявляется в понятии «искусственный интеллект», которое объединяет в рамках единого дискурса вопросы эволюции человека, социума и техногенной среды.

### **Интеллект в психологии**

В классической психологии основные результаты по исследованию проблемы способностей, интеллекта и их измерению получены в исследованиях Г.Ю. Айзенка, Р. Амтхауэра, Б.Г. Ананьева, А. Бине, Н. Бострома,

Д. Векслера, Ч. Гарднера, Д. Гилфорда, Т. Келли, Р. Кеттелла, Ж. Пиаже, К.К. Платонова, Дж. Равена, С.Л. Рубинштейна, Т. Симона, Ч. Спирмена, Р. Дж. Стернберга, Б.М. Теплова, Л. Терстоуна, Дж. Томпсона, Р.Л. Торндайка, М.А. Холодной, В. Штерна.

Определений понятия «интеллект» в психологии очень много. Самое простое и формально точное, определение Э. Боринга, интеллект – это то, что измеряется тестами интеллекта.

По мнению Г.Ю. Айзенка существует не менее трех радикально различающихся концепций интеллекта: биологический, психометрический и социальный интеллекты, каждый из которых не может быть определен без двух других [1].

В рамках концепции множественного интеллекта Ховард Гарднер определяет интеллект как «способность к решению задач или созданию продуктов, обусловленную конкретными культурными особенностями или социальной средой» [2, с. 15]. Им выделено семь базовых интеллектуальных способностей (интеллектов):

– *Вербальный интеллект* – способность к порождению речи, включающая механизмы, ответственные за фонетическую (звуки речи), синтаксическую (грамматику), семантическую (смысл) и прагматическую составляющие речи (использование речи в различных ситуациях).

– *Музыкальный интеллект* – способность человека различать смысл и значение в определенной последовательности ритмичных звуков, а также с коммуникативной целью производить такие последовательности звуков, включая механизмы, ответственные за восприятие высоты, ритма и тембра (качественных характеристик) звука.

– *Логико-математический интеллект* – способность использовать и оценивать соотношения между действиями или объектами, когда они фактически не присутствуют, т. е. к абстрактному мышлению.

– *Пространственный интеллект* – умение точно воспринимать зримый мир, выполнять трансформации и модификации его согласно первому впечатлению, а также умение воссоздавать аспекты визуального опыта даже при отсутствии соответствующего физического объекта.

– *Телесно-кинестетический интеллект* – способность использовать все части тела при решении задач или создании продуктов; включает контроль над грубыми и тонкими моторными движениями и способность к манипулированию внешними объектами, обуславливает работу тела и лежит в основе всех действий человека в мире.

– *Внутриличностный интеллект* – способность к осознанию и пониманию человеком своих чувств, намерений и мотивов.

– *Межличностный интеллект* – умение замечать и понимать различия между окружающими, способность к распознаванию и проведению различия между чувствами, взглядами и намерениями других людей [2]. Впоследствии Гарднер пришел к выводу о наличии в дополнение к вышеуказанным *экзистенциального и натуралистического* интеллектов, обеспечивающих существование человека, его способность к использованию интуиции, мышле-

ния и метапознания, распознавание и оценку взаимоотношений человека и окружающего мира.

Кубическая многофакторная модель интеллекта Джоя Пола Гилфорда, отражая дивергентный подход к мышлению, постулирует существование независимых факторов в виде интеллектуальных способностей, объединенных в единую трехмерную систему, названную автором структурой интеллекта насчитывающая 120 факторов интеллекта [3].

По мнению автора понятия «практический интеллект» Р. Стернберга интеллект – это способность гибко и эффективно адаптироваться к окружающей среде [4].

Неопределенность термина «интеллект» в психологии позволила М.А. Холодной задать вопрос «существует ли интеллект как психическая реальность?» [5]. В соответствии с предложенной Холодной теорией *интеллект по своему онтологическому статусу* – это особая форма организации индивидуального ментального (умственного) опыта в виде наличных ментальных структур, прогнозируемого ими ментального пространства и строящихся в рамках этого пространства ментальных репрезентаций происходящего. Это означает, что при исследовании интеллектуальных возможностей человека необходимо учитывать внутренние индивидуальные характеристики интеллекта. Действительной феноменологией интеллекта, с точки зрения М.А. Холодной, являются не его свойства, с высокой степенью разнообразия и вариативности обнаруживающие себя в ситуациях решения задач, а особенности структурной организации иерархии психических носителей, которые «изнутри» определяют эмпирически констатируемые проявления интеллектуальной деятельности.

По своему онтологическому статусу интеллект – это форма организации когнитивного, метакогнитивного и интенционального опыта человека. Степень сформированности этих трех форм ментального опыта определяет структурные характеристики *ментального* пространства интеллектуального отражения. Основное назначение интеллекта по М.А. Холодной – построение особого рода *ментальных репрезентаций происходящего*, связанных с воспроизводством *объектного* знания о мире [6].

Вышеизложенное понимание интеллекта, соответствует естественной форме человеческого познания, осуществляемого в конструктах субъективной реальности человека в отличие от искусственного интеллекта, непосредственно алгоритмически преобразующего информацию, поступающую от физических систем и модулей памяти. Несмотря на расцвет инженерного направления «искусственный интеллект» наблюдаемого в настоящее время, в самой психологии увлечение интеллектом прошло достаточно рано на этапе увлечения психологическими тестами на интеллект и заменено впоследствии на *исследования* познавательных стилей [7] и *частных интеллектуальных* способностей [8].

Наблюдаемая экспансия данного понятия в инженерию связана с тем, что оно в известной мере заменяет термин «эффективность». Эффективная система в технике это «умная», наделенная интеллектом система, в силу сво-

его искусственного происхождения обладающая «искусственным интеллектом». Второй причиной являться маркетинговый ход, позволяющий авторам получать ресурсы на развитие междисциплинарных исследований сложных технических систем. В пользу второго аргумента говорит то, что первые исследователи искусственного интеллекта не имели ничего общего с психологией, а являлись системными аналитиками и высококвалифицированными инженерами. Один из авторов термина «искусственный интеллект» Джон Маккарти являлся программистом, основоположником функционального программирования.

### ***Системные аспекты проблемы искусственного интеллекта***

Научно-философское обоснование проблемы искусственного интеллекта, как программно-информационного моделирования интеллектуальных функций человека в технических системах связано с работами Д.И. Дубровского автора информационной концепции психического и сознания [9]. В соответствии с нею:

1. Информация необходимо воплощена в своем физическом носителе (не существует вне определенного физического объекта, процесса).

2. Одна и та же информация (для данного типа самоорганизующихся систем) может быть воплощена (и передана) разными по своим физическим свойствам носителями. Это принцип инвариантности информации по отношению к физическим (химическим, субстратным, пространственным, временным) свойствам ее носителя.

3. Информация может служить фактором управления, причиной определенных изменений в самоорганизующейся системе.

Из принципа инвариантности информации по отношению к субстратным свойствам ее носителя следует изофункционализм систем, возможность воспроизведения одних и тех же функций на разных (по своим физическим, химическим, структурным свойствам) субстратах, что делает понятие интеллекта независимым от его физического носителя.

Классическая психология, создав метафору интеллекта как набора эффективных интеллектуальных инструментов, позволяющих решать сложные задачи, послужила основой для появления современных инженерных решений систем автоматического управления информационно-справочных и экспертных систем, средств автоматизации. Однако дальнейшее развитие технологий сложного мира вступает в противоречие с формируемой в классической психологии простой материалистической картиной мира. Требуется создание психологии сложного мира отражающей развитие человека в условиях интенсивного развития техносферы. Эта цель достигается в неклассической и постнеклассической психологии, рассматривающих субъективную реальность человека и личность с позиций включения в категориальный состав психологии принципов универсализма, относительности, случайности, системности и историзма, что позволяет решать проблемы создания сложных эргатических и робототехнических систем [10, 11].

Постнеклассическая психология использует научную методологию постнеклассической рациональности предложенную В.С. Степиным [12], в соответствии с которой постулируется:

- Целостный, глобальный взгляд на мир. Междисциплинарные и проблемно ориентированные формы исследовательской деятельности.
- Сближение физического и биологического мышления.
- Объект исследований – системы, характеризующиеся открытостью и саморазвитием: исторически развивающиеся и саморегулирующиеся.
- Гуманитаризация естественнонаучного знания, радикальное «очеловечивание» психологии. Человек в картине мира не просто активный участник, а как системообразующий принцип. Мышление человека с его целями, ценностными ориентациями сливается с предметным содержанием объекта.

В качестве базовой теории постнеклассической психологии выступает теория самоорганизации, открытых, замкнутых и операционально-замкнутых и закрытых систем. Рассматриваются нелинейность, необратимость, неравновесность, хаос [13, 14]. В объем понятия «рациональность» включены интуиция, неопределенность, эвристика и ряд других прагматических характеристик. Например, польза, удобство, эффективность.

В постнеклассической перспективе к информационному подходу добавляются представления, связанные с категорией «сложность».

Понятие «сложность» рассматривается в двух аспектах. Первый связан с субъективной сложностью, порождаемой в психике человека и обусловленной ограниченными возможностями человека по восприятию мира и обработке информации. Второй – с реальной сложностью физического и социального миров и возникающих в них феноменов. Представления о сложном отражаются в концептах: множественности; динамического разнообразия; нелинейности; неравномерности; сложности самоорганизующихся систем [15].

Множественность рассматривается как *многокомпонентность*. Она относится к описанию сложных систем непрерывно эволюционирующих и изменяющихся. Сложная система в этой парадигме предстает как процесс бесчисленного усложнения ее сущностей, возникновения новых элементов и уровней организации для новых воплощений.

Множественность в концепциях динамического разнообразия дополняется качественной характеристикой – *разнообразием*. Разнообразие связано с асинхронным существованием в среде динамически существующих и сосуществующих систем и их распадающихся элементов, которые могут образовывать в свою очередь новые системные сущности.

Динамические процессы, связанные с разнообразием, протекают как скачкообразные *нелинейные* процессы. Сложность данных систем отражает непредсказуемость появления новых качеств в новых структурах и неопределенность направления их развития. Причина появления нового качества может быть чрезвычайно малой и на первый взгляд незначительной («эффект бабочки»).

Концепт *неравномерности* отражает принципиальную неравномерность распределения в пространстве одновременно существующих различ-

ных форм материи (энергии, вещества), ведущую к локальной самоорганизации и возникновению новых систем. Постулируется холистический характер мира, который разделяется на элементы только работой механизмов человеческого сознания.

Концепция *сложности самоорганизующихся систем* отражает непрерывную динамику мира во всех ее принципиально непознаваемых количественно-качественных проявлениях.

Взгляды на сложность и сложные системы, сформулированы в исследованиях Е.Н. Князевой работающей в рамках синергетической парадигмы:

– сложность есть множество элементов системы, соединенных нетривиальным образом оригинальными связями друг с другом. Сложность есть динамическая сеть элементов, соединенных по определенным правилам;

– сложность есть внутренне разнообразие системы, разнообразие ее элементов или подсистем, которое делает ее гибкой, способной изменять свое поведение в зависимости от меняющейся ситуации;

– сложность есть *многоуровневость* системы (существует *архитектура сложности*);

– сложные системы являются открытыми системами, т. е. обменивающимися веществом, энергией и/или информацией с окружающей средой. Границы сложной системы порой трудно определить (видение ее границ зависит от позиции наблюдателя);

– сложные системы – это такие системы, в которых возникают *эмерджентные феномены* (явления, свойства), которые не могут быть «вычитаны» из анализа поведения отдельных элементов;

– сложные системы имеют *память*, для них характерно явление гистерезиса, при смене режима функционирования процессы возобновляются по старым следам (прежним руслам);

– сложные системы регулируются петлями обратной связи: отрицательной, обеспечивающей восстановление равновесия, возврат к прежнему состоянию, и положительной, ответственной за быстрый, самоподстегивающийся рост, в ходе которого расцветает сложность [16, с. 77–78].

Князева делает ряд существенных дополнений к классическим взглядам на проблему сложных систем. Во-первых, сложные системы, по ее мнению, являются системами *операционально закрытыми*. Система одновременно является открытой и замкнутой по отношению к окружающей среде. Операциональная замкнутость означает селективность системы, наличие границы упорядочивающей отношения системы со средой и окружающими системами. Система и среда проявляют взаимную активность. Среда меняет систему, но и система активно видоизменяет окружающую среду, вступая в *коэволюцию* с ней. Результатом этого процесса является *структурное сопряжение* сложной системы и среды. Эти взгляды являются развитием концепции аутопоэтических систем У. Матураны и Ф. Варелы [17].

Анализ функционирования психики человека как аутопоэтической системы позволил автору настоящей статьи выдвинуть психологический *принцип тотальной аутопоэтичности психики* [18] в соответствии с которым

все, что конструируется в психике и реальности человека, создается социальной коммуникацией и техногенной средой, носит аутопоэтический характер. Аутопоэтический характер человекообразных систем проявляется на всех уровнях и формах их жизнедеятельности и организации.

Постнеклассическая парадигма научной рациональности позволяет рассматривать интеллект в более широком контексте, как способность сложной системы к организации и самоорганизации среды, проявляющаяся в эмерджентных свойствах организованной/организуемой сущности. В наших работах [19, 20] приведено ряд общих определений и свойств интеллекта и интеллектуальных симбионтов в соответствии с постнеклассической научной методологией:

1. Интеллект есть форма и механизм активной самоорганизации сложной системы, вовлекающие погруженного в среду пользователя/наблюдателя в созидательные, целенаправленные изменения.

2. Интеллект воплощен в среду как механизм ее организации.

3. Интеллект распределен в континууме «система-среда» и воплощен в механизмы самоорганизации системы, действующей в среде.

4. Естественный интеллект представляет собой организующую сложность в организуемой среде, а искусственный интеллект – организованную сложность в организованной среде.

5. Интеллект отражает результаты селекции самоорганизующейся системой эффективных способов достижения цели в организованной среде.

### ***Техногенные и психологические следствия проблемы искусственного интеллекта***

Появление глобальной интеллектуальной социально-коммуникационной информационно-управляющей среды порождает феномены, имеющие ярко выраженную психологическую компоненту, которые нельзя описать методами классической психологии:

– нарушение межкультурного и технологического барьеров и границ между гетерохронными социальными сообществами, возникающими и эволюционирующими в средах электронной коммуникации;

– аутопоэтический характер внутрисетевых коммуникаций, формирующих формы информационного управления субъектами коммуникаций, вовлекаемых в процессы самоорганизации сетевых сообществ;

– появление техногенных интеллектуальных симбионтов [20] в сложных технических системах и средах.

Добавим отмеченные В.В. Чеклецовым феномены:

– размытие границ между цифровым и материальным бытием;

– появление новых сред и форм жизнедеятельности человека в гибридной и виртуальной реальности;

– технологическая трансформация человеческой телесности и ментальности;

– формирование специфических социальных пространств [21].

Рассматриваемые феномены порождают *новые области исследований постнеклассической психологии*, рассматривающие психологические проблемы техногенного мира, насыщенного технологиями искусственного интеллекта:

- процессы формирования маргинальных локальных сетевых сообществ и культур;
- конфликт между формальными и неформальными социальными группами;
- размывание границ личностной и социальной идентичности;
- техногенная модификация личности в сети;
- внегосударственная интеграция сетевых сообществ (формирование глобальной сетевой цивилизации вне существующих государственных структур);
- диффузия и перетекание реальной власти от государственных институтов и парламентских структур к сетевым сообществам;
- сращивание сетевых управляющих сред с глобальными средами, возникновение неравновесных состояний техногенной среды (рост аварийности);
- потеря контроля над процессами самоорганизации социальных и технологических систем, криминализация сетевой среды;
- отсутствие технологий и методов направленного управления процессами организованной сложности.

Проблема искусственного интеллекта порождает перспективные архитектуры эргатических систем, реализация которых на практике возможна лишь при взаимодействии инженеров-проектировщиков с инженерными психологами, работающими в парадигме постнеклассической психологии. К ним можно отнести:

- гибридные системы с индуцированными средами;
- искусственные когнитивные системы;
- самоорганизующиеся сетевые структуры;
- системы с многосредовой самоорганизацией;
- взаимно-ориентирующиеся системы;
- системы с самоорганизацией на базе конкурирующих структур;
- интерсубъектные системы;
- системы с формирующей социальной самоорганизацией.

Исследования данных классов эргатических систем находятся в начальной фазе. Участие психологов и инженеров на этапах формирования их системного и технического облика может послужить источником новых концепций и идей для реализации искусственного интеллекта в самоорганизующихся и развивающихся средах.

**Заключение.** В современных разработках систем и технологий искусственного интеллекта возрастает значение результатов, полученных при исследовании биологических и социальных сложных систем. При этом необходимо понимать, что это системы различного базиса – самоорганизующиеся развивающиеся системы, порождающие человеческий интеллект и алгоритмические цифровые системы, отражающие организованную сложность в случае искусственного интеллекта. Простой перенос результатов, получен-



ных при изучении живых и социальных систем на сложные инженерные системы, некорректен. Целесообразно рассмотрение гибридных форм эргатических систем, включающих цифровой и социобиологический интеллект.

### Библиографический список

1. Айзенк Г.Ю. Понятие и определение интеллекта // Вопросы Психологии. 1995. № 1. С. 111–131.
2. Гарднер Г. Структура разума: теория множественного интеллекта. М.: ООО «И.Д. Вильямс», 2007.
3. Гилфорд Дж. Три стороны интеллекта. Лекция, прочитанная в Стэнфордском университете 13 апреля 1959 года // Психология мышления, под ред. А.М. Матюшкина. М.: Прогресс, 1965. С. 433–456.
4. Практический интеллект / Р. Стернберг, Дж. Б. Форсайт, Дж. Хедланд и др. СПб.: Питер, 2002.
5. Холодная М.А. Существует ли интеллект как психическая реальность? // Вопросы психологии. 1990. № 5. С. 121–128.
6. Холодная М.А. Когнитивные стили. О природе индивидуального ума. Учебное пособие. М.: ПЕР СЭ, 2002.
7. Шкуратова И.П. Когнитивный стиль и общение. Ростов на Дону: Изд-во Ростов. пед. ун-та, 1994.
8. Теплов Б.М. Избранные труды. В 2-х томах, Т. II. М.: Педагогика, 1985.
9. Дубровский Д.И. Сознание, мозг, искусственный интеллект. М.: Стратегия-Центр, 2007. – 272 с.
10. Сергеев С.Ф. Эргономика и инженерная психология техногенного мира: вопросы методологии, теории и практики // Современные тенденции развития психологии труда и организационной психологии / Отв. ред. Л.Г. Дикая, А.Л. Журавлев, А.Н. Занковский. М.: Изд-во «Институт психологии РАН», 2015. С. 47–56.
11. Сергеев С.Ф. Психологические аспекты роботизации в эволюции техногенного мира // Актуальные проблемы психологии труда, инженерной психологии и эргономики. Выпуск 7 / Под ред. А.А. Обознова, А.Л. Журавлева. М.: Изд-во «Институт психологии РАН», 2015. С. 388–407.
12. Степин В.С. Философия и методология науки. М.: Академический проект, Альма Матер, 2015.
13. Сергеев С.Ф. На пути от биоорганизации к киберорганизации: человек в тени искусственного интеллекта // Естественный и искусственный интеллект: методологические и социальные проблемы / Под ред. Д.И. Дубровского и В.А. Лекторского. М.: «Канон+» РООИ «Реабилитация», 2011. С. 48–59.
14. Сергеев С.Ф. Психологические основания проблемы искусственного интеллекта // Мехатроника, автоматизация, управление. 2011. № 7. С. 2–6.
15. Сергеев С.Ф. Искусственный интеллект в границах исчезающей сложности // Тринадцатая национальная конференция по искусственному ин-

теллекту с международным участием КИИ-2012 (16–20 октября 2012 г., г. Белгород, Россия): Труды конференции. Т.4. Белгород: Изд-во БГТУ, 2012. С. 180–187.

16. Князева Е.Н. Темпоральная архитектура сложности // Синергетическая парадигма. «Синергетика инновационной сложности». М.: Прогресс-Традиция, 2011. С. 66–86.

17. Матурана У., Варела Ф. Древо познания. Биологические корни человеческого понимания. М.: Прогресс-Традиция, 2001. 224 с.

18. Сергеев С.Ф. Механизм тотальной аутопоэтичности человекообразных систем // Нейронаука в психологии, образовании, медицине: Сб. статей / Под науч. ред. Т.В. Черниговской, Ю.Е. Шелепина, В.М. Аллахвердова, С.Н. Костроминой, О.В. Защириной. СПб: «ЛЕМА», 2014. С.134–140.

19. Сергеев С.Ф. Искусственный и естественный интеллекты в техногенных образовательных средах // Открытое образование. 2013. № 2 (97). С. 52–60.

20. Сергеев С.Ф. Интеллектуальные симбионты организованных техногенных средств управления подвижными объектами // Мехатроника, автоматизация, управление. 2013. № 9. С. 30–36.

21. Чеклецов В.В. Чувство планеты (Интернет Вещей и следующая технологическая революция). М.: Российский исследовательский центр по Интернету Вещей, 2013. 130 с.

## ARTIFICIAL AND NATURAL INTELLIGENCE IN DIGITAL AND BIOLOGICAL SYSTEMS

*Sergeev Sergey F.*

Saint Petersburg state University

7-9 Universitetskaya Embankment, St Petersburg, Russia, 199034,  
ssfpost@mail.ru

The article provides a brief analysis of the history of the emergence and development of artificial intelligence issues and related topical areas in the field of General psychology, engineering psychology and ergonomics. When creating artificial intelligence systems and technologies generated by the evolution of the technogenic environment of human civilization, it is necessary to take into account the issues of ensuring human interaction with the organized evolving environment. The article considers promising areas of psychological research and possible problems when using the conceptual and conceptual apparatus of post-non-classical psychology and the philosophy of self-organizing complex systems in these areas.

**Keywords:** intelligence, artificial intelligence, post-non-classical psychology, self-organization, organized complexity.

# **ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ И ФИЛОСОФИЯ**

## КОМПЬЮТЕРНЫЕ СРЕДСТВА КОМПЛЕКСНОГО ТЕСТА ТЬЮРИНГА

*Алексеев Андрей Юрьевич*

МГИМО МИД РФ, Центр искусственного интеллекта, 143007,  
Московская область, г. Одинцово, ул. Ново-Спортивная, д.3,  
РУДН, Инженерная академия, департамент механики  
и процессов управления, 115419, г. Москва, ул. Орджоникидзе, д. 3  
aa65@list.ru

В статье представлены логико-лингвистические и программно-информационные средства реализации комплексного теста Тьюринга. Данная компьютерная система за счет систематизации, координации и интеграции частных тестов Тьюринга направлена на решение общего вопроса «Может ли компьютер всё?», продолжая решение оригинального тьюрингового вопроса «Может ли машина мыслить?».

**Ключевые слова:** искусственный интеллект, комплексный тест Тьюринга.

**Введение.** Комплексный тест Тьюринга (КТТ) – настоящий, истинный, правильный тест Тьюринга – это собирательное понятие для исследований компьютерной системы на предмет выявления в ней когнитивных функций. КТТ состоит из сотен версий оригинальной тьюринговой «игры в имитацию интеллекта» (1950 г.) и многочисленных версий этих версий, которые на сегодняшний день раскрывают достаточно сложную когнитивную феноменологию компьютерного мира электронной культуры: может ли машины мыслить, понимать, творить, сознавать, самосознавать, любить, быть личностью и пр.

### **Важность компьютерной поддержки комплексного теста Тьюринга**

Идея комплексного теста Тьюринга (КТТ) прорабатывалась автором на протяжении двух десятилетий и представлена в организации конференций «Философия искусственного интеллекта» [1, 2], формулировании идеи [3], написании статей (например [4, 5]), монографии [6], диссертации [7]. В одной из последних выступлений [8] доказывалось, что комплексный тест Тьюринга является формальным определением ИИ (по аналогии с тем, что машина Тьюринга является формальным определением понятия «алгоритм»), а ИИ – это квазиалгоритмическая реализация КТТ.

Пожалуй, значимость идеи КТТ в наибольшей степени подчеркивается дистинкцией между логически достаточными и логически необходимыми условиями развития ИИ. Логически достаточные условия ИИ – это те, кото-

рые способствуют развитию ИИ. К ним относится трудно регистрируемая совокупность направлений, мероприятий, проектов и др. деятельности. Например, как предлагает Технический комитет 164 «Искусственный интеллект», это проекты по распознаванию образов, категорированию, построению моделей окружающих объектов и процессов; поиску решений; реализации физических воздействий на окружающую среду; автономное движение и позиционирование в пространстве; социальные коммуникации [9]. Логически достаточных условий развития ИИ много. К этим условиям относится, например, развитие Пермской школы нейроинформатики под руководством Л.Н. Ясницкого и, скажем, развитие робототехники фирмой «Промобот».

Логически необходимых условий намного меньше. Это те условия, при отсутствии которых отсутствует само развитие ИИ. Тест Тьюринга и есть такое необходимое условие. Бесспорным является то, что работы по ИИ предваряются убеждением разработчика следующего характера: либо может компьютер (машина) мыслить, либо не может компьютер мыслить, либо когда-то сможет компьютер мыслить, либо никогда не сможет компьютер мыслить. Слово «мыслить» в данном контексте репрезентирует много других когнитивных феноменов (понимать, творить, сознавать, любить и пр.). Это – вопросы теста Тьюринга. То есть комплексный тест задает необходимые условия развития ИИ, так как поднимает принципиальные вопросы о компьютерной реализации разнообразного спектра когнитивных феноменов.

Исходя из важности изучения логически необходимых условий ИИ, существующей сложности организации КТТ (сотня тьюринговых тестов!) и сложности исследования вопросов метауровневого порядка относительно уровня вопросов изучения достаточных условий ИИ, не подлежит сомнению актуальность и важность разработки средств компьютерной поддержки КТТ.

### **Основные функции КТТ**

В литературе выделяется не менее сотни крупных версий теста Тьюринга. Мы ряд версий обобщили в [6] и предложили изучать двадцать одну версию. Эти версии обозначены как совершенные частные тесты Тьюринга. В них четко просматриваются базовые функции, предложенные в оригинальном тесте. Этим функций пять:

1) интеррогативная функция, характеризующая содержание и форму вопросов частного теста, которые задает тьюринговый судья (interrogator, наблюдатель) при изучении х-системы на предмет ее у-способностей;

2) дефинитная функция, обеспечивающая компьютерное определение когнитивного феномена;

3) конструкторская функция, раскрывающая принципы работы компьютера, способного реализовать когнитивный феномен;

4) критическая функция, отражающая суть полемики по поводу возможности компьютерной реализации когнитивного феномена;

5) конститутивная функция, позиционирующая отношение судьи (наблюдателя) к когнитивным аспектам компьютерной реальности.

## Язык описания частных тестов Тьюринга

Для формирования базы данных частных тестов тестов Тьюринга ранее в [6] был разработан язык описания частных тестов Тьюринга (ЧТТ). Язык предназначен для идентификации, систематизации, унификации, координации, обобщения, дифференциации, интеграции частных ТТ и их составляющих. Это нестрогий язык, он не предполагает четкой аксиоматики и выполняет в большей степени роль систематизации библиографических описаний ЧТТ, представленных в соответствующих работах. Так, для первого этапа игры в имитацию, на котором играют мужчина и женщина, убеждающие судью в своей гендерной принадлежности, формулировка ЧТТ представляется следующим образом:

$$\text{ИвИ1: } M[\mathcal{J}], \mathcal{J}[\mathcal{J}] | \Rightarrow C[\mathcal{C}]: (a=M \ \& \ b=\mathcal{J}) \vee (a=\mathcal{J} \ \& \ b=M),$$

где:

1) *ИвИ1*: – обозначение (уникальное имя формулы) для идентификации частного теста Тьюринга. Если далее идет скобочное выражение, то это означает, что данный тест является производным от теста, имя которого приводится в скобках;

2)  $X[Y]$  – выражение, подстановка в которое символов  $M, \mathcal{J}, K, C$  и других обозначает основное для тьюринговой игры отношение подражания (имитации):  $X$  подражает  $Y$  (или  $X$  имитирует  $Y$ );

3)  $|$  – символ, обозначающий стену Тьюринга;

4)  $\Rightarrow$  – аналог имплекативной формы, обозначающий заключение (оценку) судьи, например, для *ИвИ1* – заключение о гендерной идентичности игрока;

5)  $:-$  – отметка начала определения цели теста;

6)  $a, b...$  – индивидуальные константы для обозначения конкретных игроков;

7)  $\&, \vee, \neg \dots$  – обозначения стандартных логических связок (конъюнкции, строгой дизъюнкции, отрицания и т.д.);

Также в языке описания КТТ используются:  $x, y...$  – переменные, пробегающие по элементам класса игроков; специфические литеральные и функциональные обозначения, например,  $t$  – длина ЧТТ (время оценивания), *Архитектура (K)* – архитектура компьютера; пример диалога между судьей и  $x$ -системой: *Тест { перечень тестовых вопросов-ответов }* стандартные технические символы – скобки, запятые, точки.

### Примеры формулировок частных тестов Тьюринга

Приведем примеры формул. Вначале коснемся тестов, которые касаются оригинальной игры в имитацию. Второй этап игры в имитацию, на котором роль женщины начинает выполнять компьютер, представляется следующим образом:

$$\text{ИвИ2: } K[\mathcal{J}], \mathcal{J}[\mathcal{J}] | \Rightarrow C[\mathcal{C}]: (a=M \ \& \ b=\mathcal{J}) \vee (a=\mathcal{J} \ \& \ b=M)$$

Третий этап игры в имитацию, когда компьютер начинает выполнять роль мужчины, выглядит следующим образом:

$$\text{ИвИЗ: } K[M], M[\mathcal{J}], t, \text{ Архитектура}(K), \text{ Вероятность Ошибки}(C) \mid \Rightarrow \\ C[C]: (a=M \ \& \ b=\mathcal{J}) \vee (a=\mathcal{J} \ \& \ b=M)$$

Для него вводятся параметры времени тестирования (5 минут), вероятность расхождения случаев правильной идентификации пола игрока (30 процентов) и обсуждаются принципиальные вопросы об архитектуре компьютера. Канонические версии теста Тьюринга представляются в следующем виде. Первая форма предназначена для игры, в которой участвуют люди и компьютеры:

$$\text{ТТ1: } K[\mathcal{C}], \mathcal{C}[\mathcal{C}] \mid \Rightarrow C[\mathcal{C}]: (a=K \ \& \ b=\mathcal{C}) \vee (a=\mathcal{C} \ \& \ b=K)$$

Вторая каноническая форма выявляет человечность или компьютерность тестируемой системы.

$$\text{ТТ2: } K[\mathcal{C}] \mid \Rightarrow C[\mathcal{C}]: (a=K) \vee (a=\mathcal{C})$$

В третьем случае судья определяет X-систему (это не обязательно компьютер, может быть, планета Солярис) на предмет её «человечности»:

$$\text{ТТ3: } K[X] \mid \Rightarrow C[\mathcal{C}]: (X=\mathcal{C}) \vee \neg(X=\mathcal{C})$$

В четвертом случае изучаются некоторые неизвестные когнитивные компетенции X:

$$\text{ТТ4: } K[X] \mid \Rightarrow C[\mathcal{C}]: \{\text{когнитивные феномены}(X)\}.$$

Возможны иные вариации канонических тестов Тьюринга. Например, когда роль судьи начинает исполнять машина, т.е. вместо  $C[\mathcal{C}]$  будет  $C[K]$ , как это принято в инвертированном тесте Тьюринга С.Ватта, возникает фундаментальная трансформация всех выделенных тестов. В самом деле, искусственный интеллект (искусственное сознание) начинает задавать и определять компетенции естественного интеллекта (естественного сознания).

Для предложенного формализма возможны разные комбинаторные сочетания. Например, формула теста Мура [монография, С.] принимает следующий вид:

$\text{ТТ}_\text{Мура}: K[\mathcal{C}] \mid \Rightarrow C[\mathcal{U}]:$  *Индуктивное построение теории ИИ*, где  $\mathcal{U}$  – ученый-когнитолог, а последний член формулы означает то, что ТТ позволяет прямое или косвенное тестирование фактически всей деятельности, которая считается доказательством мышления; ТТ поощряет жестокий отбор; ТТ – это достаточное условие интеллектуальности системы.

Однако следующий тест (как ответ Д.Сталкера на предыдущий тест) отрицает последний член формулы.

$$\text{ТТ}_\text{Мура}(\text{Сталкер}): K[\mathcal{C}] \mid \Rightarrow \neg C[\mathcal{U}]: \text{Индуктивное построение теории ИИ}$$

Д. Сталкер доказывает, что интерпретация Дж.Мура в большей мере декларативная, нежели чем индуктивная. Нужна чисто механическая теория, которая не обращается к ментальным понятиям типа «интеллект», но обра-

щается к понятиям: структура компьютера, программа и физическая среда. Объяснительная теория, которая включает в себя понятия о мышлении, может относиться только к людям. Из-за фундаментальных различий между компьютерами и людьми компьютерные теории (теории искусственного интеллекта) нельзя применять для объяснения естественного интеллекта.

Данный тест можно считать предтечей тесту Сёрла, который формулируется следующим образом, когда Серль-в-комнате, работая, то как человек, то как компьютер, оценивает феномен собственного «понимания»:

*ТТ\_Серль1: (Серль-в-комнате[K] |  $\Rightarrow$  C[Китаец]):  
Понимание(a) &  $\neg$ Понимание(Серль-в-комнате[K])) &  
& (Серль-в-комнате[Ч] |  $\Rightarrow$  C[Англичанин]):  
Понимание(a) & Понимание(Серль-в-комнате[Ч]))*

Здесь используется общая формула «Китайской комнаты». Феномена «понимания» нет тогда, когда компьютер имитирует человека, и тогда, когда человек воспроизводит функционирование компьютера. Получается вторая версия теста Сёрла.

*ТТ\_Серль2: (K[Ч] |  $\Rightarrow$  C[Ч]):  $\neg$ Понимание(a) & (Ч[K] |  $\Rightarrow$  C[Ч]):  $\neg$ Понимание(a)*

Используя предложенный язык ЧТТ имеется возможность формулирования достаточно большого набора исходных формулировок частных тестов, формулировок, производных от исходных формулировок частных тестов, формулировок всевозможных комбинаций предыдущих формулировок. То есть в ходе работы компьютерных средств комплексного теста Тьюринга возможно построение весьма широкой когнитивной феноменологии компьютерного мира и оценка своей роли в ней.

**Заключение.** Компьютерные средства комплексного теста Тьюринга, стандартным образом поддерживающие язык описания частных тестов Тьюринга, представляются весьма важным средством решения вопроса, логически необходимого для запуска всевозможных проектов искусственного интеллекта. Если неубедительны предложенные частные тесты для компьютерной имитации скажем, феномена понимания или феномена сознания, то надо уменьшить границы притязаний или говорить о некоторых формах квазипонимания или квазисознания, но не о понимании и сознании как таковых.

Визуальные приложения комплексного теста Тьюринга представляются весьма родственными средствам визуального программирования. То есть при определении инструментов, каждый из которых реализует частные вариации интеррогативной, дефинитной, критической, конструирующей и конститутивной функций, мне как судье из комплексного теста Тьюринга предоставляются богатые возможности конструирования когнитивной феноменологии компьютерного мира, удобного для меня и моего социального окружения.

### Библиографический список

1. Методологические и теоретические аспекты искусственного интеллекта. Материалы студенческой конференции «Философия искусственного интеллекта», МИЭМ, 20 мая 2004 г. Под ред. А.Ю. Алексева – М.: МИЭМ, 2006. – 192 с.



2. Философия искусственного интеллекта. Материалы Всероссийской междисциплинарной конференции, г. Москва, МИЭМ, 17 – 19 января 2005 г. – М.: ИФ РАН, 2005. – 400 с.

3. Алексеев А.Ю. Возможности искусственного интеллекта: можно ли пройти тесты Тьюринга // В кн. Искусственный интеллект: Междисциплинарный подход. Под ред. Д.И. Дубровского и В.А. Лекторского. – М.: ИИнтелЛ, 2006. – 448 с. – С. 223-243.

4. Алексеев А.Ю. Роль комплексного теста Тьюринга в методологии искусственных обществ//Ежеквартальный Интернет – журнал «Искусственные общества», Том 6, номер 1-4, I-IV квартал 2011 г. – С. 18-64

5. Алексеев А.Ю., Гарбук С.В. Каким системам искусственного интеллекта можно доверять? Объективные, субъективные и интересубъективные параметры доверия// Искусственные общества. – 2022. – Т. 18. – Выпуск 2, 25.05.2022

6. Алексеев, А.Ю. Комплексный тест Тьюринга: философско-методологические и социокультурные аспекты. – М.: ИИнтелл, 2013 г. – 304 с. – ISBN: 978-5-88387-090-2

7. Алексеев А.Ю. Философия искусственного интеллекта: концептуальный статус комплексного теста Тьюринга. Диссертации на соискание ученой степени доктора философских наук по специальности 09.00.08 – философия науки и техники, философский факультет МГУ имени М.В.Ломоносова, диссертационный совет Д 501.001.37, г. Москва, 2016 г.; URL: <https://istina.msu.ru/dissertations/15422339/>

8. Алексеев А.Ю. Формальное определение искусственного интеллекта. Доклад на семинаре «Мыслить вместе со сложностью», 12 июня 2022 г., ИФ РАН; URL:[https://www.youtube.com/watch?v=O9r6zashLns&list=PLBx8d8IAB10o\\_XMiAB7vVeMn2xIs-EDvZ&index=77](https://www.youtube.com/watch?v=O9r6zashLns&list=PLBx8d8IAB10o_XMiAB7vVeMn2xIs-EDvZ&index=77), дата обращения: 1.10.2022

9. Гарбук С.В., Губинский А.М. Искусственный интеллект в ведущих странах мира: стратегии развития и военное применение. – М.: «Знание», 2020, 590 с.

## COMPUTER TOOLS FOR COMPREHENSIVE TURING TEST

*Alekseev Andrey Yu.*

MGIMO, Center for Artificial Intelligence, 143007,  
Moscow region, Odintsovo, st. Novo-Sportivnaya, 3,

RUDN University, Engineering Academy, Department of Mechanics  
and Control Processes, 115419, Moscow, st. Ordzhonikidze, 3, aa65@list.ru

The article presents the logical-linguistic and software-informational means of implementing the comprehensive Turing test. This computer system, through the systematization, coordination and integration of particular Turing tests, is aimed at solving the general question “Can a computer do everything?”, continuing the solution of the original Turing question “Can a machine think?”.

**Keywords:** artificial intelligence, comprehensive Turing test.

# **ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ В МЕДИЦИНЕ И ЗДРАВООХРАНЕНИИ**

# НЕЙРОСЕТЕВОЙ КОРРЕЛЯТОР ДЛЯ РАЗДЕЛЕНИЯ АДДИТИВНОЙ СМЕСИ КАРДИОРЕСПИРАТОРНЫХ СИГНАЛОВ ДВУХ И БОЛЕЕ ПАЦИЕНТОВ

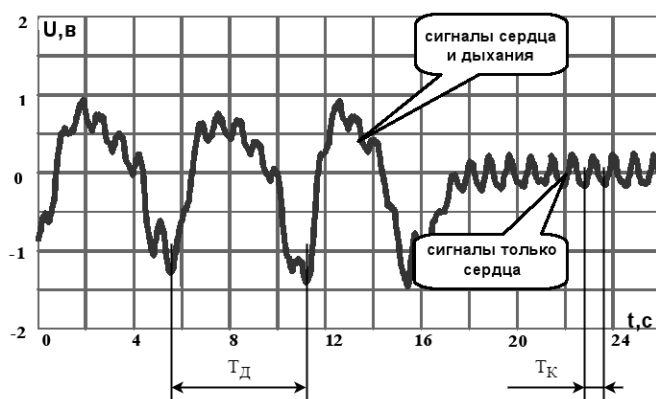
*Бутусов Андрей Владимирович, Сафронов Руслан Игоревич,  
Филист Сергей Алексеевич*

Юго-Западный государственный университет,  
305040, Россия, г. Курск, ул. 50 лет Октября, 94, SFilist@gmail.com

В статье предложена концептуальная модель выделения и идентификации динамических параметров дыхания и сердцебиения из радиоволнового сигнала, отраженного от нескольких пациентов. Модель раскрыта для случая двух пациентов, но она может быть обобщена и для большего количества пациентов. В основу модели положено разделение сигнала дыхания и электрокардосигнала от биорадиосенсора и их двумерный Фурье-анализ.

**Ключевые слова:** биорадиосенсор, сигнал дыхания, кардиосигнал, двумерный Фурье-анализ, нейронная сеть

**Введение.** Одним из наиболее эффективных методов оценки и диагностики психофизиологического состояния человека – оператора, является непрерывный анализ его функциональных состояний по сердечному ритму (кардиоинтервалометрия) и по ритму дыхания. Известно множество технических, алгоритмических и программных решений, реализующих этот метод, с использованием биорадиосенсоров [1]. Однако сигнал дыхания и сигнал ритма сердца не являются гармоническими сигналами, что вызывает трудности их идентификации частотными методами. На рисунке 1 приведен пример сигналов дыхания и кардиосигнала на выходе биорадиосенсора. В связи с тем, что они не гармонические, их спектр занимает некоторую полосу в диапазоне ритма дыхания и кардиоритма, что затрудняет измерение частоты дыхания и частоты сердечных сокращений посредством методов Фурье-анализа.



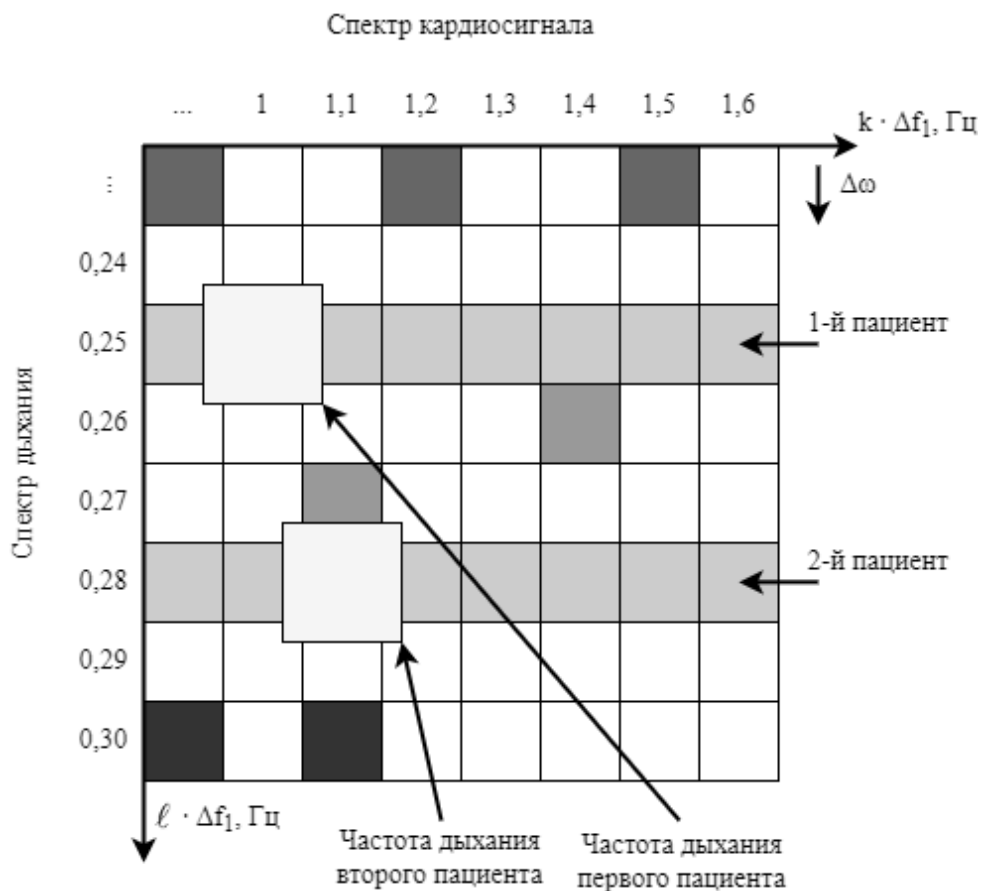
**Рисунок 1. Примеры сигналов дыхания и кардиосигнала на выходе биорадиосенсора**

Для определения истинной частоты дыхания спектру необходимо подвергнуть дальнейшей обработки, например, используя гистограммные методы или методы машинного обучения.

Если электромагнитная волна отражается от нескольких пациентов, то в спектре выходного сигнала биорадиолокатора появляются составляющие, обусловленные сложением двух гармонических составляющих близких частот, что приводит к эффекту, известному в радиотехнике как биения.

**Материалы и методы.** В [1] было показано, что для разделения сигнала биорадиорадара от нескольких пациентов целесообразно использовать вейвлет-преобразование. В данной работе для решения аналогичной задачи в качестве частотно-временного преобразования используется двумерная частотная плоскость (ДЧП) [5, 6]. Специфика использования ДЧП заключается в декомпозиции сложно модулированного сигнала на «быстрые» и «медленные» волны, спектры которых представлены как ортогональные отсчеты на ДЧП. В данном случае в качестве «быстрых» волн выступает кардиоритм, а в качестве «медленных» – ритм дыхания, так как в случае отражения радиосигнала от нескольких пациентов носителями «медленных» волн будут различные ритмы «быстрых» волн. Если имеет место суперпозиция «быстрых» ритмов, то можно предположить, что их дислокация не перекрывается в области «медленных» волн ДЧП.

На рисунке 2 представлен макет ДЧП с соответствующими «быстрыми» и «медленными» ритмами.



**Рисунок 2. Структура двумерной частотной плоскости**

Для получения ДЧП необходимо провести декомпозицию сигнала по «быстрым» ритмам. С этой целью исходный сигнал сегментируется на квазипериоды  $T_k$  (рисунок 1), которые используются в качестве окна для оконного преобразования Фурье (ОПФ). Отсчеты квазипериодов записываются в строках матрицы  $X$  с элементами  $x_{nm}$ , где  $n$ -номер квазипериода,  $m$ -номер отсчета в квазипериоде. Декомпозиция ритмов на ДЧП основана на том, что сигнал «быстрых» волн сегментирован, и каждый сегмент-квазипериод выступает как многомерный вектор дискретных отсчетов «медленной» волны. Определив спектры сегментов, можем исследовать их эволюцию во времени, которая, в свою очередь, определяет эволюцию «медленных» волн.

Таким образом, определяя в матрице  $X$  спектр каждой строчки, переходим к матрице  $Y$ , элемент которой определяются по формуле

$$y_{nk} = \frac{1}{M} \sum_{m=0}^{M-1} x_{nm} \cdot e^{-j2\pi km/M}, \quad (1)$$

где  $M$ -число отсчетов в квазипериоде.

Полагаем, что каждая гармоническая составляющая в строке матрицы  $Y$  отражает эволюцию «медленного» ритма. Поэтому спектр столбца матрицы  $Y$  будет отражать «медленные» волны, то есть ритм дыхания. Для его получения необходимо выполнить ОПФ столбцов матрицы  $Y$ , ширина окна которого определяется числом квазипериодов или числом строк матрицы  $Y$ . В итоге получаем матрицу  $D$ , элементы которой находим по формуле

$$z_{\ell k} = \frac{1}{N} \sum_{n=0}^{N-1} y_{nk} \cdot e^{-j2\pi \ell n/N}, \quad (2)$$

где  $N$ -число квазипериодов в кардиосигнале.

В результате визуализации матрицы  $Z$  получаем ДЧП.

**Результаты.** В результате двумерного спектрального преобразования на ДЧП (рисунок 2) получаем две полосы, соответствующие ритмам дыхания первого и второго пациента. Каждая двумерная частота в этих полосах (строках) характеризует вклад ритма дыхания в эволюцию соответствующей гармоники «быстрого» ритма (кардиоритма). Для идентификации ритма дыхания отдельного пациента на ДЧП, используют сигнал ритма дыхания (рисунок 1).

Несмотря на то, что отраженная волна модулируется сигналами, определяемыми механическими движениями грудной клетки и сердца, эти сигналы всегда можно разделить путем выбора соответствующей частоты дискретизации на выходе или на входе биорадиосенсора. Поэтому рассмотрим модели обработки выходного сигнала дыхания и кардиосигнала отдельно. Учитывая то, что верхняя граница спектра кардиосигнала не превышает 50 Гц, частоту дискретизации для канала кардиосигнала установим 100 Гц. Верхняя граница сигнала дыхания не превышает 1 Гц, поэтому частоту дискретизации этого сигнала установим 2 Гц. Таким образом, получим двухканальный сигнал. В первом канале имеем сигнал дыхания, а во втором канале имеем кардиосигнал, модулированный сигналом дыхания. Согласно [7, 8]

имеет место как амплитудная, так и частотная модуляция. В данной работе рассматривается только амплитудная модуляция.

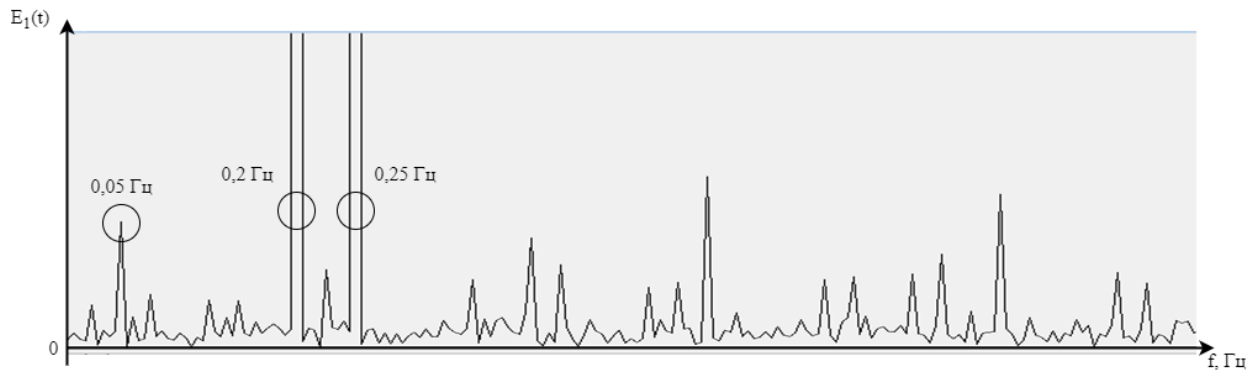
Полагаем, что частота дыхания первого пациента  $\omega_1$ , а второго  $\omega_2$ . При этом  $\Delta\omega = \omega_1 - \omega_2$  мала, то есть  $\Delta\omega \ll \omega_1, \omega_2$ . Если  $x_1(t) = A \cos \omega_1 t$ , а  $x_2(t) = A \cos(\omega_1 + \Delta\omega)t$ , то их сумма равна

$$x(t) = x_1(t) + x_2(t) \approx 2A \cos \frac{\Delta\omega}{2} t \cdot \cos \omega t. \quad (3)$$

На рисунке 3 представлен спектр модели сигнала дыхания от двух пациентов. Модель построена согласно выражению

$$U_1(t) = 100 \sin((2\pi \cdot 0,2) \cdot t) + 80 \sin((2\pi \cdot 0,25)t) \quad (4)$$

Для того, чтобы «увидеть» разностную частоту 0,05 Гц сигнал необходимо наблюдать 20 с, а для надежного выделения необходимо иметь на апертуре хотя бы 10 периодов этой частоты. ОПФ этого сигнала с отмеченными частотами дыхания двух пациентов и разностной частотой показан на рисунке 3.



**Рисунок 3. Спектр Фурье сигналов дыхания двух пациентов**

Таким образом, разделение сигналов дыхания двух пациентов к согласованию (синхронизации) одномерного спектра сигнала дыхания (рисунок 3) с двумерным спектром рисунок 2. Этот процесс назовем процессом синхронизации. Процесс синхронизации включает использование пороговых методов, решающих правил продукционного типа, а также использование обучаемых классификаторов.

В данной работе для определения синхронизма на ДЧП определялись две корреляционные функции

$$\Phi(\Delta_1, \Omega) = \int_0^{T_1} z(\omega, \Omega) \cdot z(\omega - \Delta_1) d\omega, \quad (5)$$

$$\Psi(\Delta_1, \Delta_2) = \int_0^{T_2} \Psi(\Delta_1, \Omega) \cdot \Psi(\Delta_1, \Omega - \Delta_2) d\Omega, \quad (6)$$

где  $\Delta_1 = \overline{0, (\text{int}(M/2) - 1)}$ ,  $\Delta_2 = \overline{0, (\text{int}(N/2) - 1)}$ ,  $\omega$  – горизонтальная частота ДЧП,  $\Omega$  – вертикальная частота ДЧП,  $T_1$  – длительность квазипериода ДЧП,  $T_2 = T_1 \cdot N$ .

Для построения корреляторов (5) и (6) используем многоступенчатые нейросетевые структуры. На первой ступени определяется число пациентов участвующих в формировании отраженного сигнала. Так как вероятность того, что кардицикл двух пациентов одинаков очень мала, то составляющие спектров дыхания будут смещены относительно друг друга вдоль вертикальной координаты ДЧП, как это показано на рисунке 2. Это смещение определяется разрешением по частоте вдоль горизонтальной оси ДЧП и определяется длиной квазипериода. То есть если добиваться разрешения хотя бы не менее 0,1 с, то длина квазипериода не должна быть менее 10 с. Чтобы получить частотное разрешение 0,01 Гц по вертикальной оси необходимо иметь 10 десятисекундных отсчетов, то есть ДЧП должна содержать не менее десяти строчек. Однако в таком случае наивысшая частота по вертикальной оси составит всего 0,05 Гц, что недостаточно для обнаружения сигнала дыхания. Поэтому десяти секундные строки формируются не через десять секунд, а через одну секунду, что позволяет получить верхнюю частоту по вертикальной оси 0,5 Гц, что достаточно для наблюдения на ДЧП сигнала дыхания.

На второй ступени используем нейронную сеть – двумерный коррелятор. Нейронная сеть настраивается на дата сет реальных сигналов с биорадиосенсора, полученных от двух пациентов. Также для настройки нейронной сети использовались сигналы имитационного моделирования, примеры которых показаны на рисунке 3.

В качестве входного сигнала нейронной сети используется ДЧП пациента – неизвестного образца. На выходе нейронной сети имеем матрицу  $N \times M$ , элементы которой принимают значения либо ноль, либо единица. Причем значение единица могут принять только два элемента этой матрицы (при наличии только двух пациентов), которые идентифицируют частоту сердечных сокращений и частоту дыхания. Учитывая, что  $\Delta_1$  и  $\Delta_2$  дискретны, то они и будут отражать эти показатели на выходе нейронной сети.

При необходимости от нелинейной функции активации можно перейти к линейной функции активации и тогда на выходе нейронной сети будем иметь полутоновую матрицу, которую может анализировать лицо, принимающее решение, или она может быть передана на следующую ступень обработки.

**Заключение.** В результате проведенных исследований предложена концептуальная модель выделения и идентификации динамических параметров дыхания и сердцебиения из радиоволнового сигнала, отраженного от нескольких пациентов. Модель раскрыта для случая двух пациентов, но она может быть обобщена и для большего количества пациентов. В основу модели положено разделение сигнала дыхания и кардиосигнала от биорадиосенсора и их двумерный Фурье-анализ. Введено понятие синхронизации сигнала дыхания и кардиосигнала, позволяющее по введенному критерию синхрони-

зации селективировать строки двумерной частотной плоскости, принадлежащие различным пациентам.

### Библиографический список

1. Иммореев, И.Я. Практическое использование сверхширокополосных радаров / И.Я. Иммореев // Журнал радиоэлектроники. 2009. № 9. – С.1-3.
2. Филист, С.А. Способ оценки функционального состояния организма человека на основе анализа двумерных частотных плоскостей кардиосигналов / С.А. Филист, Н.А. Корневский, Т.М. Шотланд // Системный анализ и управление в биомедицинских системах. 2003. Т. 2. № 2. – С. 85-88.
3. Филист, С.А. Методы двумерного спектрального преобразования электрокардиосигналов в ранней диагностике сердечно-сосудистых заболеваний / С.А. Филист // Биомедицинские технологии и радиоэлектроника. 2001. № 3. С. – 14-20.
4. Петрова Т.В. Предикторы синхронности системных ритмов живых систем для классификаторов их функциональных состояний // Т.В. Петрова, С.А. Филист, С.В. Дегтярев, А.В. Киселев, О.В. Шаталова // Системный анализ и управление в биомедицинских системах. 2018. Т. 17. № 3. С. 693-700.
5. Filist, S. Developing neural network model for predicting cardiac and cardiovascular health using bioelectrical signal processing/ S. Filista, R.T. Al-kasasbeh, O. Shatalova, A. Aikeyevac, N. Korenevskiya, A. Shaqadand, A. Trifonova and M. Ilyashe // Computer methods in biomechanics and biomedical engineering. Volume 25, 2022 – Issue 8. – Pages 908-921. <https://doi.org/10.1080/10255842.2021.1986486>.

## NEURAL NETWORK CORRELLATOR FOR SEPARATING ADDITIVE MIXTURE OF CARDIORESPIRATORY SIGNALS FROM TWO OR MORE PATIENTS

*Butusov Andrey V., Safronov Ruslan I., Filist Sergey A.*

Southwest State University

305040, Russia, Kursk, st. 50 years of October, 94, SFilist@gmail.com

The article proposes a conceptual model for isolating and identifying the dynamic parameters of respiration and heartbeat from a radio wave signal reflected from several patients. The model is disclosed for the case of two patients, but it can be generalized to more patients. The model is based on the separation of the respiratory signal and the electrocardiosignal from the bioradio sensor and their two-dimensional Fourier analysis.

**Keywords:** bioradiosensor, respiration signal, cardiac signal, 2D Fourier analysis, neural network



## НЕЙРОСЕТЕВАЯ СИСТЕМА ПРОГНОЗИРОВАНИЯ УРОВНЯ ВАКЦИНАЦИИ ОТ COVID-19

*Субботина Екатерина Викторовна*

Национальный исследовательский университет «Высшая школа экономики»,  
614070, Россия, г. Пермь, ул. Студенческая, 38, infoperm@hse.ru

В статье представлено описание разработки нейросетевой системы для прогнозирования уровня вакцинации от COVID-19. Система позволяет с большой точностью предсказать количество вакцинированных людей на основании данных о заболеваемости и смертности от COVID-19. С помощью разработанной интеллектуальной системы проведено исследование предметной области, выявлены закономерности, имеющие практическое значение.

**Ключевые слова:** искусственный интеллект, нейросетевые технологии, прогнозирование, вакцинация, COVID-19, коронавирусная инфекция.

**Введение.** Первый случай заражения коронавирусной инфекцией впервые был зафиксирован в декабре 2019 года, уже 30 января Всемирная организация здравоохранения объявила вспышку COVID-19 «чрезвычайной ситуацией в области общественного здравоохранения, имеющей народное значение», а 11 марта – пандемией. Из-за быстрого распространения инфекции и высокого уровня смертности от нее вопрос разработки вакцины особенно остро встал перед медициной всей стран. Первая вакцина от коронавируса «Convivisea» появилась 25 июня 2020 года в Китае для вакцинации военнослужащих. Первая вакцина, которая стала доступна для всего общества, появилась 11 августа 2020 года в России, вакцина получила название «Гам-КОВИД-Вак» или Спутник V. Уже 25 августа 2020 года вакцина «Гам-КОВИД-Вак» была введена в гражданский оборот.

Многие страны внедрили планы поэтапной вакцинации населения. По этим планам приоритет отдаётся тем, кто подвержен наибольшему риску осложнений, например, пожилым людям и тем, кто подвержен высокому риску заражения и передачи, например, медицинским работникам. Соответственно, перед каждым человеком возник вопрос: вакцинироваться или нет? Эффективность вакцин, то есть относительное снижение риска заболевания у привитых по сравнению с непривитыми, до сих пор вызывает сомнения у людей. Кто-то же просто боится вакцинироваться, так как считают, что вакцины недостаточно протестированы, кто-то просто не верит в их эффективность, у некоторых людей имеются противопоказания: аллергические реакции, хронические заболевания, беременность, возраст до 18 лет и так далее. Тем не менее, по данным на 29.11.2021 количество вакцинированных в России составляет 45,38% населения, что говорит о том, что практически половина населения предпочли поставить прививку, однако пока не ясно, когда

будет вакцинировано большинство населения, что особенно важно в период пандемии. Директор Национального исследовательского центра эпидемиологии и микробиологии Н.Ф. Гамалеи отметил, что пандемия COVID-19 закончится, когда 70-75% населения Земли будут привиты. Именно поэтому прогнозирование уровня вакцинации от COVID-19 является актуальным.

При анализе литературных источников выяснилось, что работ на тему прогнозирования уровня вакцинации от COVID-19 было проведено очень мало, однако многие ученые занимались построением нейронных сетей для прогнозирования распространения коронавирусной инфекции. Так, например, в статье [24] авторы представили результаты исследования по разработке модели нейронной сети для прогнозирования распространения COVID-19. Предиктор основан на классическом подходе с глубокой архитектурой, которая обучается с использованием модели обучения NAdam. Для обучения нейронной сети авторы использовали официальные данные из государственных и открытых репозиторий. Результаты прогнозирования выполняются для стран, а также регионов, чтобы обеспечить широкий спектр значений прогнозируемого распространения COVID-19. Результаты предложенной авторами модели показывают высокую точность, которая в некоторых случаях достигает более 99%.

В статье [5] авторами были построены математические модели распространения COVID-19 в различных группах населения. Авторами были построены модели, основанные на анализе временных рядов, модели, основанные на дифференциальных уравнениях, агентно-ориентированные модели, модели «игры среднего поля» с целью анализа и прогнозирования уровня распространения коронавирусной инфекции. Кроме того, для прогнозирования заболеваемости были реализованы искусственные нейронные сети, обученные на эмпирических данных, а именно на статистике заболеваемости COVID-19 и факторов, влияющих на нее. В результате на основе количества ПЦР-тестов и индекса самоизоляции населения, а также количества умерших и госпитализированных людей, возможно спрогнозировать уровень заболеваемости коронавирусной инфекцией.

Основная цель настоящей работы заключается в сборе множества публичных данных о количестве вакцинированных людей в разных странах, а также создание и обучение нейросетевой модели на этих данных. Конечным результатом должна быть нейросетевая система, способная прогнозировать количество вакцинированных людей в зависимости от уровня заболеваемости и смертности от COVID-19.

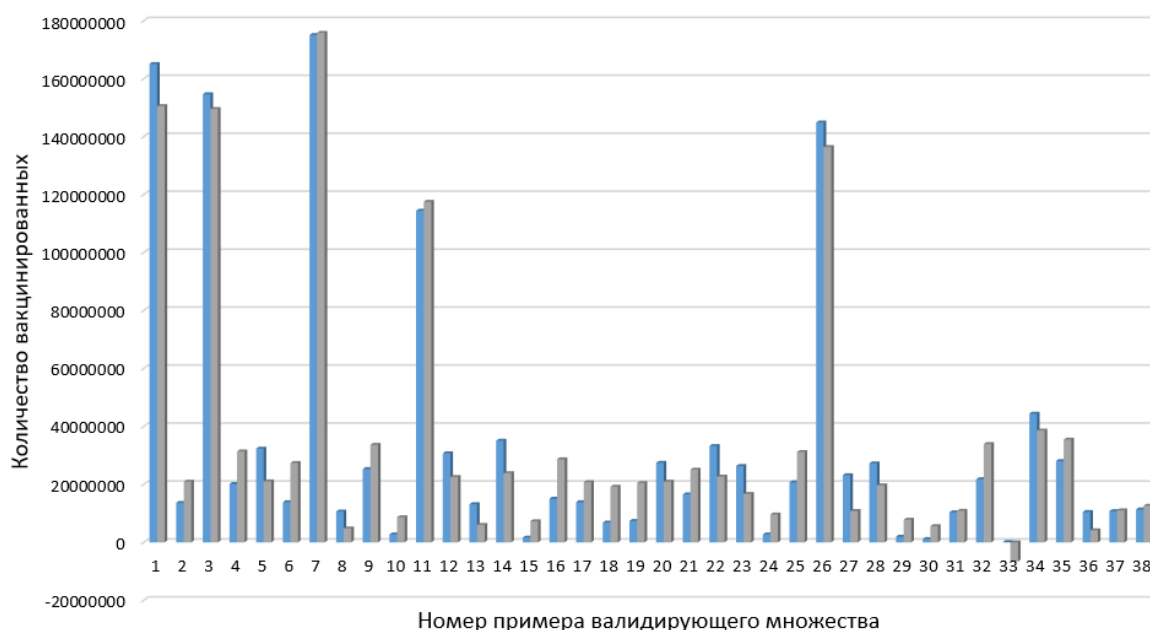
Для создания нейросетевой системы были выбраны следующие параметры:  $X_1$  – день (число),  $X_2$  – месяц,  $X_3$  – год,  $X_4$  – общее количество заболевших на эту дату,  $X_5$  – новое количество заболевших,  $X_6$  – общее количество умерших на эту дату,  $X_7$  – новое количество умерших. Выходной параметр – количество вакцинированных.

Обучающее множество было собрано с сайта kaggle [1], а также вручную. Перед переходом к проектированию и обучению нейросетевой модели, была выполнена очистка исходного множества от противоречивых примеров, выбросов [4, 6], дубликатов, пустых значений. Кроме того, был выделен пе-

риод с апреля по июнь 2021 года, данные за начало года были удалены с целью уменьшения погрешности, так как в начале массовой вакцинации происходили резкие скачки уровня вакцинированных. Данные взяты по пяти странам: Германия, Россия, Китай, США и Япония. Таким образом, объем итогового множества включает в себя 351 пример. Данное множество было разделено на обучающее, валидирующее и тестирующее в соотношении 70%, 20% и 10% соответственно.

Проектирование, обучение, тестирование нейронной сети выполнялись с помощью программы «Нейросимулятор 5» [7] по методике [8]. В результате оптимизации спроектированная нейронная сеть представляла собой персептрон, который имеет десять входных нейронов, один выходной и один скрытый слой с тремя нейронами.

Для оценки корректной работы спроектированной нейронной сети использовалось валидирующее множество, состоящее из 38 примеров. Средняя относительная ошибка тестирования составила 6.57%, что можно считать приемлемым результатом. На рисунке 1 представлена гистограмма, демонстрирующая разницу между фактическим и прогнозируемым нейросетью количеством вакцинированных.



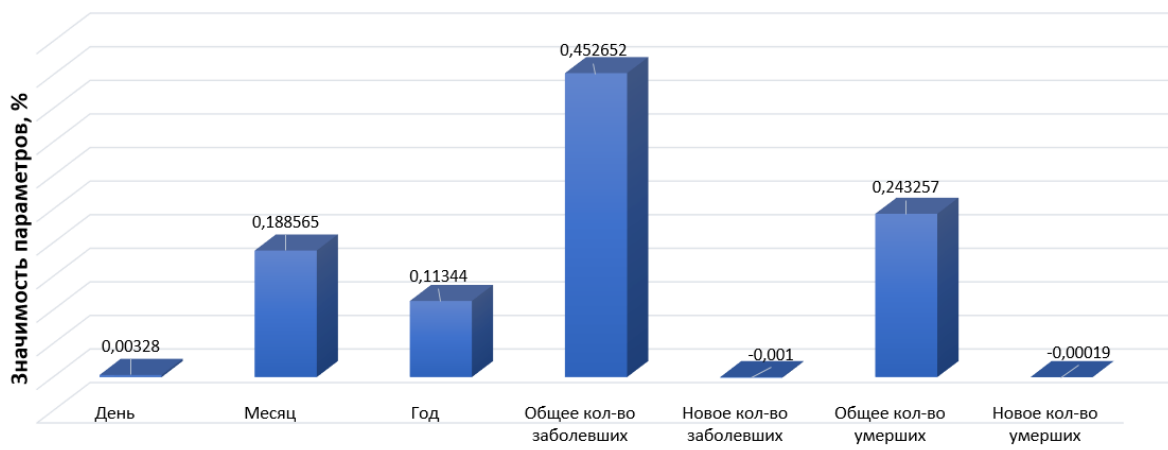
**Рисунок 1. Результат тестирования нейронной сети**

Из результатов, изображенных на рисунке 1, можно сделать вывод об адекватной работе спроектированной нейронной сети.

Оценка значимости входных параметров была выполнена с помощью программы «Нейросимулятор 5», результат отображен на рисунке 2.

Как видно из рисунка 2, наиболее значимыми являются общее количество заболевших, общее количество умерших, месяц и год, за который были взяты параметры. Как и ожидалось, наиболее влиятельным параметром является общее количество заболевших. На уровень вакцинации влияет уровень заболеваемости, так как в связи с увеличением заболеваемости страны вводят

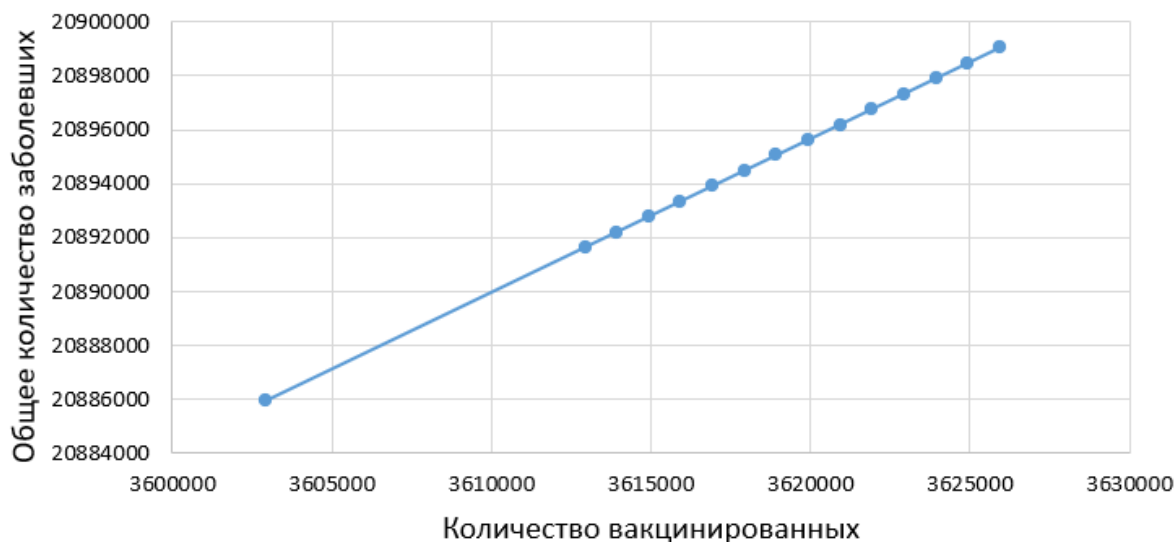
локдаун, режим QR-кодов, соответственно, люди не могут посещать общественные места без сертификата о вакцинации. Помимо этого, на вакцинацию влияет психологический фактор: много людей заболевают, многие умирают от коронавируса, а вакцина дает иммунитет, следовательно, больше людей будут вакцинироваться при высокой заболеваемости.



**Рисунок 2. Значимость входных параметров нейросетевой модели**

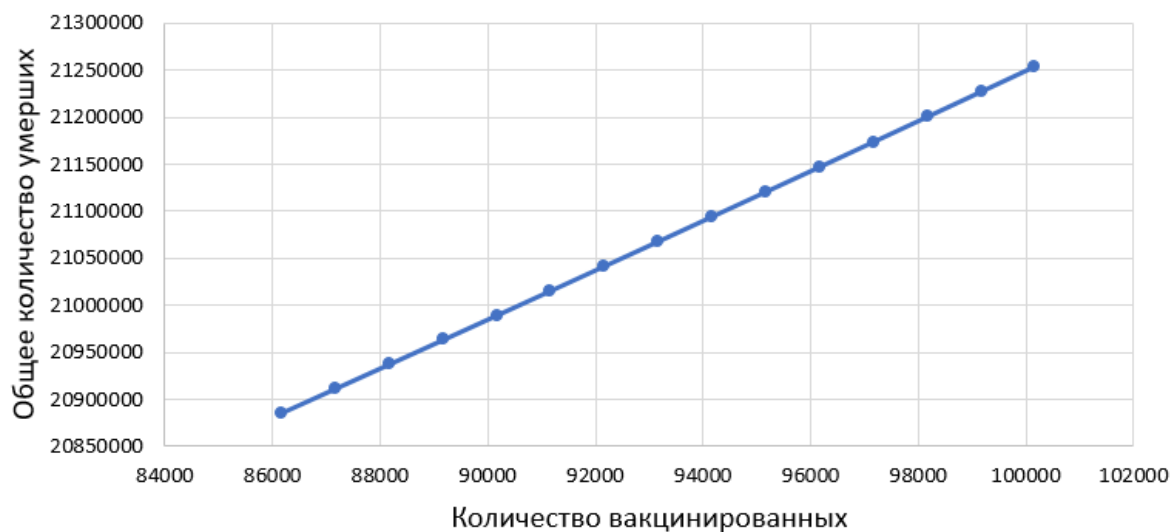
Далее было проведено исследование полученных зависимостей между входными параметрами и уровнем вакцинации. Исследование производилось с помощью метода «замораживания» [8-9], суть которого заключается в варьировании значения одного параметра и фиксации значений всех других параметров. Для этой цели были отобраны примеры, в которых возрастает один входной параметр, а остальные остаются одинаковыми.

На рисунке 3 показан график зависимости прогнозируемого количества вакцинированных человек от общего количества заболевших. В случае увеличения числа заболевших, нейросеть прогнозирует значительное увеличение количества вакцинаций.



**Рисунок 3. Зависимость прогнозируемого количества вакцинированных от общего количества заболевших**

На рисунке 4 продемонстрирована зависимость прогнозируемого количества вакцинированных человек от общего количества умерших. Можно заметить, что в случае увеличения числа умерших, нейросеть прогнозирует значительное увеличение количества вакцинаций.



**Рисунок 4. Зависимость прогнозируемого количества вакцинированных от общего количества умерших**

Полученные результаты исследований не противоречат реальности, что подтверждает, что спроектированную нейронную сеть можно считать пригодной для прогнозирования количества вакцинированных от COVID-19.

Аналогичная работа по проектированию, обучению, тестированию нейронной сети была выполнена на языке программирования Python с использованием библиотеки Keras [2]. Однако не удалось обучить сеть для прогнозирования уровня вакцинации.

Можно сделать вывод о том, программа «Нейросимулятор 5» позволяет реализовать более сложные нейронные сети, чем это возможно сделать на языке программирования Python. Нейросимулятор является более простым для изучения, более наглядным и интуитивно понятным. Отдельным достоинством нейросимулятора является то, что не нужно писать код, а также вручную стандартизировать данные.

**Заключение.** Построена система нейросетевого прогнозирования уровня вакцинации от COVID-19. Спроектированная нейросетевая модель учитывает 7 параметров: день, месяц, год, общее количество заболевших, новое количество заболевших, общее количество умерших и новое количество умерших от COVID-19. Методом сценарного прогнозирования построены графики зависимостей прогнозируемого уровня вакцинации от изменения входных параметров. Применение такого набора параметров в модели позволяет с высокой точностью прогнозировать количество вакцинированных от COVID-19 людей в зависимости от уровня заболеваемости и смертности. Данный набор параметров может быть изменен для прогнозирования уровня вакцинации от других болезней.

## Библиографический список

1. Kaggle.com – COVID-19 World Vaccination Progress. [Электронный ресурс]. Режим доступа: [https://www.kaggle.com/gpreda/covid-world-vaccination-progress?select=country\\_vaccinations.csv](https://www.kaggle.com/gpreda/covid-world-vaccination-progress?select=country_vaccinations.csv)
2. Ru-keras.com – библиотека глубокого обучение на Python. [Электронный ресурс]. Режим доступа: <https://ru-keras.com/home/>
3. Wiczorek M, Siłka J, Woźniak M. Neural network powered COVID-19 spread forecasting model. Chaos Solitons Fractals. 2020; 140: 110203. doi: 10.1016/j.chaos.2020.110203.
4. Yasnitsky L.N. Algorithm for searching and analyzing abnormal observations of statistical information based on the Arnold – Kolmogorov – Hecht-Nielsen theorem // International Journal of Advanced Trends in Computer Science and Engineering. 2020. 9(2). Pp. 1814-1819. DOI: 10.30534/ijatcse/2020/139922020 <http://www.warse.org/IJATCSE/static/pdf/file/ijatcse139922020.pdf>
5. Криворотько О. И., Кабанихин С. И. Математические Модели Распространения Covid-19. <https://arxiv.org/pdf/2112.05315>.
6. Черепанов Ф.М., Ясницкий Л.Н. Нейросетевой фильтр для исключения выбросов в статистической информации // Вестник Пермского университета. Математика. Механика. Информатика. – Пермь: Изд. Пермского университета, 2008. – Вып.4 (20). – С.151-155.
7. Черепанов Ф.М., Ясницкий Л.Н. Нейросимулятор 4.0. Свидетельство о регистрации программы для ЭВМ RUS 2014612546. Заявка № 2014610341 от 15.01.2014.
8. Ясницкий Л.Н. Введение в искусственный интеллект. М.: Издательский центр «Академия», 2005. – 176 с.
9. Ясницкий Л.Н., Думлер А.А., Полещук А.Н., Богданов К.В., Черепанов Ф.М. Нейросетевая система экспресс-диагностики сердечно-сосудистых заболеваний // Пермский медицинский журнал. – 2011. – №4. – С. 77-86.

## NEURAL NETWORK SYSTEM FOR PREDICTING COVID-19 VACCINATION RATES

*Subbotina Ekaterina V.*

National Research University "Higher School of Economics"  
Str. Studencheskaya, 38, Perm, Russia, 614070, infoperm@hse.ru

The article presents a description of the development of a neural network system for predicting the level of vaccination against COVID-19. The system makes it possible to predict the number of vaccinated people with great accuracy based on data on morbidity and mortality from COVID-19. With the help of the

developed intellectual system, a study of the subject area was carried out, patterns of practical importance were identified.

**Keywords:** artificial intelligence, neural network technologies, forecasting, vaccination, COVID-19, coronavirus infection.

УДК 004.032.26

## НЕЙРОСЕТЕВАЯ СИСТЕМА ПРОГНОЗИРОВАНИЯ ДИАБЕТА

*Копнин Матвей Вадимович*

Национальный исследовательский университет «Высшая школа экономики»,  
614060, Россия, г. Пермь, ул. б-р Гагарина, 37А, [copnin.matwei@yandex.ru](mailto:copnin.matwei@yandex.ru)

В статье представлено описание разработки нейросетевой системы для прогнозирования диабета у пациентов. Система позволяет с большой точностью предсказать наличие диабета у пациента, анализируя показатели представленных параметров. С помощью разработанной интеллектуальной системы проведено исследование предметной области, выявлены закономерности, имеющие практическое значение.

**Ключевые слова:** искусственный интеллект, нейросетевые технологии, прогнозирование, диабет.

**Введение.** Сахарный диабет — это заболевание, характеризующееся повышением уровня сахара в крови, приводящее к поражению почек и нервной системы. Более того, заболевание нарушает зрение и влияет на сосудистую систему. Диабет бывает разных типов, может поражать как молодых людей, так и развиваться с возрастом [1].

Главными причинами возникновения диабета у человека является генетическая предрасположенность (известно, что если родители или близкие родственники болели сахарным диабетом, то риск его развития повышается на 5-10%), избыточная масса тела и ожирение и повышенный уровень глюкозы в крови.

Основная цель работы заключается в сборе данных о пациентах, которые имеют или могут иметь данное заболевание, а также создание и обучение нейросетевой модели на этих данных. Конечным результатом должна быть нейросетевая система, способная прогнозировать наличие диабета у пациента.

Для создания нейросетевой системы были выбраны данные из источника [2]. Наборы данных состоят из нескольких медицинских предиктивных (независимых) переменных и одной целевой (зависимой) переменной — результата. Независимые переменные включают количество беременностей, которые были у пациента, концентрацию глюкозы, артериальное давление,

толщину кожи, ИМТ (индекс массы тела), уровень инсулина, возраст и наследственную предрасположенность к диабету. Важно отметить, что данная система прогнозирования будет использована только для женщин.

При разработке нейросетевой модели были использованы следующие входные параметры:

- X1 – Количество беременностей.
- X2 – Глюкоза (концентрация глюкозы в плазме крови)
- X3 – Диастолическое артериальное давление (мм рт. ст.).
- X4 – Толщина кожной складки трицепса (мм).
- X5 – Инсулин (2-часовой сывороточный инсулин (мю Ед/мл)).
- X6 – Индекс массы тела (вес в кг/(рост в м)<sup>2</sup>).
- X7 – Наследственная предрасположенность к диабету.
- X8 – Возраст (годы).

Выходным параметром модели является параметр D1 – наличие диабета (0 – отсутствует, 1 – диабет).

Проектирование, обучение, тестирование нейронной сети выполнялись с помощью программы «Нейросимулятор 5» по методике [3]. В результате оптимизации спроектированная нейронная сеть представляла собой перцептрон, который имеет восемь входных нейронов, один выходной и один скрытый слой с тремя нейронами.

Изначально объём множества включал в себя 769 примеров, однако, при обучении средняя относительная погрешность составляла 30%, что не является приемлемым результатом. В связи с этим по методике [6] был произведен поиск и удаление статистических выбросов в исходном множестве. Всего было обнаружено и удалено 240 примеров.

На рисунке 1 представлены выбросы, которые были удалены. Для этого значения столбца E1(разница между D1 и Y1) была отсортирована и все крупные числа были удалены.

X1	X2	X3	X4	X5	X6	X7	X8	D1	Y1	E1
9	102	76	37	0	32,9	0,665	46	1	0,2683	0,7317
7	103	66	32	0	39,1	0,344	31	1	0,2812	0,7188
10	90	85	32	0	34,9	0,825	56	1	0,2856	0,7144
8	105	100	36	0	43,3	0,239	45	1	0,2871	0,7129
6	166	74	0	0	26,6	0,304	66	0	0,7118	0,7118
5	112	66	0	0	37,8	0,261	41	1	0,2922	0,7078
4	125	80	0	0	32,3	0,536	27	1	0,3037	0,6963
1	144	82	40	0	41,3	0,607	28	0	0,6855	0,6855
10	108	66	0	0	32,4	0,272	42	1	0,3168	0,6832
0	118	84	47	230	45,8	0,551	31	1	0,3176	0,6824
7	114	64	0	0	27,4	0,732	34	1	0,3186	0,6814
2	127	58	24	275	27,7	1,6	25	0	0,6805	0,6805
4	109	64	44	99	34,8	0,905	26	1	0,3205	0,6795
5	139	80	35	160	31,6	0,361	25	1	0,3208	0,6792
7	142	60	33	190	28,8	0,687	61	0	0,6781	0,6781
2	155	74	17	96	26,6	0,433	27	1	0,3243	0,6757
1	167	74	17	144	23,4	0,447	33	1	0,325	0,675
9	154	78	30	100	30,9	0,164	45	0	0,6741	0,6741
11	127	106	0	0	39	0,19	51	0	0,6718	0,6718

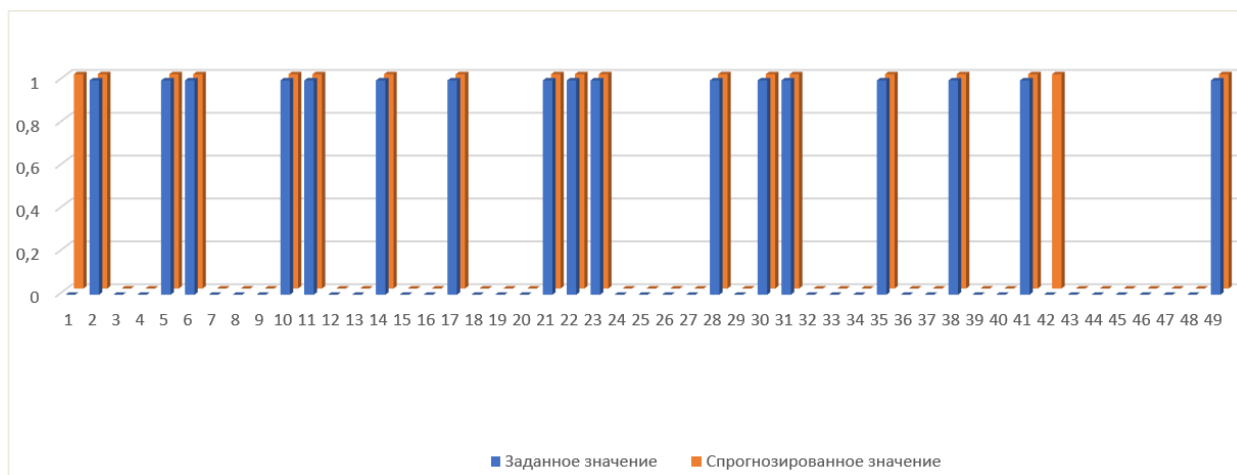
**Рисунок 1. Выбросы**



Объём итогового множества включает в себя 529 примеров, 49 из которых составляют тестовое множество, а 480 – обучающее множество.

Средняя относительная погрешность составила 4.08%

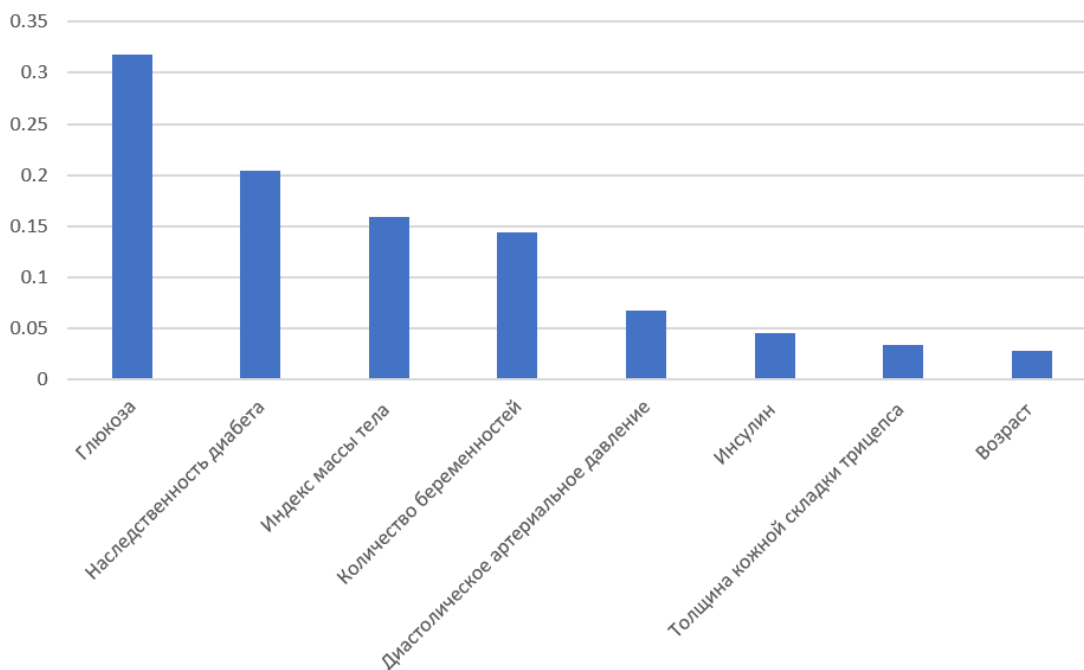
На рисунке 1 представлена гистограмма, демонстрирующая разницу между фактическим и прогнозируемым нейросетью значением наличия диабета для примеров из тестирующего множества.



**Рисунок 2. Результат тестирования нейронной сети**

Из результатов, изображенных на рисунке 2, можно сделать вывод об адекватной работе спроектированной нейронной сети.

Оценка значимости входных параметров была выполнена с помощью программы «Нейросимулятор 5» [5], результат отображен на рисунке 2.



**Рисунок 3. Значимость параметров**

Исходя из рисунка 3, важнейшим параметром является концентрация глюкозы, затем наследственная предрасположенность к диабету и индекс

массы тела. Это означает, что упомянутые в начале причины появления диабета подтвердились. Главные показатели наличия диабета у пациента: повышенная глюкоза в крови, генетика, ожирение.

Далее было проведено исследование полученных зависимостей между входными параметрами и диагнозом. Исследование производилось с помощью метода «замораживания» [3], который заключается в варьировании значения одного параметра и фиксации значений всех других параметров.

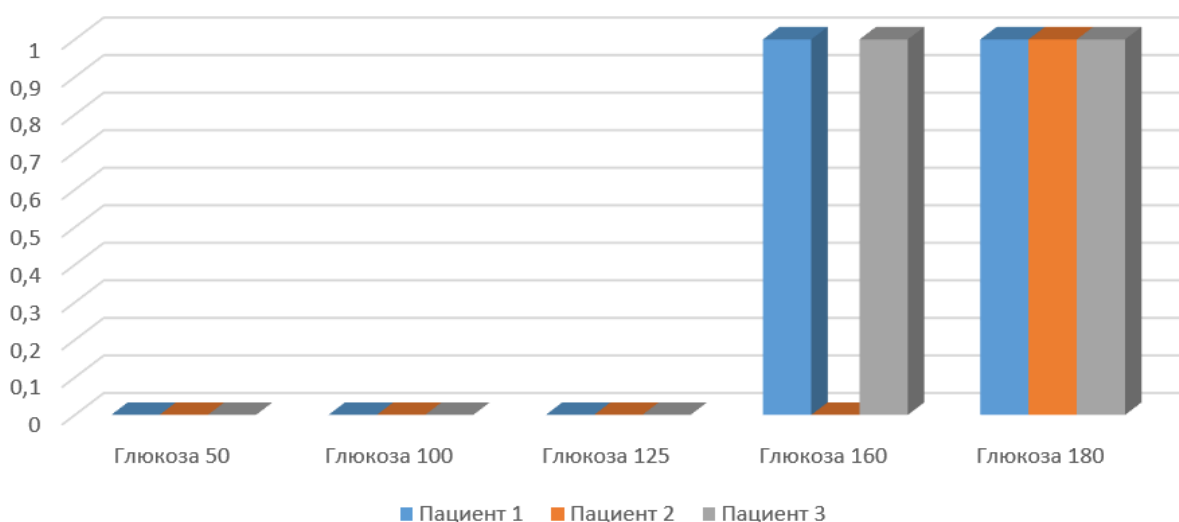
На таблице ниже представлены показатели и диагноз трех пациентов. С помощью метода «замораживания» будут проверены показатели, от которых зависит диагноз пациента.

**Таблица 1**

**Изначальные показатели пациентов**

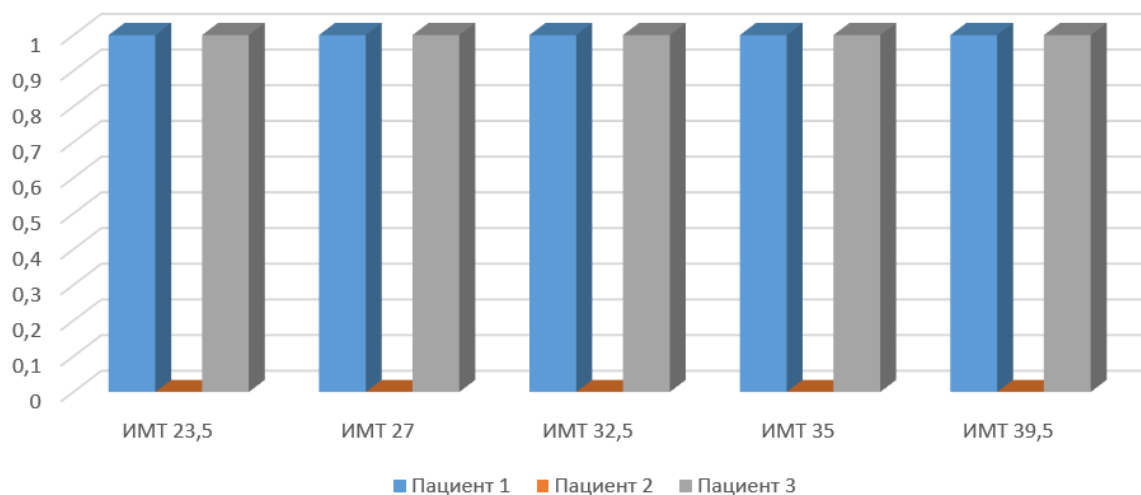
	X1	X2	X3	X4	X5	X6	X7	X8	D1
Пациент 1	0	179	50	36	159	37,8	0,455	22	1
Пациент 2	1	111	94	0	0	32,8	0,265	45	0
Пациент 3	6	195	70	0	0	30,9	0,328	31	1

На рисунке 4 показан график зависимости диабета от количества глюкозы в крови. Можно заметить, что изначально высокие показатели у первого и третьего пациента влияют на их положительный результат, но если глюкозы в крови было бы 125 и меньше, то вероятность диабета сводится к 0. При показателе 180 глюкозы у каждого пациента обнаружен диабет.



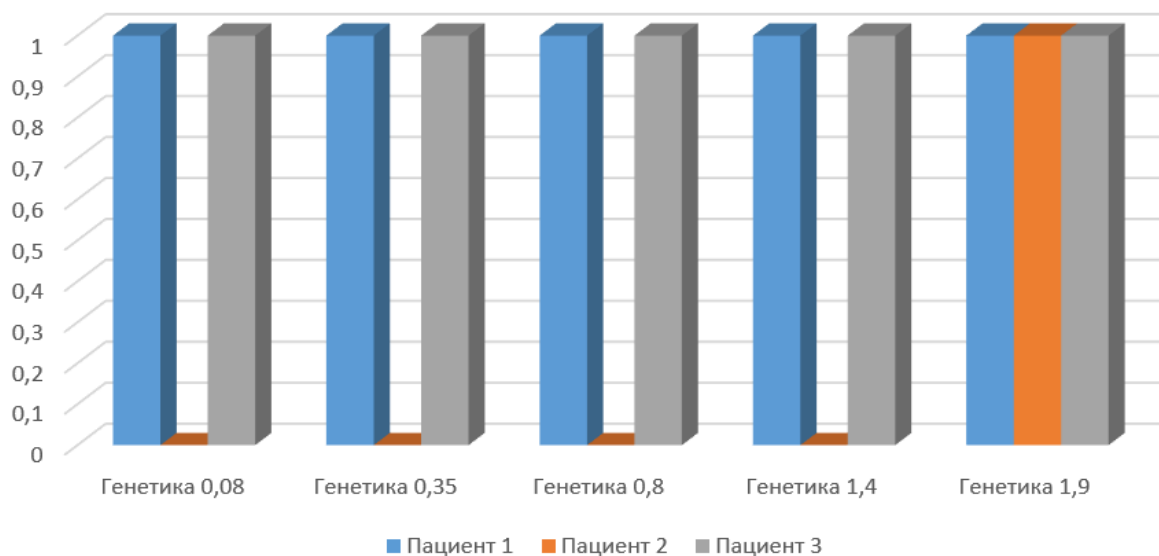
**Рисунок 4. Зависимость результата от глюкозы**

На рисунке 5 изображен график зависимости диабета от индекса массы тела. Эти данные свидетельствуют о том, что несмотря ни на какое соотношение роста и веса у каждого из пациентов, их диагноз не изменится. Значит, ИМТ не играет никакой роли в диагнозе этих пациентов, однако, является важным показателем.



**Рисунок 5. Зависимость результата от ИМТ**

На рисунке 6 показан график зависимости диабета от наследственной предрасположенности. Исходя из этих показателей, у пациентов 1 и 3 генетика не влияет на диагноз, но в корне меняет жизнь второго пациента. При наибольшей вероятности наследственной предрасположенности, его диагноз становится положительным.



**Рисунок 6. Зависимость результата от генетики**

Исходя из результатов исследования, можно сделать вывод, что важнейшей причиной возникновения диабета у пациентов является глюкоза. Низкие показатели ИМТ и вероятности наследственной предрасположенности никак не повлияли на диагноз 1 и 3 пациентов, а при высоких показателях глюкозы и вероятности наследственной предрасположенности второй пациент обречен на возникновение диабета.

**Заключение.** Построена система нейросетевого прогнозирования наличия диабета у пациентов. Нейросетевая модель учитывает 8 параметров:

количество беременностей, которые были у пациента, концентрацию глюкозы, артериальное давление, толщину кожи, ИМТ (индекс массы тела), уровень инсулина, возраст и наследственную предрасположенность к диабету. Методом сценарного прогнозирования построены графики зависимостей прогнозируемого положительного результата на диабет от изменения входных параметров. Применение такого набора параметров в модели позволяет с высокой точностью прогнозировать наличие диабета. Данный набор параметров может быть изменен для прогнозирования наличия диабета среди мужчин.

### **Библиографический список**

1. Сахарный диабет: признаки, симптомы, лечение, питание (диета при диабете). [Электронный ресурс]. Режим доступа: <https://www.mgb1-74.ru/novosti/154-saharnyj-diabet-priznaki-simptomu-lechenie-pitanie-dieta-pri-diabete.html>
2. Hüseyin Özdemir «Diabetes Prediction with Artificial Neural Network» [Электронный ресурс]. Режим доступа: <https://www.kaggle.com/ozdemirh/diabetes-prediction-with-artificial-neural-network/data>.
3. Ясницкий Л.Н. Интеллектуальные системы: учебник. М.: Лаборатория знаний, 2016. 221 с.
4. ВОЗ [Электронный ресурс]. Режим доступа: <https://www.who.int/ru/news-room/fact-sheets/detail/diabetes>
5. Черепанов Ф.М., Ясницкий Л.Н. Нейросимулятор 4.0. Свидетельство о регистрации программы для ЭВМ RUS 2014612546. Заявка № 2014610341 от 15.01.2014.
6. Yasnitsky L.N. Algorithm for searching and analyzing abnormal observations of statistical information based on the Arnold – Kolmogorov – Hecht-Nielsen theorem. International Journal of Advanced Trends in Computer Science and Engineering. 2020. 9(2). Pp. 1814-1819.

## **NEURAL NETWORK SYSTEM FOR PREDICTING DIABETES**

*Kopnin Matvej V.*

National Research University "Higher School of Economics"  
Str. Gagarina Boulevard, 37A, Perm, Russia, 614060, kopnin.matwei@yandex.ru

The article describes the development of neural network system for predicting diabetes in patients. The system allows to accurately predict the presence of diabetes in a patient by analyzing the indicators of the presented parameters. With the help of the developed intelligent system, a study of the subject area has been carried out and the regularities of practical importance have been revealed.

**Keywords:** artificial intelligence, neural network technology, prognosis, diabetes.

## НЕЙРОСЕТЕВАЯ СИСТЕМА ПРОГНОЗИРОВАНИЯ ОБНАРУЖЕНИЯ САХАРНОГО ДИАБЕТА НА ОСНОВЕ ДАННЫХ АНАМНЕЗА

*Лавров Кирилл Михайлович*  
*Никитина Анна Леонидовна*

Пермский государственный национальный исследовательский университет,  
ПМИ. 614990, Россия, г. Пермь, ул. Букирева, 15.  
tgidk1812@gmail.com, fullofevilclowns@mail.ru

В статье представлено описание разработки нейросетевой системы для прогнозирования результатов диагностики сахарного диабета на основе анамнеза пациента. Система позволяет с большой точностью предсказать вероятность выявления диабета при дальнейших лабораторных исследованиях. С помощью разработанной интеллектуальной системы проведено исследование предметной области и выявлены закономерности, имеющие практической значение.

**Ключевые слова:** искусственный интеллект, нейросетевые технологии, прогнозирование, сахарный диабет, анамнез.

**Введение.** Достаточно легко прогнозировать наличие сахарного диабета на основе результатов анализов человека, но не у каждого есть возможность и/или желание периодически посещать больницу. Сложнее предсказывать наличие диабета на основе субъективных показаний (ответов на вопросы), но это самый доступный вариант для большинства людей. Поэтому было принято решение работать с анамнезом предполагаемых больных, которых после опроса врачи полноценно обследовали на наличие сахарного диабета. Далее результаты обследования сопоставлялись с результатами опроса. Таким образом, любой человек может ответить на простые вопросы о своем состоянии и узнать насколько ему необходимо обратиться в больницу для профессионального обследования.

При анализе литературных источников выяснилось, что работ на тему прогнозирования вероятности сахарного диабета на основе анамнеза было проведено достаточно мало. В различных материалах интернета (картинки, популярные и научные статьи про симптомы, на которые стоит обратить внимание) встречаются в основном те же вопросы для сбора анамнеза, что и в данном проекте (можно заметить небольшие различия в критериях в разных источниках, но основная часть, представленная в проекте, встречается в абсолютном большинстве материалов).

Основная цель настоящей работы заключается в сборе множества публичных данных об анамнезе людей и результатах наличия (отсутствия) у них

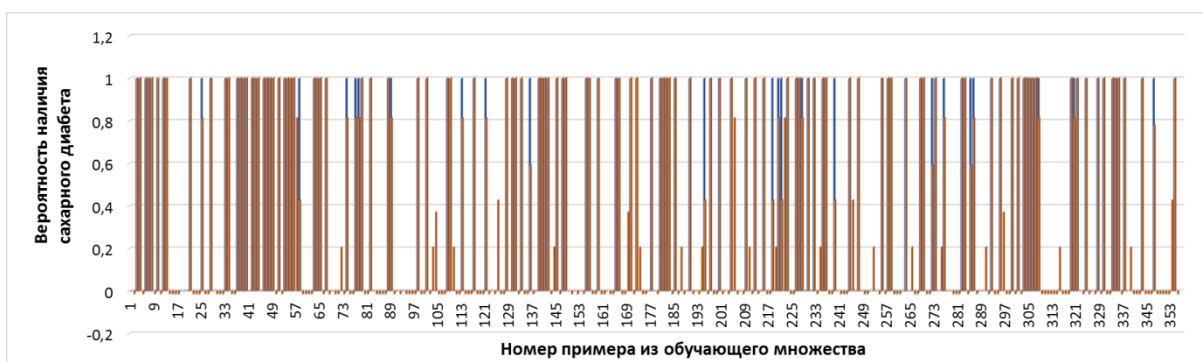
сахарного диабета на основе полноценного врачебного исследования, а также создание и обучение нейросетевой модели на этих данных. Конечным результатом должна быть нейросетевая система, способная корректно прогнозировать вероятность наличия у человека сахарного диабета в больше чем 80% случаев.

Для создания нейросетевой системы были выбраны следующие параметры:

- X1 – возраст,
- X2 – пол,
- X3 – увеличенное мочеобразование,
- X4 – чрезмерная жажда,
- X5 – внезапная потеря веса,
- X6 – слабость,
- X7 – повышенный аппетит,
- X8 – генитальная молочница,
- X9 – ухудшение зрения,
- X10 – зуд,
- X11 – раздражительность,
- X12 – замедленное заживление ран,
- X13 – слабость в мышцах,
- X14 – неподвижность, онемение мышц,
- X15 – выпадение волос,
- X16 – ожирение.

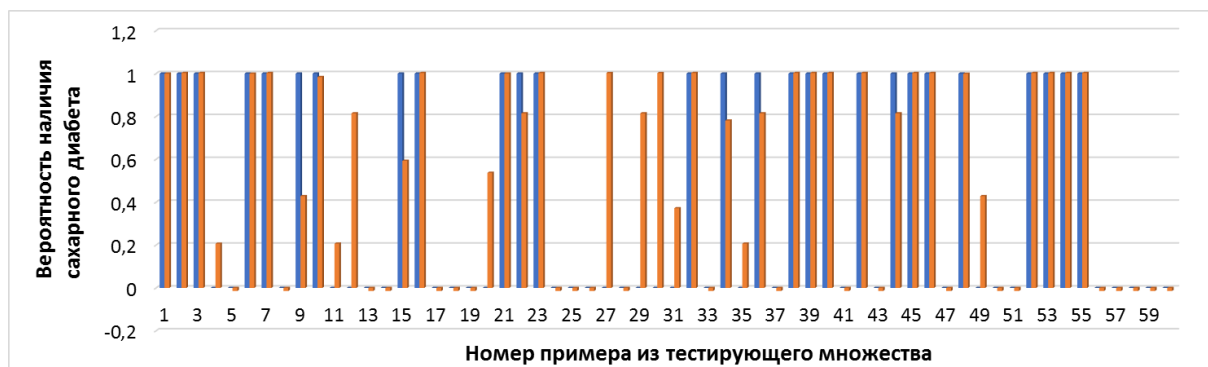
Выходной параметр – сведения о наличии/отсутствии сахарного диабета.

Обучающее множество было собрано с помощью интернет-ресурса kaggle [1]. Изначально в множестве было 520 примеров. Перед переходом к проектированию и обучению нейросетевой модели была необходима очистка исходного множества от противоречивых примеров и выбросов. Обучив нейронную сеть на основе обучающего и валидирующего множеств, была проведена проверка обучающего множества (вместо тестового) для поиска выбросов в обучающем множестве. Результат этих действий представлен на рисунке 1.



**Рисунок 1. Поиск выбросов**

Обучив систему на основе исходного множества, был получена крайне неточная система с высоким процентом ошибок, гистограмма со сравнением полученного результата с правильным ответом на основе тестового множества представлена на рисунке 2. Среднеквадратичная относительная ошибка составила 28,4%. Самыми некорректными являются результаты тестов на примерах 9, 12, 27, 29 и 30.



**Рисунок 2. Результат обучения на исходном множестве**

Было принято решение очистить множество от некорректных примеров: удалялись примеры с самыми некорректными результатами тестирования, после чего система заново обучалась и проводилась оценка результатов тестирования на общем множестве до получения удовлетворительных погрешностей по методике Ясницкого Л.Н.[6].

Таким образом, объем итогового очищенного множества включает в себя 491 пример. Данное множество было разделено на обучающее, валидирующее и тестирующее в соотношении 17%, 71% и 12% соответственно.

Проектирование, обучение, тестирование нейронной сети выполнялись с помощью программы «Нейросимулятор 5» [2] по методике [3]. В результате оптимизации спроектированная нейронная сеть представляла собой персептрон, имеющий шестнадцать входных нейронов, один выходной и один скрытый слой с четырьмя нейронами.



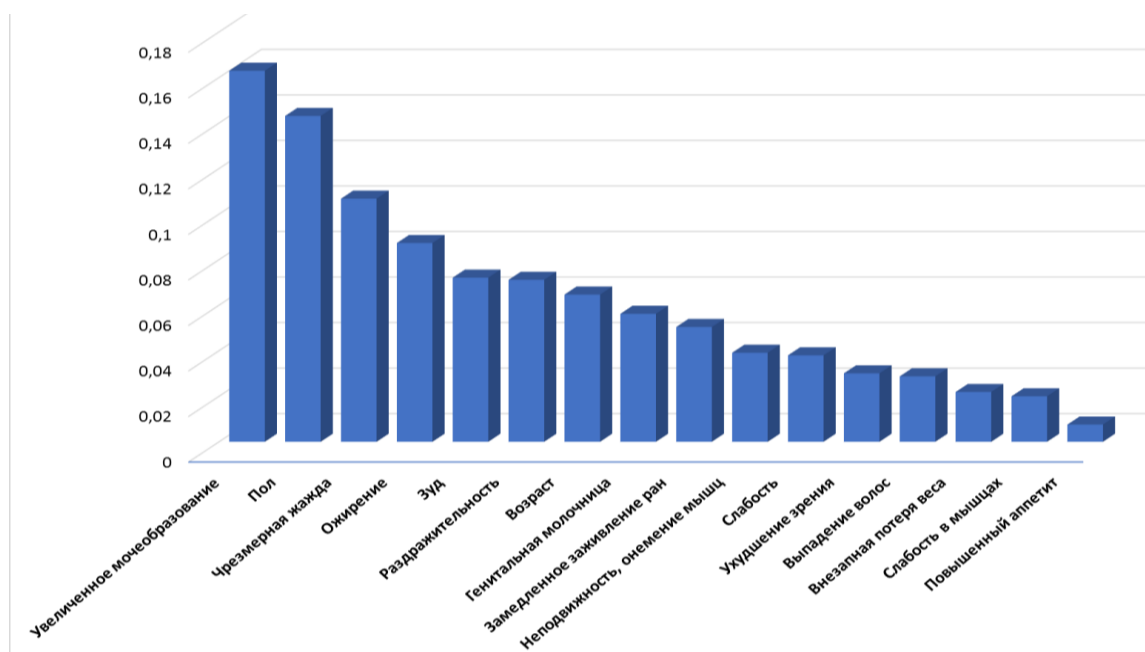
**Рисунок 3. Результат обучения на очищенном от выбросов множестве**

Для оценки корректной работы спроектированной нейронной сети использовалось тестирующее множество, состоящее из 57 примеров. Средняя относительная ошибка тестирования составила 1.2%, что можно считать при-

емлемым результатом. На рисунке 3 представлена гистограмма, демонстрирующая разницу между фактическим и прогнозируемым нейросетью результатом исследованию по выявлению сахарного диабета у пациента.

Из результатов, изображенных на рисунке 3, можно сделать вывод о корректной работе спроектированной нейронной сети.

Оценка значимости входных параметров была выполнена с помощью программы «Нейросимулятор 5», результат отображен на рисунке 4.



**Рисунок 4. Значимость входных параметров нейросетевой модели**

Как видно из рисунка 4, наиболее значимыми являются увеличенное мочеобразование, пол и чрезмерная жажда. Неожиданно, что наиболее влиятельным параметром оказалось именно увеличенное мочеобразование. Ожидалось, что самым влиятельным параметром будет ожирение, так как рацион с повышенной калорийностью и вероятное присутствие большого количества сладкой и жирной еды провоцирует не только ожирение, но и приводит к нарушениям в организме, ведущим за собой сахарный диабет (чаще 2 типа).

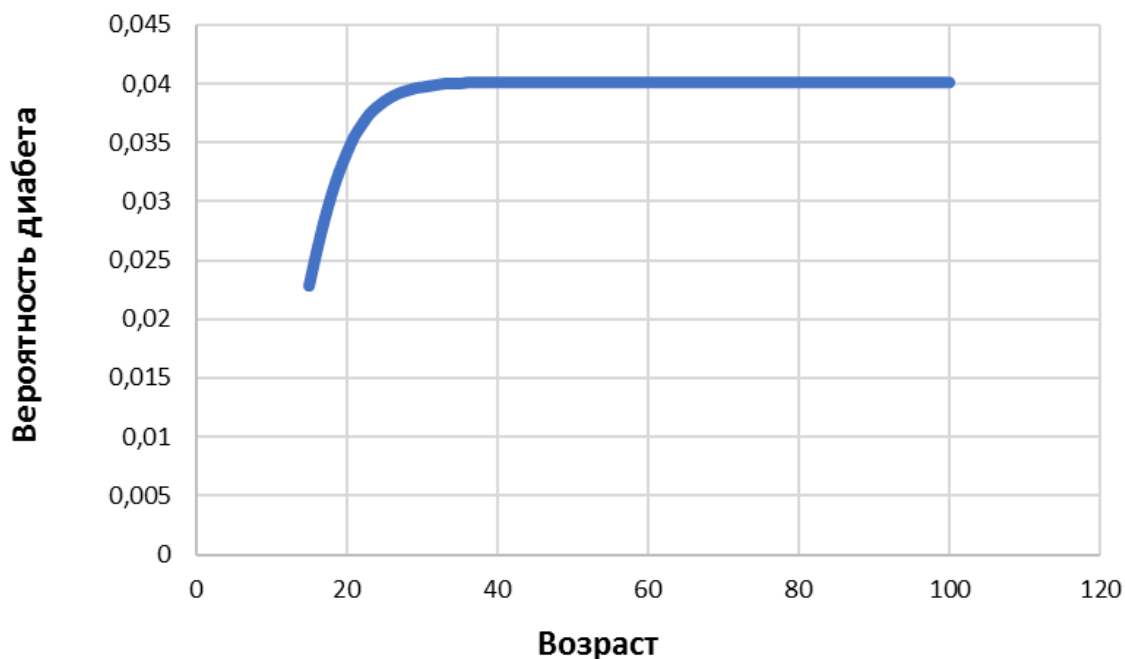
Далее было проведено исследование полученных зависимостей между входными параметрами результатом медицинского обследования. Исследование производилось с помощью метода «замораживания», суть которого заключается в варьировании значения одного параметра и фиксации значений всех других параметров. Для этой цели был отобран «нейтральный пример», для которого нейросеть давала ответ, средне отличный от 0 и 1.

На рисунке 5 показан график зависимости прогнозируемого результата от возраста. С увеличением возраста вероятность наличия диабета при прочих равных увеличивается.

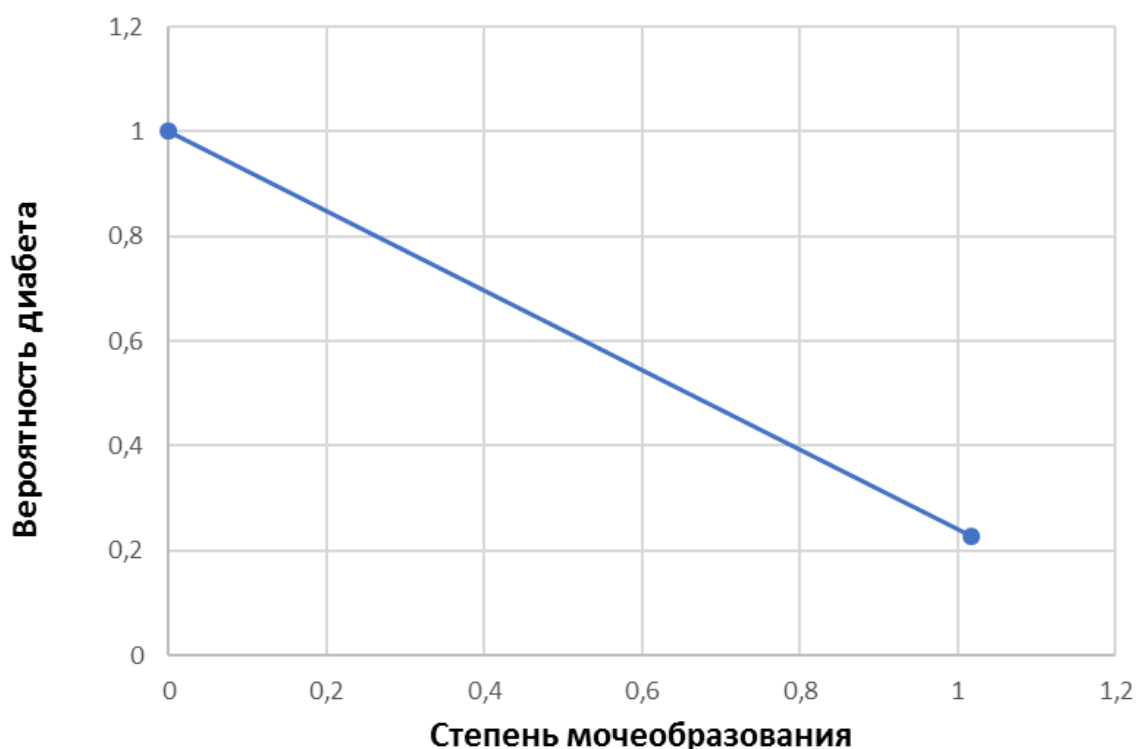
На рисунке 6 продемонстрирована зависимость прогнозируемого результата от отсутствия или наличия увеличенного мочеобразования. Можно заметить, что в данном случае зависимость обратная.



На рисунке 7 изображен график зависимости результата от наличия чрезмерной жажды. Как видно из графика, результат обратно пропорционально зависит от этого параметра.



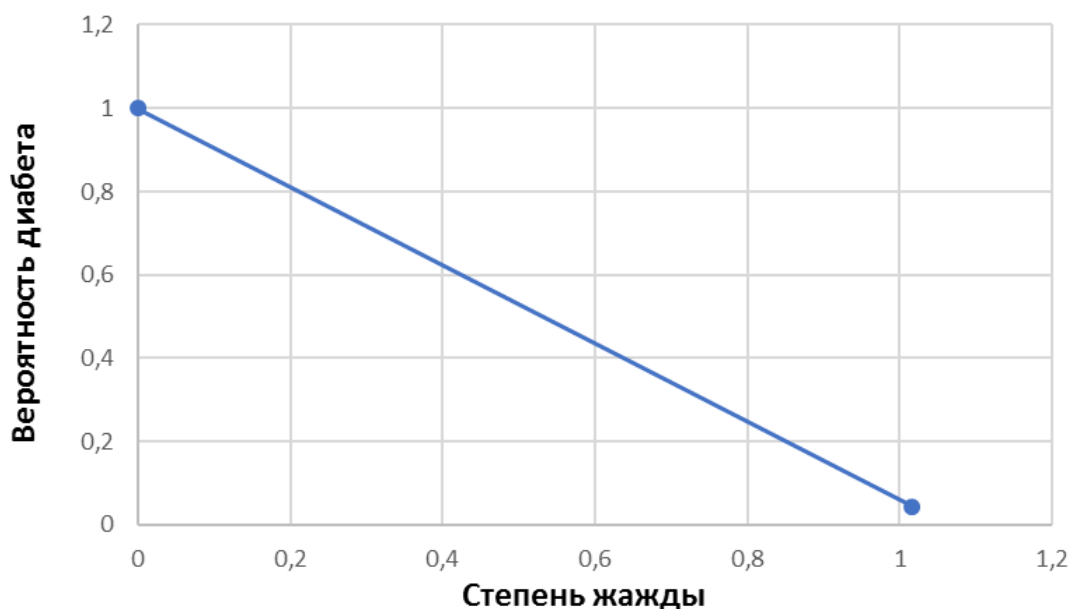
**Рисунок 5. Зависимость прогнозируемого результата от возраста**



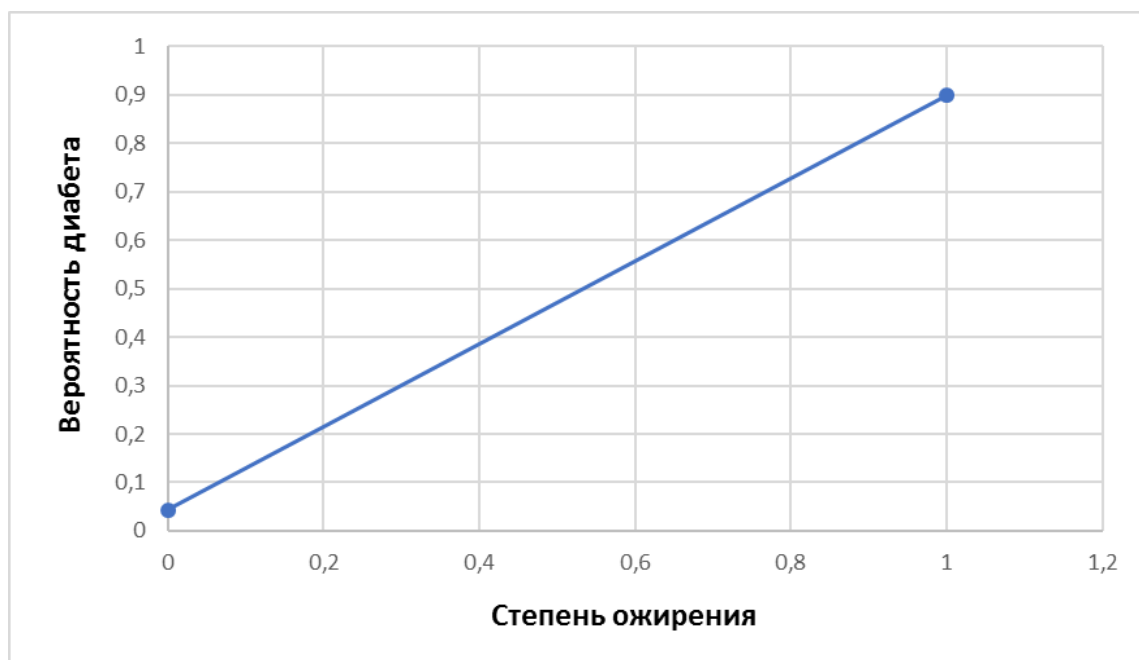
**Рисунок 6. Зависимость прогнозируемого результата от увеличенного мочеобразования**

На рисунке 8 продемонстрирована зависимость прогнозируемого результата от ожирения. Как видно из графика, ожирение значительно по-

вышает риск заболевания сахарным диабетом. Это совпадает с первоначальным ожиданием: ожирение как самого влиятельный параметр, с точки зрения медицины это весьма обоснованно: слишком калорийный рацион, обилие сладкого и жирного приводят к серьезным сбоям в организме, в том числе к сбою равномерного выделения инсулина, что является серьезной причиной развития диабета.

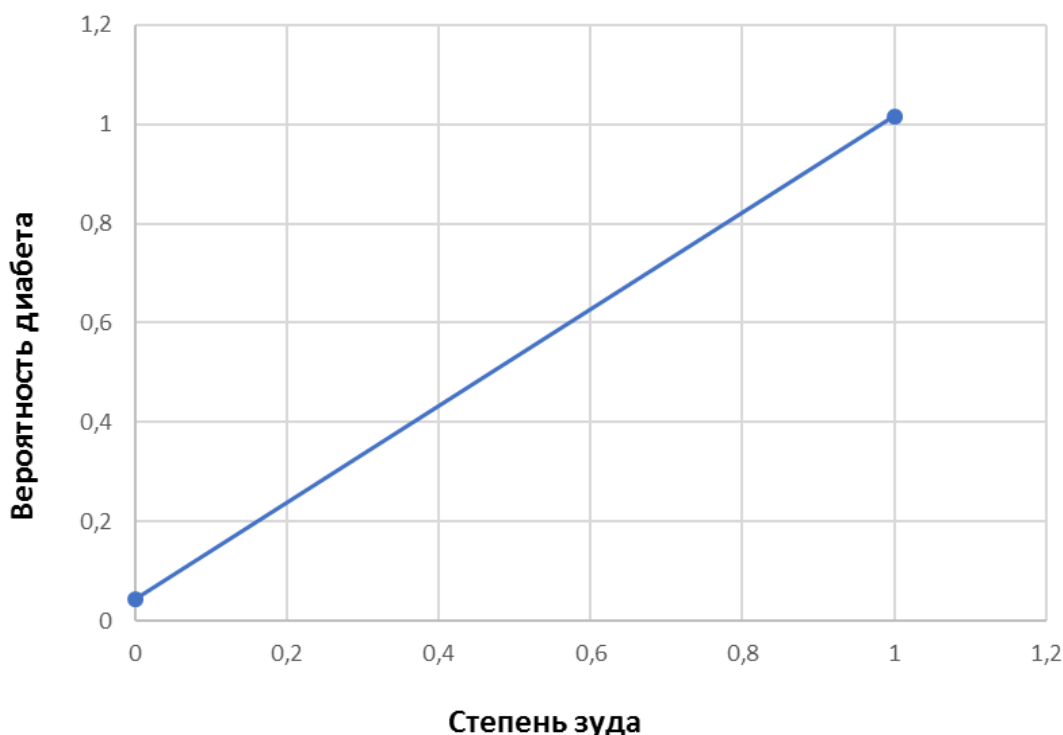


**Рисунок 7. Зависимость прогнозируемого результата от жажды**



**Рисунок 8. Зависимость прогнозируемого результата от ожирения**

На рисунке 9 продемонстрирована зависимость прогнозируемого результата от наличия зуда. Зуд наблюдается примерно у четверти больных сахарным диабетом пациентов, а причиной зуда являются повреждения нервных волокон из-за специфики заболевания.[5]



**Рисунок 9. Зависимость прогнозируемого результата от зуда**

Полученные результаты исследований не противоречат реальности, что подтверждает: спроектированную нейронную сеть можно считать пригодной для прогнозирования результатов выявления медицинским обследованием сахарного диабета на основе анамнеза пациента.

**Заключение.** Построена система нейросетевого прогнозирования результата обследования пациента на наличие сахарного диабета на основе анамнеза. Спроектированная нейросетевая модель учитывает 16 параметров: возраст, пол, увеличенное мочеобразование, чрезмерная жажда, внезапная потеря веса, слабость, повышенный аппетит, генитальная молочница, ухудшение зрения, зуд, раздражительность, замедленное заживление ран, слабость в мышцах, неподвижность, онемение мышц, выпадение волос, ожирение. Методом сценарного прогнозирования построены графики зависимостей прогнозируемой вероятности обнаружения сахарного диабета у пациента от изменения входных параметров. Применение такого набора параметров в модели позволяет с высокой точностью прогнозировать вероятность наличия сахарного диабета.

### **Библиографический список**

1. kaggle.com – базы данных. [Электронный ресурс]. Режим доступа: <https://www.kaggle.com>
2. Черепанов Ф.М., Ясницкий Л.Н. Нейросимулятор 4.0. Свидетельство о регистрации программы для ЭВМ RUS 2014612546. Заявка № 2014610341 от 15.01.2014.
3. Ясницкий Л.Н. Введение в искусственный интеллект. М.: Издательский центр «Академия», 2005. – 176 с.

4. Л.Н. Ясницкий, А.А. Думлер, Ф.М. Черепанов Новые возможности применения методов искусственного интеллекта для моделирования появления и развития заболеваний и оптимизации их профилактики и лечения.

5. Полинейропатия диабетическая – симптомы и лечение Коптенко Н. В. 18 июля 2020 <https://probolezny.ru/polinyayropatiya-diabeticheskaya/>

6. Yasnitsky L.N. Algorithm for searching and analyzing abnormal observations of statistical information based on the Arnold – Kolmogorov – Hecht-Nielsen theorem // International Journal of Advanced Trends in Computer Science and Engineering. 2020. 9(2). Pp. 1814-1819. DOI: 10.30534/ijatcse/2020/139922020 <http://www.warse.org/IJATCSE/static/pdf/file/ijatcse139922020.pdf>

## NEURAL NETWORK SYSTEM FOR PREDICTION OF DETECTION OF DIABETES MELLITUS BASED ON ANAMNESIS

*Lavrov Kirill M., Nikitina Anna L.*

Perm State University

Str. Bukireva, 15, Perm, Russia, 614990,

tgidk1812@gmail.com, fullofevilclowns@mail.ru

The article presents a description of the development of a neural network system for predicting the results of diagnosing diabetes based on anamnesis. The system makes it possible to predict the probability of detecting diabetes in further laboratory tests with great accuracy. With the help of the developed intellectual system, a study of the subject area was carried out, and the results of the consideration of practical significance were identified.

**Keywords:** artificial intelligence, neural network technologies, forecasting, diabetes mellitus, anamnesis.

УДК 004.032.26

## НЕЙРОСЕТЕВАЯ СИСТЕМА ПРОГНОЗИРОВАНИЯ ЗЛОКАЧЕСТВЕННОЙ ОПУХОЛИ

*Мокеева Анастасия Сергеевна*

Национальный исследовательский университет «Высшая школа экономики»,

614990, Россия, г. Пермь, ул. Бульвар Гагарина, 37

mokeeva\_\_as@mail.ru

В статье представлено описание разработки нейросетевой системы для прогнозирования обнаружения злокачественной опухоли в организме человека. Система позволяет с большой точностью предсказать на основе

различных характеристик, полученных при обследовании, является ли опухоль злокачественной. С помощью разработанной интеллектуальной системы проведено исследование предметной области, выявлены закономерности, имеющие практическое значение.

**Ключевые слова:** искусственный интеллект, нейросетевые технологии, прогнозирование, рак.

**Введение.** Нейросети изменили многие отрасли и области науки. Их возможности можно применить для решения проблем обнаружения злокачественной опухоли, когда анализ огромного количества данных выходит за рамки возможностей традиционных статистических анализов и инструментов. Злокачественная опухоль – это патологическое новообразование, возникающее при нарушении механизмов деления и роста клетки. Злокачественные опухоли возникают в результате злокачественного перерождения (малигнизации) нормальных клеток, которые начинают бесконтрольно размножаться. В настоящее время непосредственная причина появления злокачественных опухолей неизвестна. Эта опухоль характеризуется достаточно быстрым ростом, способностью прорасти в другие органы и ткани, а также давать метастазы по всему организму. Именно поэтому особенно важно обнаружить ее на ранних этапах развития.

Применение нейросетевых технологий имеет большой потенциал в медицине. На тему обнаружения рака с помощью нейронных сетей написано множество статей [1, 2]. MedicalXpress приводит в статье положительный опыт использования нейросети для обнаружения рака на ранних стадиях. Нейросети были обучены обрабатывать рентгеновские снимки и компьютерную томографию. На основе полученных данных ученым удалось выделить наиболее значимые признаки, которые потенциально способны предсказать рак на раннем этапе.

Основная цель настоящей работы заключается в сборе публичных данных об ультразвуковых исследованиях злокачественной опухоли, а также диагнозов, поставленных на их основе. Конечным результатом должна быть нейросетевая система, способная прогнозировать наличие злокачественной опухоли у пациента с вероятностью 90%.

В работе использованы методические приемы и опыт Пермской научной школы искусственного интеллекта в части применения нейросетевых технологий в области медицины [7-10].

Для создания нейросетевой системы были выбраны следующие параметры:

- X1 – среднее расстояний от центра до точек на периметре
- X2 – стандартное отклонение значений шкалы
- X3 – средний размер основной опухоли
- X4 – площадь опухоли
- X5 – среднее локальное изменение длины радиуса
- X6 – среднее значение периметра
- X7 – среднее значение выраженности вогнутых участков контура

- X8 – среднее число вогнутых частей контура
  - X9 – симметричность опухоли
  - X10 – среднее значение фрактальной размерности
  - X11 – стандартная ошибка среднего значения расстояний от центра до точек на периметре
  - X12 – стандартная ошибка для стандартного отклонения значений шкалы
  - X13 – стандартная ошибка для стандартного отклонения периметра опухоли
  - X14 – стандартная ошибка для стандартного отклонения площади опухоли
  - X15 – стандартная ошибка для локального изменения длины радиуса
  - X16 – стандартная ошибка для периметра
  - X17 – стандартная ошибка серьезности вогнутых частей контура
  - X18 – стандартная ошибка количества вогнутых частей контура
  - X19 – стандартная ошибка симметрии
  - X20 – стандартная ошибка фрактальных изменений
  - X21 – «наихудшее» или наибольшее среднее значение среднего расстояния от центра до точек на периметре
  - X22 – «худшее» или наибольшее среднее значение стандартного отклонения значений шкалы
  - X23 – «худшее» или наибольшее среднее значение стандартного отклонения периметра
  - X24 – «худшее» или наибольшее среднее значение стандартного отклонения площади
  - X25 – «наихудшее» или наибольшее среднее значение локального изменения длины радиуса
  - X26 – «худшее» или наибольшее среднее значение периметра
  - X27 – «наихудшее» или наибольшее среднее значение выраженности вогнутых участков контура
  - X28 – «наихудшее» или наибольшее среднее значение количества вогнутых частей контура
  - X29 – «наихудшее» или наибольшее среднее значение симметрии
  - X30 – наихудшая или наибольшая фрактальная размерность
- Выходной параметр – диагноз опухоли:
- 1 – злокачественный
  - 2 – доброкачественный

Обучающее множество было взято с сайта [www.kaggle.com](http://www.kaggle.com) [3]. Перед переходом к проектированию и обучению нейросетевой модели, была выполнена очистка исходного множества от противоречивых примеров, дубликатов, а также выбросов методом ящичковых диаграмм. Объем итогового множества включает в себя 569 примеров. Данное множество было разделено на обучающее, валидирующее и тестирующее в соотношении 70%, 20% и 10% соответственно.

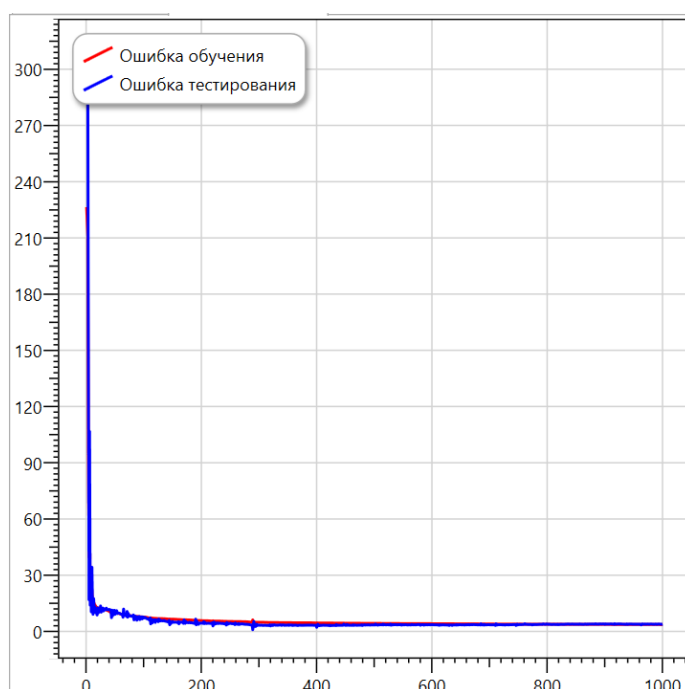
Проектирование, обучение, тестирование нейронной сети выполнялись с помощью программы «Нейросимулятор 5» [4]. В результате оптимизации

спроектированная нейронная сеть представляла собой персептрон, который имеет 30 входных нейронов. Для выходного слоя была выбрана функция активации: Тангенс гиперболический.



**Рисунок 1. Графическое представление сети**

Для обучения были отобраны обучающее и валидирующее множество 397 и 114 примеров соответственно. После обучения и оптимизации нейронной сети среднеквадратичное отклонения тестирования составило 9,5998%.



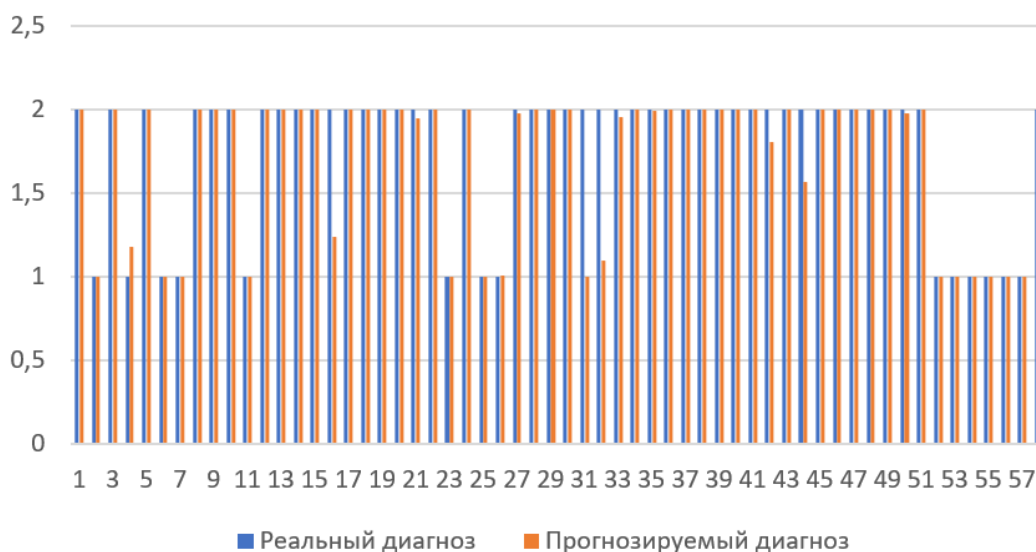
**Рисунок 2. Результат обучения сети**

Для оценки корректной работы спроектированной нейронной сети использовалось тестирующее множество, состоящее из 58 примеров. Средняя относительная ошибка тестирования составила 6,3%, что можно считать отличным результатом. На рисунке 3 представлена гистограмма, демонстрирующая разницу между фактическим и прогнозируемым нейросетью диагнозом для определения злокачественности опухоли. Из результатов, изображенных на рисунке 3, можно сделать вывод об адекватной работе спроектированной нейронной сети.

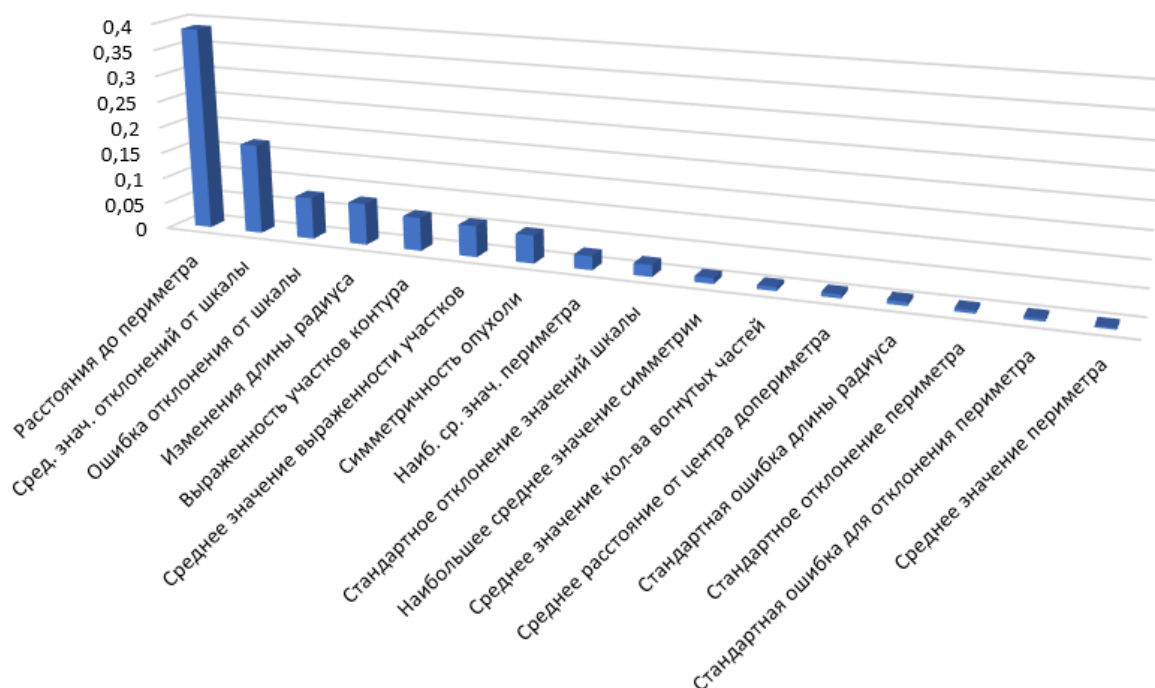
В разработке нейросети для обнаружения злокачественной опухоли были использованы 30 параметров. На основе данных значимости параметров, вычисленных нейросетью можно выделить 16 основных параметров (рис. 4). Среди них наиболее важными являются:

1. Наибольшее среднее значение среднего расстояния от центра опухоли до точек на периметре;

2. Наибольшее среднее значения стандартного отклонения значений шкалы;
3. Стандартная ошибка для стандартного отклонения значений шкалы;
4. Наибольшее среднее значение локального изменения длины радиуса опухоли;
5. Средне значение выраженности вогнутых участков контура;
6. Наибольшее среднее значение выраженности вогнутых участков контура;
7. Симметричность опухоли.



**Рисунок 3. График оценки степени достоверности полученных данных**



**Рисунок 4. Диаграмма значимости параметров**



Для более наглядной демонстрации влияния самого значимого параметра «наибольшее среднее значение расстояния от центра до точек на периметре» на выходной параметр «диагноз опухоли» был построен график (рис. 5). Исходя из полученных результатов можно сделать вывод, что при увеличении среднего расстояния от центра до точек на периметре частота злокачественной опухоли растет. Переломный момент наступает после расстояния, превышающего значение 18,33. При достижении значения в 21,93 нейросеть прогнозирует наличие злокачественной опухоли.



**Рисунок 5. Влияние расстояния до центра периметра на злокачественность**

Рассмотрим влияния второго по значимости параметра наибольшего среднего значения стандартного отклонения шкалы на наличие злокачественной опухоли в организме человека (рис. 6). При прочих равных условиях можно сделать вывод о том, что при отклонении от значений шкалы более чем на 55 единиц шанс обнаружить злокачественную опухоль существенно возрастает.



**Рисунок 6. Зависимость наличия злокачественной опухоли от отклонений шкалы**

При исследовании остальных параметров было выявлено, что они имеют слишком незначительное влияние на обнаружение злокачественной опухоли.

**Заключение.** Реализована математическая модель на основе технологий нейросетевого моделирования. Система, основываясь на параметрах выявления рака, позволяет определить какой является опухоль: злокачественной или доброкачественной. Таким образом, врачи могут на более ранних стадиях назначить подходящее лечение и существенно повысить шанс пациента на выздоровление.

### Библиографический список

1. Тасс.ру [Электронный ресурс] // URL: <https://tass.ru>. (Дата обращения 06.03.2022)
2. MedicalXpress [Электронный ресурс] // URL: <https://medicalxpress.com>. (Дата обращения 06.03.2022)
3. Kaggle.com [Электронный ресурс] // URL: <https://www.kaggle.com>. (Дата обращения 06.03.2022)
4. Черепанов Ф.М. Симулятор нейронных сетей для вузов. Пермь, 2012
5. Ясницкий Л.Н. Интеллектуальные системы : учебник. М.: Лаборатория знаний, 2016. 224 с.
6. Ясницкий Л.Н. Введение в искусственный интеллект. М.: Издательский центр «Академия», 2005. – 176 с
7. Yasnitsky L.N., Dumler A.A., Cherepanov F.M. Dynamic artificial neural networks as basis for medicine revolution. *Advances in Intelligent Systems and Computing*. 2018. Т. 850. С. 351-358.
8. Ясницкий Л.Н., Зайцева Н.В., Гусев А.Л., Шур П.З. Нейросетевая модель региона для выбора управляющих воздействий в области обеспечения гигиенической безопасности. *Информатика и системы управления*. 2011. № 3 (29). С. 51-59.
9. Ясницкий Л.Н., Грацилёв В.И., Куляшова Ю.С., Черепанов Ф.М. Возможности моделирования предрасположенности к наркозависимости методами искусственного интеллекта. *Вестник Пермского университета. Философия. Психология. Социология*. 2015. № 1 (21). С. 61-73.
10. Yasnitsky L.N., Dumler A., Cherepanov F.M. The capabilities of artificial intelligence to simulate the emergence and development of diseases, optimize prevention and treatment thereof, and identify new medical knowledge. *Journal of Pharmaceutical Sciences and Research*. 2018. Т. 10. № 9. С. 2192-2200.

# NEURAL NETWORK SYSTEM FOR MALIGNANT TUMOR PREDICTION

*Mokeyeva Anastasia S.*

National Research University “Higher School of Economics”  
Russia, Perm, 614990, Gagarin Boulevard st., 37,  
mokeyevanas@gmail.com

The article presents a description of the development of a neural network system for predicting the detection of a malignant tumor in the human body. The system makes it possible to predict with great accuracy whether a tumor is malignant based on various characteristics obtained during the examination. With the help of the developed intellectual system, a study of the subject area was carried out, patterns of practical importance were identified.

**Keywords:** artificial intelligence, neural network technologies, prediction, cancer.

УДК 004.032.26; 615.4

## МНОГОМАСШТАБНЫЕ СПЕКТРАЛЬНЫЕ ПРЕОБРАЗОВАНИЯ В КЛАССИФИКАТОРАХ СНИМКОВ ВИДЕОРЕЯДА

*Кондрашов Дмитрий Сергеевич, Белозеров Владимир Анатольевич,  
Горбачев Игорь Николаевич, Аль-Дарраджи Часиб Хасан*

Юго-Западный государственный университет  
30540, Россия, г. Курск, ул. 50 лет Октября, 94, SFilist@gmail.com

Разработан метод и программное обеспечение для классификации снимков видеоряда на основе многомасштабного преобразования Уолша-Адамара. Апробация метода и программного обеспечения осуществлена на УЗИ снимках поджелудочной железы.

**Ключевые слова:** сегментация снимка видеоряда, преобразование Уолша-Адамара, классификация снимков видеоряда, поджелудочная железа.

**Введение.** Методы анализа и классификация снимков видеоряда являются превалирующим инструментом при выполнении мониторинга посредством беспилотных летательных аппаратов, диагностики заболеваний по УЗИ снимкам, в системах безопасности и т.д. Особенности снимков видеоряда заключаются в том, что для их классификации требуется глобальная информация о картине мира, тогда как снимки видеоряда несут фрагментарную информацию. Поэтому лицом, принимающим решение (ЛПР), приходится

анализировать множество разноракурсных и разномасштабных изображений, которые трудно сопоставимы. Следовательно, из-за особенностей субъективного зрительного восприятия, снижается качество классификации снимков видеоряда. Поэтому возникает научно-техническая задача повышения достоверности и качества интерпретации снимков видеоряда за счет применения специализированного программного обеспечения, позволяющего снизить риск ошибок первого и второго рода.

**Методы и материалы.** Метод может быть использован для классификации снимков видеоряда от любых объектов и предполагает возможность двухэтапной классификации, на первом этапе определяется только наличие интереса к данному снимку или сегменту (класс ROI или класс HE ROI), а на втором этапе выносится окончательное решение по вопросу принадлежности снимка или сегмента к конкретному классу. При этом возникают трудности, как на первом, так и на втором уровнях классификации [1, 2, 3].

Для решения поставленных в работе задач используются методы классификации на основе «сильных» и «слабых» классификаторов. Для формирования дескрипторов для «слабых» классификаторов было предложено использовать преобразование Уолша-Адамара [4, 5, 6].

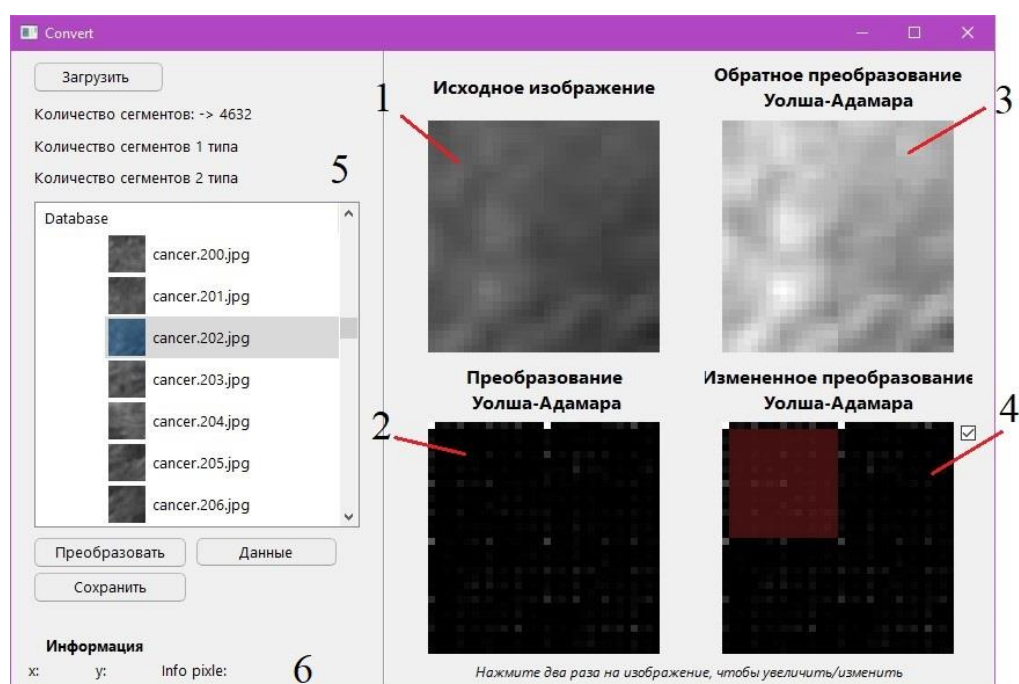
Изображение УЗИ разбивается на сегменты заданного размера с размером кратными двум. После этого ЛПР формирует базу данных. Для отнесения сегмента к заданному классу ЛПР использует глобальное изображение. Затем ЛПР отправляет сегмент в базу данных, которая и хранит сегмент с соответствующим маркером класса [7, 8, 9].

В качестве дата сет для полносвязной нейронной сети были использованы спектральные коэффициенты Уолша – Адамара. Коэффициенты Уолша-Адамара были разделены на «хранители» и «вершители». К «хранителям» были отнесены такие коэффициенты, частота появления которых наиболее высока в выборке определенного класса. Однако такой подход не позволяет судить о специфичности спектральных коэффициентов, так как одни и те же спектральные коэффициенты могут встречаться в выборках как того, так и другого класса. Еще одним способом определить «хранители» является пороговая селекция спектральных коэффициентов. Но для использования этого способа требуется классификатор, то есть признаком того, что спектральный коэффициент является «хранителем», является то, что при его исключении из вектора дескрипторов образец переходит в индифферентный класс.

Для определения «вершителей» необходимо взять две выборки, например, норма и рак. По этим выборкам сформировать базу данных по спектральным коэффициентам. По каждой выборке определяем матрицу средних значений спектральных коэффициентов по сегменту. Итого имеем две матрицы средних для двух классов. Из двух матриц средних получаем матрицу «вершителей». Каждый элемент матрицы «вершителей» определяется как процент разности соответствующих элементов этих матриц. Разность элементов матриц средних по классам нормируется по элементам одной из этих матриц, например, по элементам из матрицы норма, и умножается на 100. К «вершителям» отнесем те элементы, у которых процент разности превысит

некоторое пороговое значение. Спектральные коэффициенты, отнесенные к «вершителям» и будут вектором информативных признаков нейронной сети.

**Результаты.** Для обучения классификаторов использовалась база данных, содержащие аннотированные сегменты изображений УЗИ с онкологией, панкреатитом и индифферентным классом. Для создания базы данных, было разработано программное обеспечение, позволяющее загружать и разбивать на сегменты изображения УЗИ поджелудочной железы. Программное обеспечение позволяет задавать размеры сегментов любого размера. Однако настройка классификатора требует одинаковых размеров сегментов, поэтому для обучения классификаторов был определен и взят размер сегментов равный 32x32 пикселя.



**Рисунок 1. Изображение интерфейса программного обеспечения:**

**1) окно отображения исходного сегмента; 2) окно отображения двумерных спектральных коэффициентов Уолша; 3); окно отображения обратного преобразования Уолша 4); окно отображения 5) область закружки базы данных и выбора сегментов; 6) область отображения информации о пикселе**

Основой вычислительных процедур является программный модуль вычисления преобразования Уолша. На рисунке 1 представлен интерфейс разработанного программного обеспечения, который позволяет проводить разведочный анализ сегментов по исследованию релевантности двумерных спектральных коэффициентов различных классов патологии. На этом изображении в области 1 отображается исходный сегмента, вырезанный из исходного изображения, он загружается путем выбора мышью из области 5 выбора сегментов, вычисленные двумерные спектральные характеристики преобразования Уолша отображаются в области 2. В программном обеспечении присутствует режим изменения содержания сегмента, при активировании, которого ЛПП предоставляется возможность в ручном режиме выделить пиксели, которые на его усмотрение являются не релевантными спектральными

коэффициентами Уолша и не несут в себе необходимой информации, тем самым ЛПР сокращает вектор информативных признаков путем их обнуления, который отображается в области 4, где есть возможность, путем активирования «флажка» изменять вид отображения, между удаленными характеристиками и наложенным на исходный двумерный спектр Уолша из области 2, при этом в области 3 отобразится результат удаления не релевантных характеристик двумерного спектра Уолша.

В результате технология формирования дескриптора информативных признаков состоит в анализе двумерных частот Уолша и нахождения значимых спектральных коэффициентов двумерного спектрального разложения, путем визуального анализа изменений, которые внесли удалённые не релевантные характеристики Уолша. Причем значимые коэффициенты отбираются по пороговому и информационному критериям.

Наибольшая энергия двумерного спектра Уолша обычно находится в нулевой строке или столбце. Поскольку нулевой столбец нулевой строки равен взвешиванию изображения в окне с нулевой функцией Уолша, которая равна единице, нулевой столбец нулевой строки равен средней яркости изображения в окне.

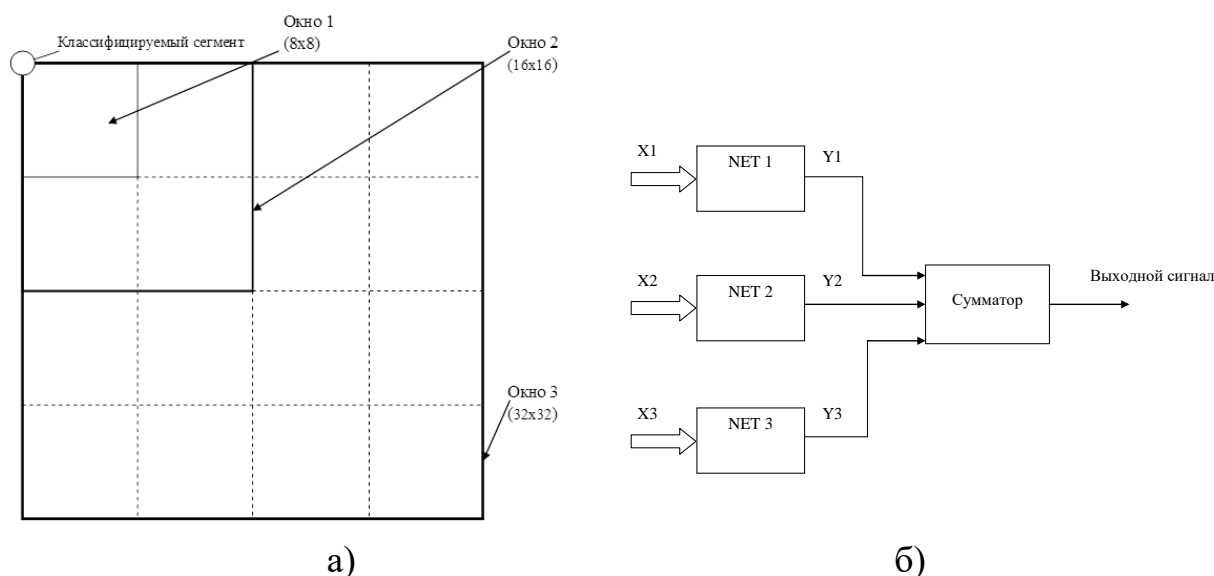
Одной из задач исследования является дифференциальная диагностика морфологических образований «онкология», «панкреатит» и индифферентного класса, на снимках УЗИ поджелудочной железы.

В классифицируемом сегменте формируются вложенные окна размерами 8x8, 16x16, 32x32 пикселя. В каждом из этих окон выполняется преобразование Уолша. На основании полученного множества окон и соответствующих им оконных спектров формируются обучающие выборки для классификаторов [10, 11, 12, 13].

Классификатор состоит из трех независимо обучаемых нейронных сетей. Для объединения выходов нейронных сетей используется простой блок усреднения по ансамблю. Такая схема классификатора представлена на рисунке 2б. Отличие её от ассоциативной машины состоит в том, что на входы параллельных нейронных сетей поступают одинаковые по размерности и различные по величине данные.

Вектор  $X_1$  – это спектральные коэффициенты 16 окон, вектор  $X_2$  – спектральные коэффициенты восьми окон, вектор  $X_3$  – спектральные коэффициенты в окне 3 рисунок 2а.

Целью исследования является обнаружение на изображениях УЗИ поджелудочной железы морфологических образований, соответствующих классов: «онкология», «панкреатит», индифферентный класс и минимизация ложноотрицательных срабатываний. В соответствии с выбранной целью были сформированы обучающие выборки для двухальтернативных классификаторов. На основе обучающих выборок построены три двухальтернативных классификатора со структурой, представленной на рисунке 2б.



**Рисунок 2. а) Спектральные окна для классификации сегмента;  
б) Блок схема классификатора сегмента**

**Заключение.** Проведена экспериментальная апробация программного обеспечения интеллектуальной системы по классификации УЗИ изображений поджелудочной железы по классам «нет области интереса» или «область интереса». При этом в класс «область интереса» входили снимки, классифицируемые как «онкология» или «панкреатит». Эксперименты на контрольных выборках показали диагностическую эффективность по классам «нет области интереса» – «область интереса» не ниже 90%.

При дифференциальной диагностике по классам «онкология» и «панкреатит» в индифферентный класс также включались патологические образования, не принадлежащие к нулевой гипотезе. При этом показатели диагностической эффективности на контрольных выборках по этим классам были не ниже 85%.

### Библиографический список

1. Филист, С.А. Автоматизированная система для классификации снимков видеопотоков / С.А. Филист, М.В. Шевцов, В.А. Белозеров, Д.С. Кондрашов, И.Н. Горбачев, Н.А. Корсунский // Известия Юго-Западного государственного университета. Серия: Управление, вычислительная техника, информатика. Медицинское приборостроение. – 2021. – Т. 11. № 4. – С. 85-105.
2. Томакова, Р.А. Интеллектуальные технологии сегментации и классификации биомедицинских изображений: монография / Р.А. Томакова, С.Г., Емельянов С.А. Филист. – Курск: Юго-Зап. гос. ун-т, 2012. – 222 с.
3. Дюдин, М.В. Автоматические классификаторы сложно структурируемых изображений на основе мультиметодных технологий многокритериального выбора / М.В. Дюдин, И.В. Зуев, С.А. Филист, С.М. Чудинов // Вопросы радиоэлектроники. Серия: Системы и средства отображения информации и управления спецтехникой (СОИУ). – 2015. – Выпуск 1. – С.130-141.

4. Кудрявцев, П.С. Моделирование морфологических образований на рентгенограммах грудной клетки в интеллектуальных диагностических системах медицинского назначения / П.С. Кудрявцев, А.А. Кузьмин, Д.Ю. Савинов, С.А. Филист, О.В. Шаталова // Прикаспийский журнал: управление и высокие технологии. – 2017. – № 3 (39). – С. 109-120.

5. Dabagov, A.R. An Automated System for Classification of Radiographs of the Breast / A.R. Dabagov, V.A. Gorbunov, S.A. Filist, I.A. Malyutina, D.S. Kondrashov // Biomedical Engineering. – 2020. – Vol. 53, No. 6. – Pp. 425-428. – URL: <https://doi.org/10.1007%2Fs10527-020-09957-7>.

6. Дабагов, А.Р. Многослойные морфологические операторы для сегментации сложноструктурируемых растровых полутонковых изображений / А.Р. Дабагов, И.А. Малютина, Д.С. Кондрашов и др. // Известия Юго-Западного государственного университета. Серия: Управление, вычислительная техника, информатика, медицинское приборостроение. – 2019. – Т.9, №3. – С. 44-63.

7. Филист, С.А. Метод каскадной сегментации рентгенограмм молочной железы / С.А. Филист, А.Р. Дабагов, Р.А. Томакова, И.А. Малютина, Д.С. Кондрашов // Известия Юго-Западного государственного университета. Серия: Управление, вычислительная техника, информатика. Медицинское приборостроение. – 2019. – Т. 9, № 1 (30). – С. 49-61.

8. Малютина И.А. Методы и алгоритмы анализа рентгенограмм грудной клетки, использующие локальные окна в задачах обнаружения патологий / И.А. Малютина, А.А. Кузьмин, О.В. Шаталова // Прикаспийский журнал: управление и высокие технологии. – 2017. – №3(39). – С.131-138.

9. Кассим, К.Д.А. Формирование признакового пространства для задач классификации сложноструктурируемых изображений на основе спектральных окон и нейросетевых структур / Кассим К.Д.А., Кузьмин А.А., Шаталова О.В., Е.А. Алябьев // Известия Юго-Западного государственного университета. – 2016. – №4(67). – С. 56-68.

10. Шаталова, О.В. Метод классификации сложноструктурируемых изображений на основе самоорганизующихся нейросетевых структур / О.В. Шаталова, А.А. Кузьмин, К.Д.А. Кассим, С.А. Филист // Радиопромышленность. – 2016. – №4. – С. 57-65.

11. Filist, S.A. Hybrid Intelligent Models for Chest X-Ray Image Segmentation / S.A. Filist, R.A. Tomakova, S.V. Degtyarev, A.F. Rybochkin // Biomedical Engineering. – 2018. – Vol. 51, No. 5. – Pp. 358-363. – URL: <https://doi.org/article/10.1007/s10527-018-9748-5>.

12. Филист, С.А. Гибридная нейронная сеть с макрослоями для медицинских приложений / С.А. Филист, О.В. Шаталова, М.А. Ефремов // Нейрокомпьютеры. Разработка, применение. – 2014. – №6. – С. 35-39.

13. Томакова, Р.А. Программное обеспечение автоматической классификации рентгенограмм грудной клетки на основе гибридных классификаторов / Р.А. Томакова, С.А. Филист, И.В. Дураков // Экология человека. – 2018. – № 6. – С. 59-64.



# MULTI-SCALE SPECTRAL TRANSFORMATIONS IN CLASSIFIERS OF VIDEO IMAGES

*Kondrashov Dmitry S., Belozеров Vladimir A.,  
Gorbachev Igor N., Al-Darraji Chasib Hasan*

Southwestern State University

Str. 50 years of October, 94, Kursk, Russia, 30540, SFilist@gmail.com

A method and software have been developed for classifying video sequence images based on the multiscale Walsh-Hadamard transform. Approbation of the method and software was carried out on ultrasound images of the pancreas.

**Keywords:** video sequence image segmentation, Walsh-Hadamard transform, video sequence image classification, pancreas

УДК 004.891.3

## МАТЕМАТИЧЕСКИЕ МОДЕЛИ КОЛИЧЕСТВЕННОЙ ОЦЕНКИ УРОВНЯ ЗАЩИТЫ ОРГАНИЗМА ОТ ВНЕШНИХ ФАКТОРОВ

*Крикунова Елена Владимировна<sup>1</sup>, Кадырова София<sup>1</sup>,  
Песок Валерия Вячеславовна<sup>1</sup>, Сафронов Руслан Игоревич<sup>2</sup>*

<sup>1</sup> Юго-Западный государственный университет», Курск  
30540, Россия, г. Курск, ул. 50 лет Октября, 94, SFilust@gmail.com

<sup>2</sup> Курская государственная сельскохозяйственная академия  
имени И.И. Иванова, 305021, Россия, г. Курск, ул. Карла Маркса, 70

В работе показано, что для количественной оценки уровня защиты организма от внешних факторов риска можно использовать показатели адаптационного потенциала, энергетического разбаланса биологически активных точек, адаптационного соответствия и ряда иммунологических показателей. Модель обеспечивает уверенность в принимаемых решениях оценки состояния функции ЦНС не ниже 0,9.

**Ключевые слова:** защита организма, центральная нервная система, воздействие внешних факторов, нечеткие гибридные модели.

**Введение.** Одной из важных задач адаптологии является донозологическая диагностика на основе знаний об адаптационном потенциале, являющимся показателем уровня защиты организма от внешних воздействий. В данном исследовании решается задача распространения показателей, характеризующих уровень защиты организма (УЗО) на органном и системном уровнях, на решение задач ранней диагностики заболеваний нервной систе-

мы, определяемых по характеру совместного воздействия экзогенных и эндогенных факторов риска (ФР).

Исследования показали, что показатель УЗО на общесистемном уровне обладает низкой специфичностью, которую можно увеличить, введя дополнительные признаки, характеризующие ФС и функциональный резерв органов (ФРО) и систем мишеней для рассматриваемых типов внешних воздействий.

Для оценки ФС и ФРО на уровнях органов, подсистем и систем использовались два блока показателей: электрические характеристики БАТ, «связанные» с исследуемыми биологическими структурами с идентификатором  $\ell$ , и показатели, общепринятые в медицинской практике (адаптационный потенциал, индекс напряжения по Р. Баевскому, реакция на комбинированные стимулы и др).

**Методы.** Для синтеза решающих правил оценки уровня защиты центральной нервной системы (ЦНС) необходимо учитывать особенности исследуемого класса заболеваний. В традиционной медицине выделяется несколько основных групп типов заболеваний ЦНС: заболевания инфекционной природы (клещевой энцефалит, корь, сифилис и др.); отравление организма (алкоголь, лекарственные препараты, ядовитые растения и др.); нарушение кровообращения сосудов головного мозга (инсульт, аневризма, ишемия, спазмы сосудов и др.); черепно-мозговые травмы, сильные ушибы; наследственные (приобретенные) заболевания ЦНС. Одним из мощных инструментов оценки функциональных состояний ЦНС является электроэнцефалография (ЭЭГ-обследование). Дополнительные диагностические возможности обеспечиваются методом вызванных потенциалов (ВП) [1, 2].

Эксперты для каждого свойства внимания: концентрированность (КВ), объем (ОВ), селективность (СВ), переключаемость (ПВ), распределяемость (РВ) и устойчивость (УВ) [3, 4, 5, 6], построили графики уровня защиты в виде функций принадлежности, которые описываются выражениями:

$$UZ(KB) = \begin{cases} 0,5 & , \text{если } KB < 150; \\ -0,004KB + 1,1, & \text{если } 150 \leq KB < 250; \\ 0,1 & , \text{если } 250 \leq KB. \end{cases}$$

$$UZ(OB) = \begin{cases} 0,5 & , \text{если } OB < 4500; \\ -0,00013OB + 1,1, & \text{если } 4500 \leq OB < 7500; \\ 0,1 & , \text{если } 7500 \leq OB. \end{cases}$$

$$UZ(CB) = \begin{cases} 0,55 & , \text{если } CB < 0,5; \\ -0,45CB + 0,775, & \text{если } 0,5 \leq CB < 1,5; \\ 0,1 & , \text{если } 1,5 \leq CB. \end{cases}$$

$$UZ(PB) = \begin{cases} 0,5 & , \text{ если } PB < 250; \\ -0,0025PB + 1,23, & \text{ если } 250 \leq PB < 450; \\ 0,1 & , \text{ если } 450 \leq PB. \end{cases}$$

$$UZ(PB) = \begin{cases} 0,1 & , \text{ если } PB < 400; \\ 0,00275PB - 1, & \text{ если } 400 \leq PB < 600; \\ 0,65 & , \text{ если } 600 \leq PB. \end{cases}$$

$$UZ(YB) = \begin{cases} 0,4 & , \text{ если } YB < 250; \\ -0,0015YB + 0,775, & \text{ если } 250 \leq YB < 450; \\ 0,1 & , \text{ если } 450 \leq YB. \end{cases}$$

Уровень защиты когнитивной функции внимания в целом определяется уравнением

$$UZV = \frac{1}{6} [UZ(KB) + UZ(OB) + UZ(CB) + UZ(PB) + UZ(PB) + UZ(YB)]. \quad (1)$$

Рассмотрим, как используя электрические характеристики БАТ оценить уровень защиты когнитивной функции внимания. Анализ атласов меридиан показал, что состояние ЦНС представлено значительным числом биологически активных точек, причем на них достаточно хорошо дифференцируются различные типы поражений головного мозга.

Исследования, проведенные в работе [3] показали, что «прямых связей» БАТ с функцией внимания нет, однако в силу достаточно сильной функциональной зависимости между свойствами внимания и памяти энергетические характеристики БАТ «связанных» с памятью можно использовать для решения задач прогнозирования и ранней диагностики нарушений функции внимания.

В этой же работе на статистическом материале доказана диагностическая эффективность соответствующих прогностических и диагностических решающих правил. С учетом этих результатов в данном разделе по точкам С3, С7, VС14, V15, V43 во французской классификации синтезируются решающие правила оценки уровня защиты функции внимания. В качестве ДЗТ в работе [3] выбраны точки меридиана сердца С3 и С7 они же оставлены как информативные БАТ, которых достаточно для решения практических задач.

ЭР  $EY_{sj}$  БАТ с именем  $Y_{sj}$  по выбранным как информативные, по отношению к ФС функции внимания точкам, эксперты определили следующим образом:

$$EY_{C3}^B = \begin{cases} 0, & \text{ если } \delta R_{C3} < 10\% ; \\ 0,0067\delta R_{C3} - 0,067, & \text{ если } 10\% \leq \delta R_{C3} < 70\% ; \\ 0,4, & \text{ если } \delta R_{C3} \geq 70\% , \end{cases}$$

$$EY_{C7}^B = \begin{cases} 0, & \text{если } \delta R_{C7} < 10\% ; \\ 0,016\delta R_{C7} - 0,16, & \text{если } 10\% \leq \delta R_{C7} < 60\% ; \\ 0,8, & \text{если } \delta R_{C7} \geq 60\% . \end{cases}$$

С учетом представлений экспертов о роли выбранных точек (точка C7 является главной (седативная и пособник), а C3 – обычная точка) в оценке уровня функционального состояния функции внимания уровень ЭР по этой паре точек определяется выражением:

$$ER_{BB} = EY_{C3}^B + EY_{C7}^B - EY_{C3}^B \cdot EY_{C7}^B . \quad (2)$$

Электрический разбаланс  $ER_{BB}$  выбирается в качестве базовой переменной для функции принадлежности  $\mu_{BB}(ER_{BB})$  к лингвистической переменной – нормальное функциональное состояние по функции внимания

График функции принадлежности  $\mu_{BB}(ER_{BB})$ , построенный экспертами с использованием метода Дельфы, описывается выражением:

$$\mu_{BB}(ER_{BB}) = \begin{cases} 0,85, & \text{если } ER_{BB} < 0,2 ; \\ -0,87ER_{BB} + 1,04, & \text{если } 0,2 \leq ER_{BB} < 0,6 . \\ 0,1, & \text{если } ER_{BB} \geq 0,6 ; \end{cases} \quad (3)$$

Используя  $\mu_{BB}(ER_{BB})$  как базовую переменную эксперты построили график уровня защиты функции внимания, определяемого по ФС этой когнитивной функции с помощью БАТ, который описывается формулой

$$UZV_B = \begin{cases} 0,1, & \text{если } \mu_{BB}(ER_{BB}) < 0,1 ; \\ 0,5\mu_{BB}(ER_{BB}) + 0,05, & \text{если } 0,1 \leq \mu_{BB}(ER_{BB}) < 0,7 . \\ 0,4, & \text{если } \mu_{BB}(ER_{BB}) \geq 0,7 . \end{cases} \quad (4)$$

Функциональный резерв определяется двумя частными функции уровня функционального резерва  $f_{FR\ell}(ON_\ell)$  и  $f_{FR\ell}(VV_\ell)$ , по которым рассчитывается ФРО для системы (органа) с идентификатором  $\ell$ .

Для решаемой задачи график нормирующей функции  $f_{FRV}(ON_V)$ , предложенный экспертами, описывается выражением:

$$f_{FRV}(ON_V) = \begin{cases} 0,5, & \text{если } ON_V < 1,0 ; \\ -0,16 ON_V + 0,66, & \text{если } 1,0 \leq ON_V < 3,5 . \\ 0,1, & \text{если } ON_V \geq 3,5 . \end{cases}$$

График нормирующей функции  $f_{FRV}(VV_V)$  описывается выражением:

$$f_{FRV}(VV_V) = \begin{cases} 0,4, & \text{если } VV_V < 1,0 ; \\ -0,05 VV_V + 0,45, & \text{если } 1,0 \leq VV_V < 7 . \\ 0,1, & \text{если } VV_V \geq 7 . \end{cases}$$

Используя показатель  $UZV_{FR}$ , как базовую переменную, определяется уровень защиты функции внимания, определяемого по ФРО этой когнитивной функции с помощью БАТ по показателю  $UZV_{FR}$ , который описывается выражением:

$$UZV_{FR} = \begin{cases} 0,1, & \text{если } FR_V < 0,1 \\ 0,67FR_V - 0,067, & \text{если } 0,1 \leq FR_V < 0,7. \\ 0,5, & \text{если } FR_V \geq 0,7 \end{cases} \quad (5)$$

Уровень защиты функции внимания  $UZV$  по показателям ФС и ФРО определяемый по электрическим характеристикам БАТ вычисляется по формуле:

$$UZV = UZV_B + UZV_{FR} - UZV_B \cdot UZV_{FR}. \quad (6)$$

С учетом  $UZB$  и  $UZV$  уровень защиты когнитивной функции внимания в соответствии с рекомендациями МСГНПП определяется выражением:

$$UZFV = UZB + UZV - UZB \cdot UZV. \quad (7)$$

Аналогичная стратегия может быть использована для синтеза решающих правил для других элементов центральной нервной системы.

**Результаты.** Для экспериментальной проверки качества работы полученных решающих правил были сформированы репрезентативные контрольные выборки по которым оценивались правильные и ошибочные решения диагностических моделей с расчетом таких статистических показателей как диагностическая чувствительность (ДЧ), специфичность (ДС) и эффективность (ДЭ), а также таких показателей, как прогностическая значимость положительных (ПЗ<sup>+</sup>) и отрицательных (ПЗ<sup>-</sup>) результатов.

Для контроля качества «работы» модели (7) оценки уровня защиты когнитивной функции внимания по показателю  $UZFV$  эксперты сформировали две группы испытуемых. Первая группа из 100 человек состояла из относительно здоровых студентов технического университета с нормальным уровнем защиты (нр). Во вторую группу были включены пациенты с кардиологическими проблемами, которых эксперты отнесли к классу «неудовлетворительный уровень защиты СУСР (н)». У всех обследуемых были определены показатели необходимые для расчета показателя  $UZFV$  (модель (7)). Для этой модели эксперты определили порог классификации на уровне 0,5. При  $UZFV \geq 0,5$  обследуемые относились к классу неудовлетворительный уровень защиты СУСР и наоборот. Показатели качества классификации принимают следующие значения: ДЧ=0,97; ДС=0,99; ДЗ<sup>+</sup>=0,99, ДЗ<sup>-</sup>=0,97, ДЭ=0,98.

Статистический анализ качества работы решающих правил (7) показал, что они достаточно надежно разделяют выбранные экспертами классы уровней защиты нервной системы.

**Заключение.** В ходе проведенных исследований была исследована роль защитных функций организма в противодействии различным внутренним и внешним ФР, выбрана система показателей, определены объект, методы и средства исследования. В качестве базового математического аппарата использована методология синтеза гибридных нечетких решающих правил.

2. Получена математическая модель оценки уровня защиты когнитивной функции внимания отличающаяся использованием гибридной нечеткой модели, включающей в себя показатели характеризующие различные свойства внимания в сочетании с электрическими характеристиками биологически активных точек, параметры которых зависят от состояния исследуемой функции позволяющая повысить качество оценки состояния этой функции ЦНС.

3. Синтезированы гибридные нечеткие модели диагностики патологии системы управления когнитивной функции внимания, обеспечивающие уверенность в принимаемых решениях не ниже 0,9, что делает оправданным применение полученных результатов в практической медицине.

### **Библиографический список**

1. Курзанов, А.Н. Совершенствование оценки функциональных резервов организма – приоритетное направление развития донологической диагностики преморбидных состояний / А.Н. Курзанов, А.Н. Заболотских, Д.В. Ковалев, Д.А. Бузиашвили // Международный журнал экспериментального образования. – 2015.-№ 10-1.- С. 67-70.

2. Филист, С.А. Модель формирования функциональных систем с учетом менеджмента адаптационного потенциала / Филист С.А., Шуткин А.Н., Шкатова Е.С., Дегтярев С.В., Савинов Д.Ю.// Биотехносфера 1(55) 2018. С.32-37.

3. Поляков А.В. Методы и средства прогнозирования и ранней диагностики когнитивной функции внимания // Дисс. канд. техн. наук. Курск. 2020. – 153 с.

4. Филист, С.А. Модели нечетких нейронных сетей с трех-стабильным выходом в инструментарии для психологических и физиологических исследований / С.А. Филист, Абдул Рахим Салем Халед, О.В. Шаталова, В.В. Руденко // Системный анализ и управление в биомедицинских системах. – 2007. – Т.6, №2. – С. 475-479. – URL: <https://www.elibrary.ru/item.asp?id=11666227>.

5. Филист, С.А. Гибридные решающие системы для прогнозирования послеоперационных осложнений у больных с доброкачественной гиперплазией предстательной железы / С.А. Филист, К.Д.А. Кассим, Р.В. Руцкой // Известия Юго-Западного государственного университета. – 2013. – № 5 (50). – С. 40-49.

6. Al-Kasasbeh R.T. Influence of Ergonomics of Electric Power Industry Enterprises on Nervous System Diseases/ Riad Taha Al-Kasasbeh,, Nikolay Korenevskiy, Altyn A. Aikeyeva, Mahdi S. Alshamasin, Sofia N. Rodionova, Ashraf Shaqdan, Sergey Filist and Yousif Eltous // Healthcare and Medical Devices, Vol. 51, 2022, 165–175. – URL: <https://doi.org/10.54941/ahfe1002113>.

# MATHEMATICAL MODELS FOR QUANTITATIVE ASSESSMENT OF THE LEVEL OF PROTECTION OF THE ORGANISM FROM EXTERNAL FACTORS

*Krikunova Elena Vl.<sup>1</sup>, Kadyrova Sofia<sup>1</sup>, Pesok Valeriya V., Safronov Ruslan I.<sup>2</sup>*

<sup>1</sup> Kursk Southwestern State University

St. 50 years of October, 94, Kursk, Russia, 30540, SFilist@gmail.com

<sup>2</sup> Kursk State Agricultural Academy named after I.I. Ivanova

Karl Marx street, 70, Kursk, Russia, 305021

The paper shows that for a quantitative assessment of the level of protection of the body from external risk factors, indicators of adaptive potential, energy imbalance of biologically active points, adaptive compliance and a number of immunological indicators can be used. The model provides confidence in the decisions made to assess the state of the CNS function of at least 0.9.

**Keywords:** body defense, central nervous system, external factors, fuzzy hybrid models.

УДК 004.032.26

## МОДЕЛЬ ДИАГНОСТИКИ И ОБОСНОВАНИЯ КОРОНАВИРУСНОЙ ИНФЕКЦИИ ПО ДАННЫМ МИКРОВОЛНОВОЙ РАДИОТЕРМОМЕТРИИ

*Попов Илларион Евгеньевич*

Волгоградский государственный университет

400062, г. Волгоград, пр-т Университетский, 100, popov.larion@volsu.ru

*Сайгин Павел Алексеевич*

Волгоградский государственный университет

400062, г. Волгоград, пр-т Университетский, 100, pashasaigin@mail.ru

В работе построено признаковое пространство по данным микроволновой радиотермометрии для алгоритмов классификации. Приведены результаты вычислительных экспериментов по алгоритмам нейронной сети и случайного леса. Предложен алгоритм обоснования поставленного диагноза по результату работы данных классификаторов.

**Ключевые слова:** искусственный интеллект, машинное обучение, микроволновая радиотермометрия.

**Введение.** На сегодняшний день опасность пандемии коронавирусной инфекции (covid-19) остается все такой же актуальной. По данным оператив-

ного штаба ежедневно обнаруживается около 500 000 новых случаев заболевания, так, например, на 3 сентября 2022 года по всему миру обнаружено 375 605 новых случаев заболеваний коронавирусной инфекцией, и такая тенденция сохраняется [5]. До сих пор основные методы диагностики – ПЦР-тест и компьютерная томография, но данные методы имеют свои недостатки. Так, например, ПЦР-тест не может показать степень тяжести болезни, а компьютерная томография не рекомендуется к использованию для обследования одного и того же пациента чаще, чем один раз в полгода. Из-за этого возникает необходимость в альтернативных методах диагностики коронавирусной инфекции, которые будут давать результат сразу или информировать о том, что нужно провести еще одно обследование, а также будут безопасны при частом использовании. Таким методом является микроволновая радиотермометрия, с помощью которой измеряются кожные и глубинные температуры тела [4]. По полученным данным определяются температурные аномалии в организме пациента, на основе чего ставится диагноз или сообщается о необходимости дальнейшей проверки основными методами диагностики.

Микроволновая радиотермометрия является методом диагностики, основанным на измерении интенсивности собственного излучения тканей в микроволновом диапазоне. Отличительной особенностью метода является возможность неинвазивно выявлять тепловые аномалии на глубине нескольких сантиметров, при этом метод полностью безвреден и безопасен. Перед началом обследования выполняется сбор первичных данных. После анализа данных обследования выносится заключение о состоянии здоровья, либо о необходимости проведения дальнейшего обследования иными методами.

**Материалы и методы.** Вычислительные эксперименты проводились по базе данных, полученной в результате медицинских обследований легких методом микроволновой радиотермометрии [4]. Схема измерений температур представлена на рисунке 1.

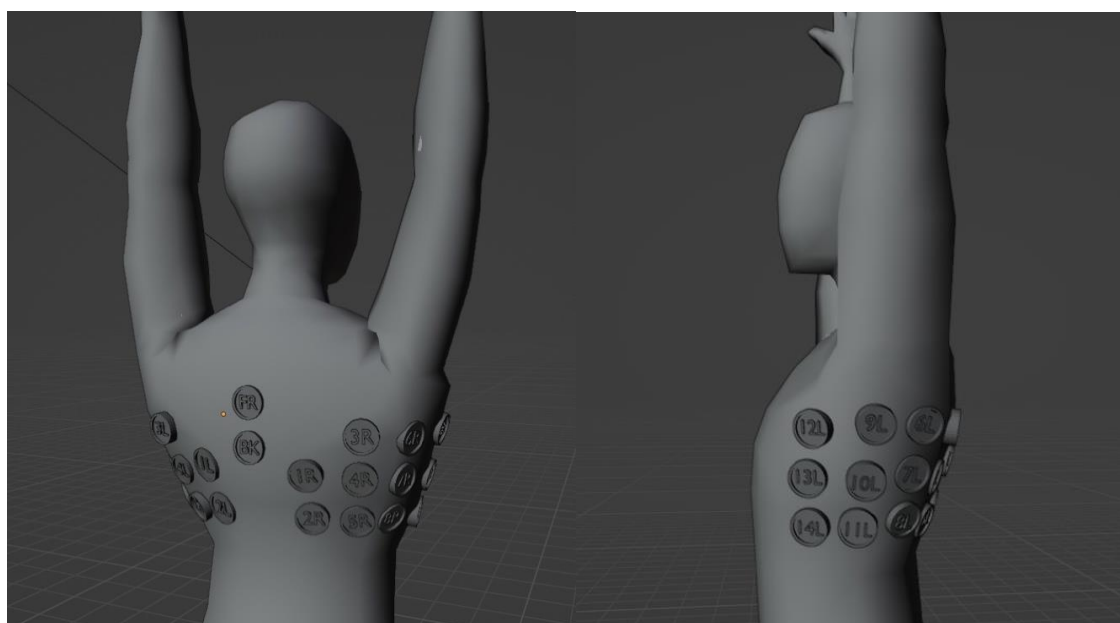


Рисунок 1. Схема обследования



По результаты обследований в базу данных записываются:

- кожные температуры правого легкого

$$T_r^{i,ir} = \{T_{1,r}^{i,ir}, T_{2,r}^{i,ir}, \dots, T_{14,r}^{i,ir}\}, \quad (1)$$

- кожные температуры левого легкого

$$T_l^{i,ir} = \{T_{1,l}^{i,ir}, T_{2,l}^{i,ir}, \dots, T_{14,l}^{i,ir}\}, \quad (2)$$

- глубинные температуры правого легкого

$$T_r^{i,mw} = \{T_{1,r}^{i,mw}, T_{2,r}^{i,mw}, \dots, T_{14,r}^{i,mw}\}, \quad (3)$$

- глубинные температуры левого легкого

$$T_l^{i,mw} = \{T_{1,l}^{i,mw}, T_{2,l}^{i,mw}, \dots, T_{14,l}^{i,mw}\}, \quad (4)$$

- кожные температуры опорных точек

$$T_p^{i,ir} = \{T_{0,p}^{i,ir}, T_{1,p}^{i,ir}\}, \quad (6)$$

- глубинные температуры опорных точек

$$T_r^{i,mw} = \{T_{0,p}^{i,mw}, T_{1,p}^{i,mw}\}. \quad (7)$$

Здесь  $i$  – номер пациента,  $ir$  – кожная температура,  $mw$  – глубинная температура,  $1, 2, \dots, 14$  – номер точки, согласно схеме на рисунке 1. У опорных точек индекс 0 соответствует точке FR на схеме, 1 – ВК.  $l$  – точки левого легкого,  $r$  – правого,  $p$  – опорные точки.

Помимо температур, измеренных методом микроволновой радиотермометрии, в базе данных содержится информация о диагнозе пациента (болен или здоров), характере обследования (проводились измерения у пациента через одежду или без нее), а также измеренная температура в подмышечной впадине термометром.

Для более эффективной диагностики строилось признаковое пространство. Так как концептуальная модель, описывающая поведение температур в области легких при заболевании covid-19, на данный момент слабо изучена, признаковое пространство было построено на основе предыдущих исследований [2]. Для этого были посчитаны дополнительные группы соотношений температур:

- термоассиметрия – разница между температурами в соответствующих точках измерений правого и левого легких. Например, термоассиметрия кожных температур в точке 0:

$$T_{0,l}^{i,ir} - T_{0,r}^{i,ir} \quad (8)$$

- внутренний градиент – разница между глубинной и кожной температурами в точке измерения. Например, внутренний градиент правого легкого в точке 0:5

$$T_{0,r}^{i,mw} - T_{0,r}^{i,ir} \quad (9)$$

Далее к данным группам температур и их соотношений (1)–(9) применялись различные функции для описания предполагаемых характерных особенностей заболевания (например, повышенная температура в одном из легких или повышенная глубинная температура). Применялись такие функции, как: среднее значение по данной группе температур или их соотношений, максимальное значение, высчиталось среднее квадратичное отклонение и т. д. Таким образом было построено признаковое пространство, состоящее из 84 признаков.

В качестве классификатора была выбрана нейронная сеть, разделяющая пациентов на две категории (1 – болен, 0 – здоров). За основу был взят персептрон, в котором использовалось 12 слоев. Входные данные были пронормированы на отрезок с центром в точке среднего значения и с радиусом равным среднее квадратичному отклонению. На входной слой подавались все 87 значений признаков (температурные признаки, метка одежды). На каждом из слоев на данные воздействовали: функция активации слоев (нами был выбран гиперболический тангенс), зашумление Гаусса (нами было выбрано значение 0,2) и предотвращение взаимoadaptации нейронов на этапе обучения. На выходном слое мы получаем вектор данных с предсказанием диагноза пациента из тестовой выборки.

Для обоснования поставленного нейронной сетью диагноза дополнительно обучался классификатор случайный лес. Данный классификатор представляет собой ансамбль деревьев решений, каждое из которых производит диагностику, по итогам которых происходит голосование. Алгоритм обоснования предлагается следующий:

1. Производится диагностика пациента нейронной сетью.
    - 1.1. В результате диагностики выявляется предполагаемый класс пациента (здоров или болен).
  2. Производится диагностика пациента алгоритмом случайный лес.
    - 2.1. Среди ансамбля алгоритмом ищется дерево решений, результат которого наиболее близок к результатам диагностики нейронной сетью. Степень близости считается по вероятности принадлежности пациента к классу из пункта 1.1.
  3. По выбранному дереву решений строится обоснования [3]. Для этого:
    - 3.1. Составляется множество узлов дерева решений, в которые пациент «попадает».
    - 3.2. Проверяется, выполняется ли условие в каждом из узлов.
    - 3.3. Если выполняется, то выводится информация узла, в случае выполнения.
    - 3.4. Иначе выводится информация узла в случае невыполнения.
- Заметим, что информация в узлах каждого из деревьев решений заполняется заранее, так как их структура остаётся неизменной после обучения.

**Результаты.** Вычислительные эксперименты проводились методом повторной кроссвалидации. Для этого изначальная выборка несколько раз разбивалась на 10 одинаковых подвыборок с сохранением соотношения больных и здоровых пациентов в каждой из них. Одна из подвыборок помещалась тестовой выборкой, остальные объединялись в обучающую выборки, на которой настраивались нейронная сеть и случайный лес. На тестовой выборке измерялась точность алгоритмов классификации и затем в неё помещалась следующая подвыборка.

По результатам вычислительных экспериментов были получены следующие показатели для нейронной сети:

- специфичность – 86.75%;
- чувствительность – 83.81%;
- эффективность – 84.65%.

Здесь специфичность – доля вернодиагностированных здоровых пациентов, чувствительность – доля вернодиагностированных больных пациентов, эффективность – среднегеометрическое значение по данным показателям.

Для случайного леса были получены следующие показатели:

- специфичность – 64.54%;
- чувствительность – 81.10%;
- эффективность – 72.24%.

Приведём пример полученного обоснования для одного здорового пациента и одного больного. Для больного пациента получилась следующая характеристика: среднее значение внутренних градиентов левого легкого в норме, температура в подмышечной впадине повышена, повышенное значение кожной температуры в опорной точке 0.

И для здорового пациента получилась следующая характеристика: значение кожной температуры в опорной точке 0 в норме, среднее значение глубинных температур и левого и правого легких в норме. Температурных аномалий не выявлено.

**Заключение.** В заключении отметим, что предложенный алгоритм обоснования не объясняет принцип работы нейронной сети, напротив, нейронная сеть и случайный лес работают в паре. Нейронная сеть с высокой точностью выявляет класс пациента и сообщает случайному лесу, какой из классификаторов ансамбля стоит использовать для обоснования.

Также отметим, что нейронная сеть эффективна при применении комбинированной диагностики. Так, при добавлении в признаковое пространство результатов измерения сатурации пациента, специфичность, чувствительность и, соответственно, эффективность, увеличиваются до 96%.

### **Библиографический список**

1. Goryanin I, Karbainov S, Shevelev O, Tarakanov A, Redpath K, Vesnin S, Ivanov Y. Passive microwave radiometry in biomedical studies. *Drug Discov Today*. 2020 Apr;25(4):757-763. doi: 10.1016/j.drudis.2020.01.016. Epub 2020 Jan 28. PMID: 32004473.

2. Levshinskii V. V. Mathematical models for analyzing and interpreting microwave radiometry data in medical diagnosis. Journal of Computational and Engineering Mathematics, Vol 8, No 1, 2021, pp. 3-14.

3. Losev, A.G., Popov, I.E., Petrenko, A.Y. et al. Some Methods for Substantiating Diagnostic Decisions Made Using Machine Learning Algorithms. Biomed Eng 55, 442–447 (2022). <https://doi.org/10.1007/s10527-022-10153-y>

4. Osmonov B, Ovchinnikov L, Galazis C, Emilov B, Karaibragimov M, Seitov M, Vesnin S, Losev A, Levshinskii V, Popov I, Mustafin C, Kasymbekov T, Goryanin I. Passive Microwave Radiometry for the Diagnosis of Coronavirus Disease 2019 Lung Complications in Kyrgyzstan. Diagnostics. 2021; 11(2):259. <https://doi.org/10.3390/diagnostics11020259>

5. Коронавирус: статистика [Электронный ресурс] // Коронавирус: статистика. <https://yandex.ru/covid19/stat> (дата обращения: 20.09.2022)

## MODEL OF DIAGNOSTICS AND JUSTIFICATION OF CORONAVIRUS INFECTION ACCORDING TO MICROWAVE RADIOMETRY

*Popov Illarion E., Saigin Pavel A.*

Volgograd State University

Prospekt Universitetskiy, 100, Volgograd, 400062, [popov.ilarion@volsu.ru](mailto:popov.ilarion@volsu.ru)

In this paper, a feature space is constructed according to microwave radiometry data for classification algorithms. The results of computational experiments on neural network and random forest algorithms are presented. An algorithm for substantiating the diagnosis based on the results of the work of these classifiers is proposed.

**Keywords:** artificial intelligence, machine learning, microwave radiometry.

УДК 004.032.26

## НЕЙРОСЕТЕВАЯ СИСТЕМА ПРОГНОЗИРОВАНИЯ СЕРДЕЧНОГО ПРИСТУПА

*Завгородних Екатерина Дмитриевна*

Национальный исследовательский университет «Высшая школа экономики»,  
614060, Россия, г. Пермь, ул. б-р Гагарина, 37А, [katya.zavgar@yandex.ru](mailto:katya.zavgar@yandex.ru)

В статье представлено описание разработки нейросетевой системы для прогнозирования сердечного приступа. Система позволяет предсказать вероятность сердечного приступа по показателям состояния здоровья.

С помощью нейросетевой системы были выявлены закономерности в данной предметной области, и проведено их исследование.

**Ключевые слова:** нейросетевая система, сердечный приступ, прогнозирование, вероятность.

**Введение.** На сегодняшний день сердечно-сосудистые заболевания находятся в десятке ведущих причин смертности. Людям нередко диагностируют сердечный приступ в момент его наступления. Позднее обнаружение приступа приводит к необратимым последствиям.

Нейросетевые системы способны диагностировать вероятность приступа по показателям состояния здоровья, а также являются системой проверки заключений врача [1]. Доклиническая диагностика с помощью нейросетевых систем предотвращает постановку неправильных диагнозов, дает возможность пациенту и врачу минимизировать риски сердечного приступа в силу раннего обнаружения вероятности приступа [1]. Для настоящей работы были изучены труды в области прогнозирования заболеваний [2-4]. Для обучения моделей за основу был взят принцип представления медицинских знаний в виде закономерностей.

Целью настоящей работы является создание набора данных [6] с релевантными примерами и обучение нейросетевой модели. Методики обучения взяты из вышеупомянутых трудов [2-5].

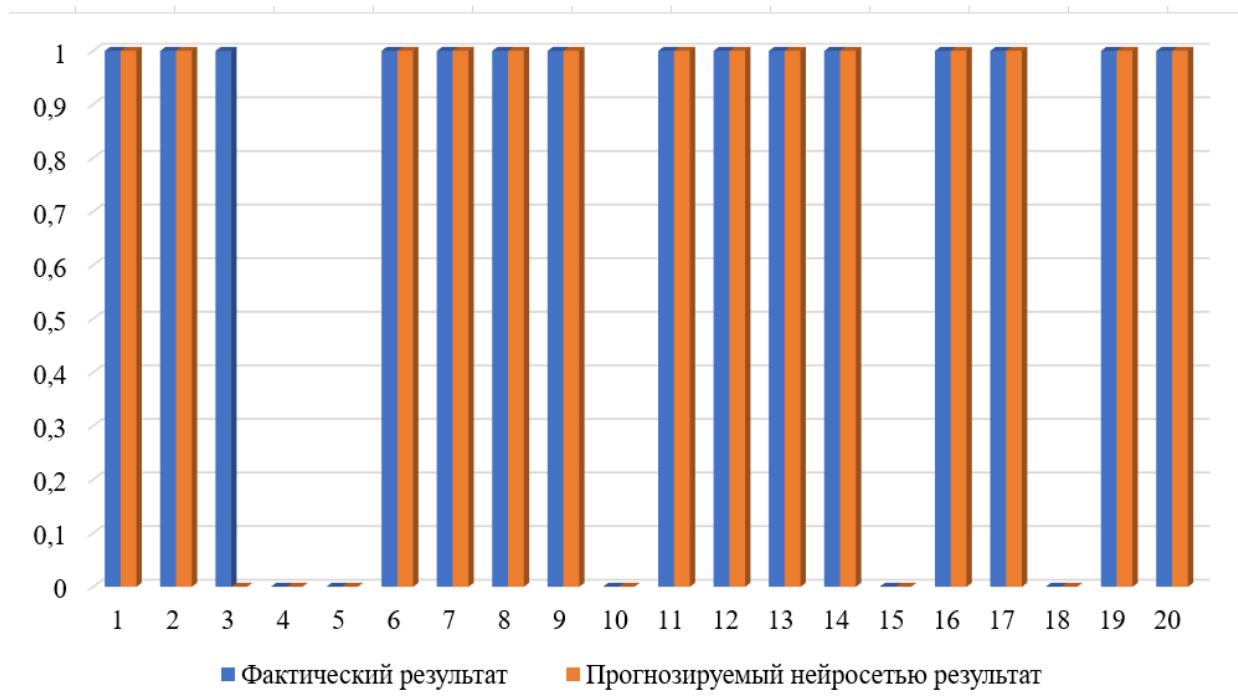
Для создания нейросетевой системы были выбраны следующие параметры: X1 – Возраст, X2 – Пол, X3 – Тип боли в груди, X4 – Артериальное давление в покое, X5 – Уровень холестерина в сыворотке крови, X6 – Сахар в крови натощак > 120 мг/дл, X7 – Результаты электрокардиографии в состоянии покоя, X8 – Максимальная достигнутая частота сердечных сокращений, X9 – Стенокардия, вызванная физической нагрузкой, X10 – Депрессия ST, вызванная физической нагрузкой по сравнению с покоем, X11 – Наклон сегмента ST пика физической нагрузки, X12 – Количество крупных сосудов (0-3), окрашенных при флуороскопии, X13 – Крутизна пульса.

Некоторые параметры необходимо было закодировать, «Пол»: 1 – мужчина, 0 – женщина, «Тип боли в груди»: 1 – типичная стенокардия, 2 – атипичная стенокардия, 3 – неангинальная боль, 4 – бессимптомно. Также параметры «Сахар в крови натощак > 120 мг/дл» и «Стенокардия, вызванная физической нагрузкой» были закодированы таким образом: 1- да, 0 – нет. Параметр «Результаты электрокардиографии в состоянии покоя» также был представлен в закодированном формате: 0 – нормальные, 1 – наличие аномалии ST-T, 2 – указание вероятной или определенной гипертрофии левого желудочка по критериям Эстеса.

Данные для обучающего множества были взяты с интернет-ресурса Kaggle [6]. Размер обучающего множества составил 158 примеров.

Создание нейросети осуществлялось с помощью программы «Нейросимулятор 5» [7] по методике [5]. В результате оптимизации спроектированная нейронная сеть представляла собой персептрон, который состоит из тринадцати входных нейронов, одного скрытого слоя с пятью нейронами и одного выходного нейрона.

Для оценки корректной работы спроектированной нейронной сети использовалось тестирующее множество, состоящее из 20 примеров. Средняя относительная ошибка тестирования составила 5%, что считается допустимым результатом для дальнейшей работы с нейросетью. На Рисунок 1 представлена гистограмма, которая наглядно демонстрирует разницу между фактическим и прогнозируемым сетью результатом вероятности возникновения сердечного приступа.



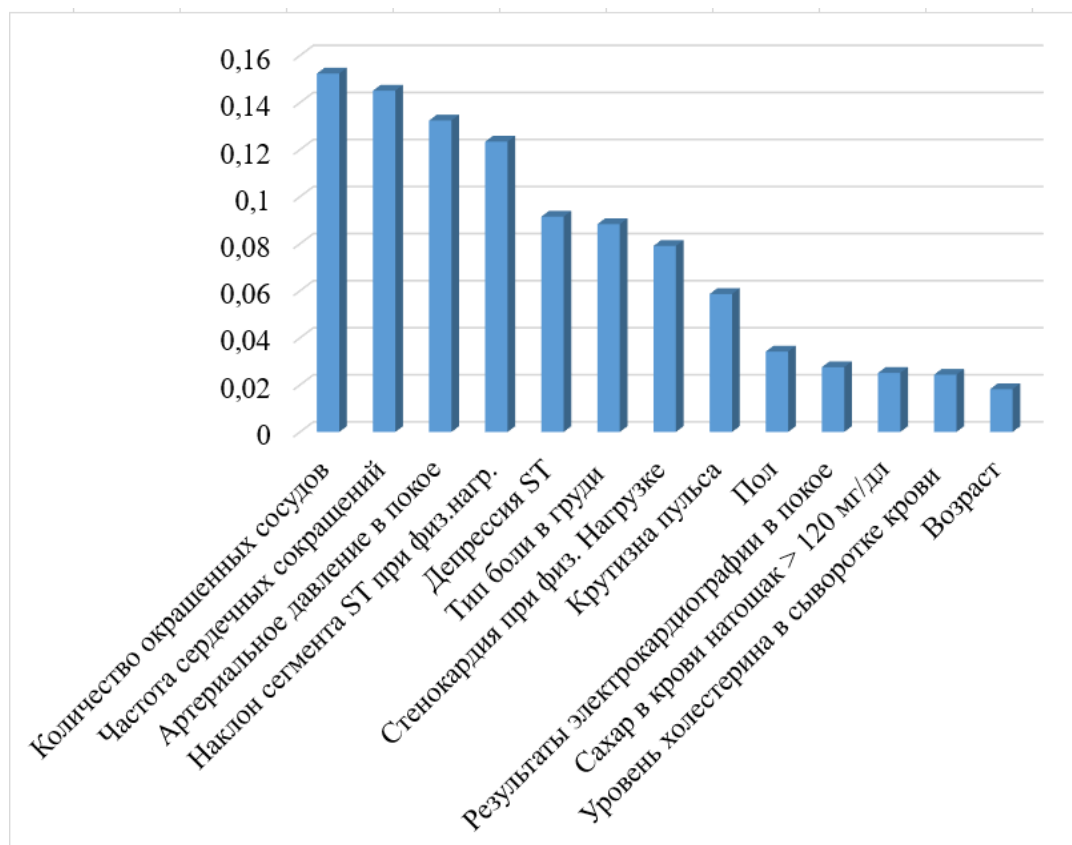
**Рисунок 1. Тестирующее множество**

По результатам гистограммы на Рисунке 1 можно сделать вывод об адекватности работы спроектированной нейросети.

Оценка значимости входных параметров была выполнена с помощью программы «Нейросимулятор 5». Для наглядности значимости параметров представлена гистограмма на Рисунок 2.

Из гистограммы видно, что самыми значимыми параметрами являются: количество окрашенных при флуороскопии сосудов, максимальная частота сердечных сокращений и артериальное давление в покое. Это вполне ожидаемый результат, поскольку состояние сосудов, частота сердечных сокращений и давление – это основные направления, которые необходимо наблюдать при сердечно – сосудистых заболеваниях.

После доказательства адекватности обученной модели необходимо провести исследование предметной области на конкретных примерах. Проанализировать зависимости входных и выходного параметра. Исследование производилось с помощью метода «замораживания» [7], суть которого заключается в варьировании значения одного параметра и наблюдении за выходным параметром при фиксировании значений всех других входных параметров. Для исследования были отобраны 4 пациента, данные которых представлены ниже в Таблица 1.



**Рисунок 2. Значимость параметров**

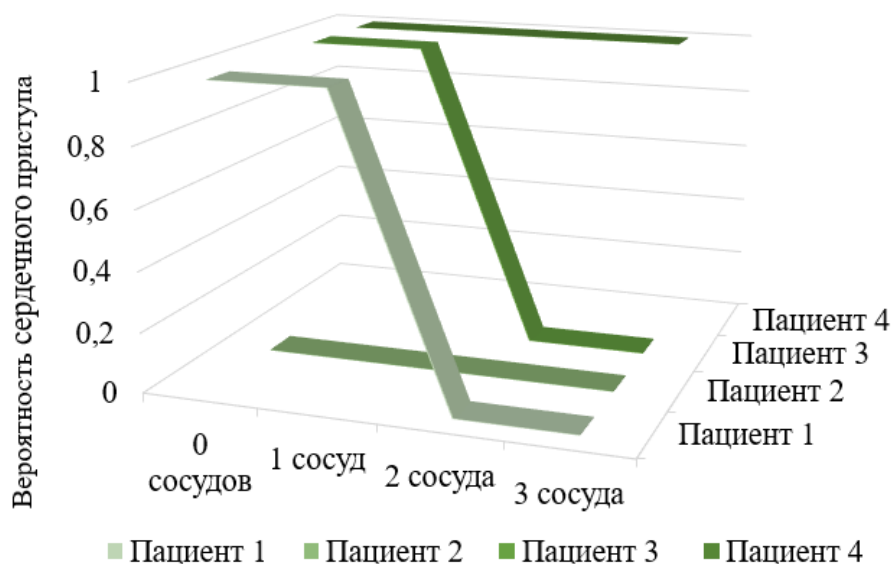
**Таблица 1**

**Данные пациентов для исследования зависимостей**

№ Пациент	x1	x2	x3	x4	x5	x6	x7	x8	x9	x10	x11	x12	x13
Пациент 1	64,0	1,0	0,0	128,0	263,0	0,0	1,0	105,0	1,0	0,2	1,0	1,0	3,0
Пациент 2	67,0	1,0	2,0	152,0	212,0	0,0	0,0	150,0	0,0	0,8	1,0	0,0	3,0
Пациент 3	56,0	1,0	3,0	120,0	193,0	0,0	0,0	162,0	0,0	1,9	1,0	0,0	3,0
Пациент 4	46,0	0,0	1,0	105,0	204,0	0,0	1,0	172,0	0,0	0,0	2,0	0,0	2,0

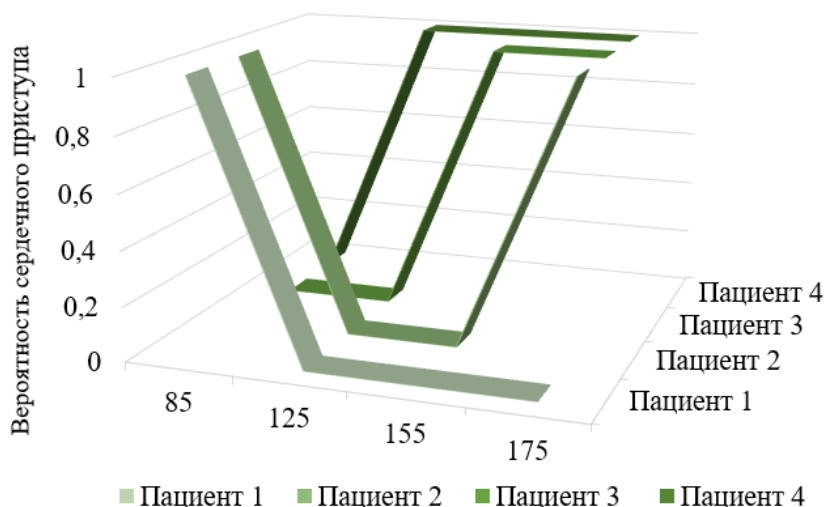
В первую очередь необходимо исследовать и наглядно представить зависимость выходного параметра (вероятность сердечного приступа) от «Количества крупных сосудов, окрашенных при флюороскопии», поскольку это самый значимый параметр. Для прогнозирования у пациентов необходимо было изменять параметр количества сосудов от 0 до 3. Ниже представлен график (Рисунок 3) влияния значимого параметра «Количество окрашенных сосудов при флюороскопии» на результат. Можно сделать вывод о следующей зависимости: чем больше крупных сосудов у пациента окрашивается при флюороскопии, тем меньше вероятность сердечного приступа.

Далее нужно проанализировать зависимость выходного параметра от «Максимальная достигнутая частота сердечных сокращений». Для поиска зависимости параметр частоты сердечных сокращений был 85, 125, 155, 175. Результаты представлены на



**Рисунок 3. Влияние количества окрашенных сосудов на вероятность сердечного приступа**

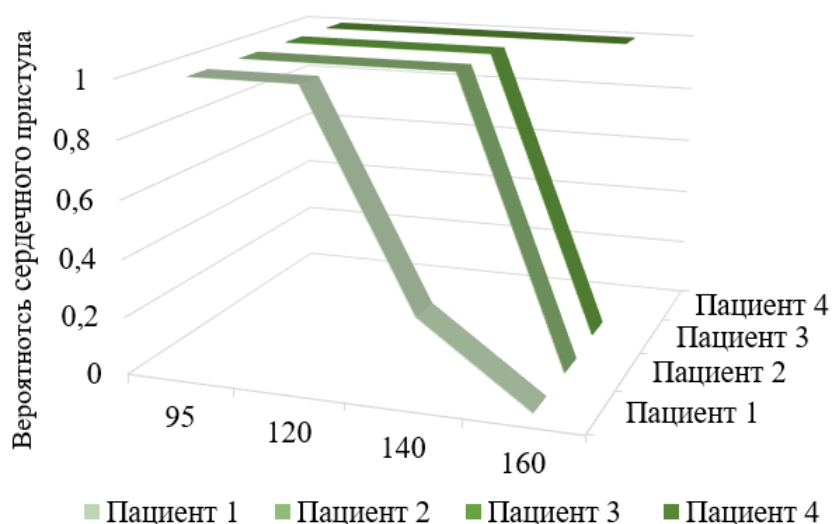
Рисунок 4. Нельзя однозначно говорить о прямой зависимости частоты сердечных сокращений и выходного параметра. Однако, в трех из четырех случаев оказалось, что, чем больше частота сердечных сокращений, тем выше вероятность сердечного приступа.



**Рисунок 4. Влияние максимальной частоты сердечных сокращений на вероятность сердечного приступа**

Завершающим значимым для нейросетевой модели параметром является «Артериальное давление в покое». На примере четырех пациентов параметр давления будет меняться следующим образом: 95,120,140,160. Необходимо проанализировать, меняется ли выходной параметр при изменении уровня артериального давления. Результат прогнозирования нейросетью представлен на Рисунок 5. В 75% случаев при росте артериального давления в покое вероятность сердечного приступа уменьшается.





**Рисунок 5. Влияние артериального давления в покое на вероятность сердечного приступа**

Прогнозируемые нейросетью результаты не противоречат реальности и являются пригодными для использования в диагностике сердечных приступов по показателям здоровья.

**Заключение.** Построена система нейросетевого прогнозирования сердечного приступа. Нейросетевая модель основывается на 13-ти параметрах: пол, возраст, тип боли в груди, артериальное давление в покое, уровень холестерина в сыворотке крови в мг/дл, сахар в крови натощак > 120 мг/дл, результаты электрокардиографии в состоянии покоя, максимальная достигнутая частота сердечных сокращений, стенокардия, вызванная физической нагрузкой, депрессия ST, вызванная физической нагрузкой по сравнению с покоем, наклон сегмента ST пика физической нагрузки, количество крупных сосудов (0-3), окрашенных при флюороскопии, крутизна пульса. Методом сценарного прогнозирования построены графики зависимостей выходного параметра от наиболее значимых параметров. Данный набор параметров позволяет диагностировать возможность сердечного приступа до его наступления.

### Библиографический список

1. Ясницкий Л.Н., Думлер А.А., Полещук А.Н., Богданов К.В., Черепанов Ф.М. Нейросетевая система экспресс-диагностики сердечно-сосудистых заболеваний // Пермский медицинский журнал. 2011. Т. 28. № 4. С. 77-86.
2. Yasnitsky L.N., Dumler A.A., Cherepanov F.M. Dynamic artificial neural networks as basis for medicine revolution // Advances in Intelligent Systems and Computing. 2018. Т. 850. С. 351-358. 5.
3. Yasnitsky L.N., Dumler A., Cherepanov F.M. The capabilities of artificial intelligence to simulate the emergence and development of diseases, optimize prevention and treatment thereof, and identify new medical knowledge // Journal of Pharmaceutical Sciences and Research. 2018. Т. 10. № 9. С. 2192- 2200.
4. Ясницкий Л.Н., Зайцева Н.В., Гусев А.Л., Шур П.З. Нейросетевая модель региона для выбора управляющих воздействий в области обеспечения

гигиенической безопасности // Информатика и системы управления. 2011. № 3 (29). С. 51-59.

5. Ясницкий Л.Н., Сичинава З.И. Нейросетевые алгоритмы анализа поведения респондентов // Нейрокомпьютеры: разработка, применение. 2011. № 10. С. 59-64.

6. Набор данных для обучения [Электронный ресурс]. URL: <https://www.kaggle.com/pritsheta/heart-attack>

7. Нейросимулятор 4.0. Свидетельство о регистрации программы для ЭВМ RUS 2014612546. Заявка № 2014610341 от 15.01.2014.

## **A NEURAL NETWORK SYSTEM FOR PREDICTING A HEART ATTACK**

*Zavgorodnikh Ekaterina D.*

National Research University “Higher School of Economics”  
Str. Gagarina, 37a, Perm, Russia, 614060, katya.zavgar@yandex.ru

This article describes the development of neural network system for heart attack prediction. The system allows to predict the probability of a heart attack by health indicators. With the help of a neural network system were identified patterns in the subject area, and conducted their study.

**Keywords:** neural network system, heart attack, prediction, probability.

УДК 004.032.26

## **ПРОГНОЗИРОВАНИЕ ДИАБЕТА МЕТОДОМ НЕЙРОСЕТЕВОГО МОДЕЛИРОВАНИЯ**

*Куликова Елена Александровна*

Пермский государственный национальный исследовательский университет,  
ПМИ. 614990, Россия, г. Пермь, ул. Букирева, 15,  
kulikova-lena1998@yandex.ru

В статье представлено описание разработки нейросетевой системы для прогнозирования диабета. Система позволяет с большой точностью предсказать, болен ли человек диабетом в зависимости от симптомов и признаков.

**Ключевые слова:** искусственный интеллект, нейросетевые технологии, прогнозирование, сахарный диабет, симптомы диабета, прогнозирование диабета.

**Введение.** Диабет – серьезное хроническое заболевание, при котором люди теряют способность эффективно регулировать уровень глюкозы в крови, что может привести к снижению качества и продолжительности жизни

[1]. После того, как во время пищеварения различные продукты расщепляются на сахар, сахар попадает в кровоток. Это сигнализирует поджелудочной железе о высвобождении инсулина. Инсулин помогает клеткам в организме использовать эти сахара в кровотоке для получения энергии. Диабет обычно характеризуется либо тем, что организм не вырабатывает достаточно инсулина, либо не может использовать инсулин, который вырабатывается настолько эффективно, насколько это необходимо.

По данным ВОЗ [2], на сегодняшний день диабетом страдает около 422 млн человек, что составляет 6,028% от всего населения планеты. Статистика заболеваемости диабетом ежегодно растёт. Если ситуация будет развиваться теми же темпами, то к 2025 году количество пациентов с диабетом увеличится в 2 раза. К 2030 году сахарный диабет станет 7-й причиной смерти во всём мире.

Осложнения, такие как сердечные заболевания, потеря зрения, ампутация нижних конечностей и заболевание почек, связаны с хронически высоким уровнем сахара, остающимся в кровотоке у людей с диабетом. Хотя от диабета нет лекарства, такие стратегии, как похудание, здоровое питание, активный образ жизни и лечение, могут уменьшить вред от этого заболевания для многих пациентов. Ранняя диагностика может привести к изменению образа жизни и более эффективному лечению, что делает модели прогнозирования риска диабета важными инструментами для государственных и общественных органов здравоохранения.

*Основная цель* настоящей работы заключается в обработке и отборе данных из датасета [3], а также создание и обучение нейросетевой модели на этих данных. Конечным результатом должна быть нейросетевая система, способная прогнозировать удовлетворенность посетителя отеля в больше чем 70% случаев.

Для создания нейросетевой системы были выбраны следующие параметры: X1 – возраст, X2 – пол, X3 – полиурия (чрезмерное мочеиспускание), X4 – полидипсия (чрезмерная жажда), X5 – внезапная потеря веса, X6 – слабость, X7 – полифагия (нарушение пищевого поведения), X8 – грибковая инфекция, X9 – нечеткость зрения, X10 – зуд, X11 – раздражительность, X12 – замедленное заживление ран, X13 – ослабление мышц, X14 – ригидность (неподатливость) мышц, X15 – алопеция (выпадение волос), X16 – ожирение.

Среди входных параметров модели имеются качественные и количественные факторы. Качественные были зашифрованы следующим образом: «Пол»: мужчина – 0, женщина – 1; «полиурия»: не было чрезмерного мочеиспускания – 0, было чрезмерное мочеиспускание – 1; «полидипсия»: пациент не испытывал чрезмерную жажду – 0, пациент испытывал чрезмерную жажду – 1; «внезапная потеря веса»: не было – 0, было – 1; «слабость»: нет – 0, да – 1; «полифагия»: не было нарушения пищевого поведения – 0, было нарушение пищевого поведения – 1; «грибковая инфекция»: не было – 0, было – 1; «нечеткость зрения»: не было – 0, было – 1; «зуд»: не было – 0, было – 1; «раздражительность»: не было – 0, было – 1; «замедленное заживление ран»: не было замечено – 0, было замечено – 1; «ослабление мышц»: не было – 0, было – 1; «ригидность мышц»: не было – 0, было – 1; «алопеция»: не было – 0, было – 1; «ожирение» (исходя из его индекс массы тела): нет – 0, да – 1;

Количественный показатель «Возраст» представляют собой целые числа.

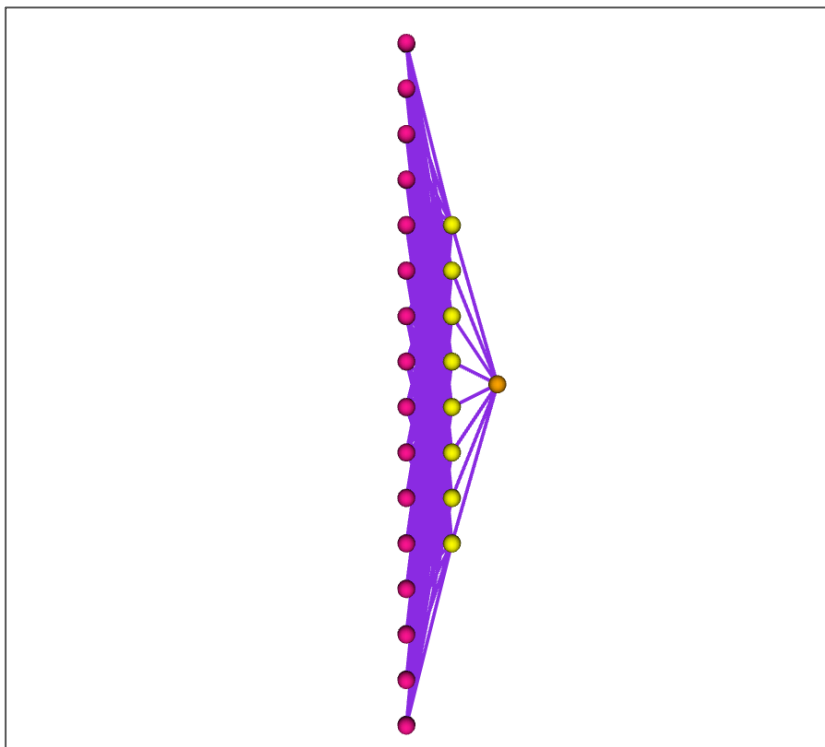
На выходе нейронная сеть выдает вероятность диабета значением от 0 до 1.

Обучающее множество было взято из социальной сети специалистов по обработке данных и машинному обучению Kaggle [4] и включала выборку, собранных с помощью прямых анкет и результатов диагностики у пациентов диабетической больницы в Силхете (Бангладеш) [3].

Перед переходом к проектированию и обучению нейросетевой модели, были выполнены следующие шаги:

1. очистка исходного множества от противоречивых примеров,
2. очистка исходного множества от выбросов,
3. очистка исходного множества от дубликатов.

Таким образом, объем итогового множества включает в себя 468 примера. Данное множество было разделено на обучающее и тестовое множества в соотношении 85% и 15%, соответственно. Таким образом, в обучающее множество попало 398 примера, а в тестовое – 70 пример.

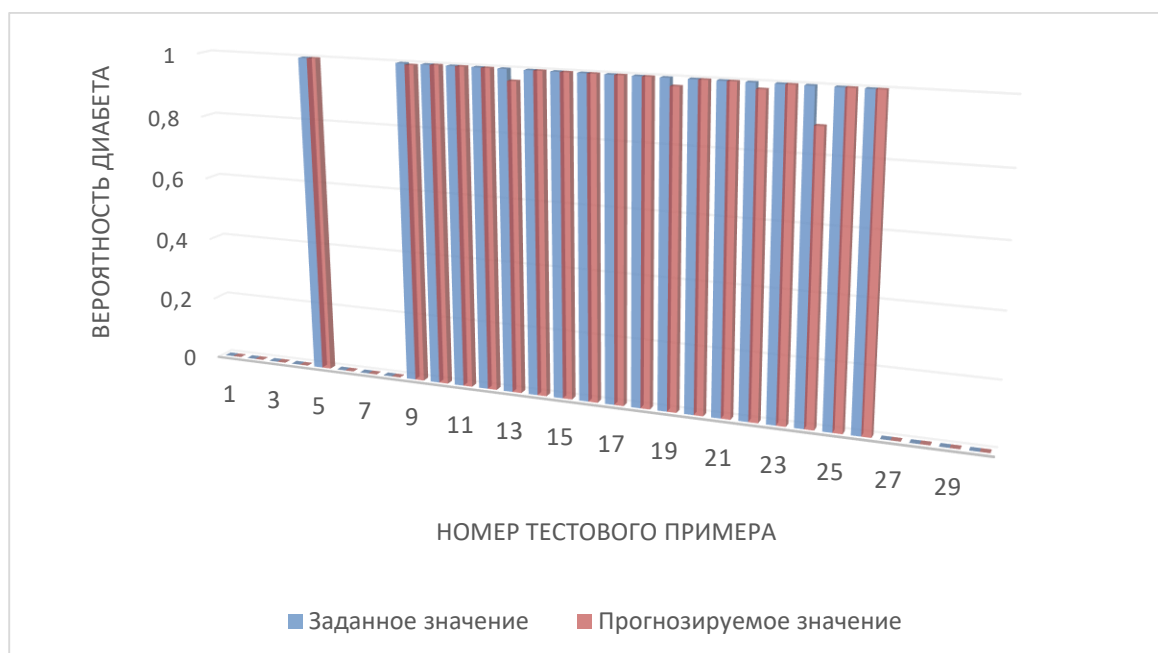


**Рисунок 1. Графическое представление нейронной сети**

Проектирование, обучение и тестирование нейросетевой модели выполнялись с помощью программы «Нейросимулятор 5» [5] по методике [6]. В результате оптимизации спроектированная нейронная сеть представляла собой персептрон, который имеет 16 входных нейронов, один скрытый слой с 8-ью нейронами и 1 выходной нейрон.

На рисунке 1 представлено графическое представление данной нейронной сети.

Средняя относительная ошибка тестирования составила 5,134%, что можно считать приемлемым результатом. На рисунке 2 представлена гистограмма, демонстрирующую разницу между фактическим и прогнозируемым нейросетью результатом.



**Рисунок 2. Результат тестирования обученной нейросетевой модели**

Из результатов, изображенных на рисунке 1, можно сделать вывод об адекватной работе спроектированной нейронной сети.

Оценка значимости входных параметров была выполнена с помощью программы «Нейросимулятор 5», результат отображен на рисунке 3.

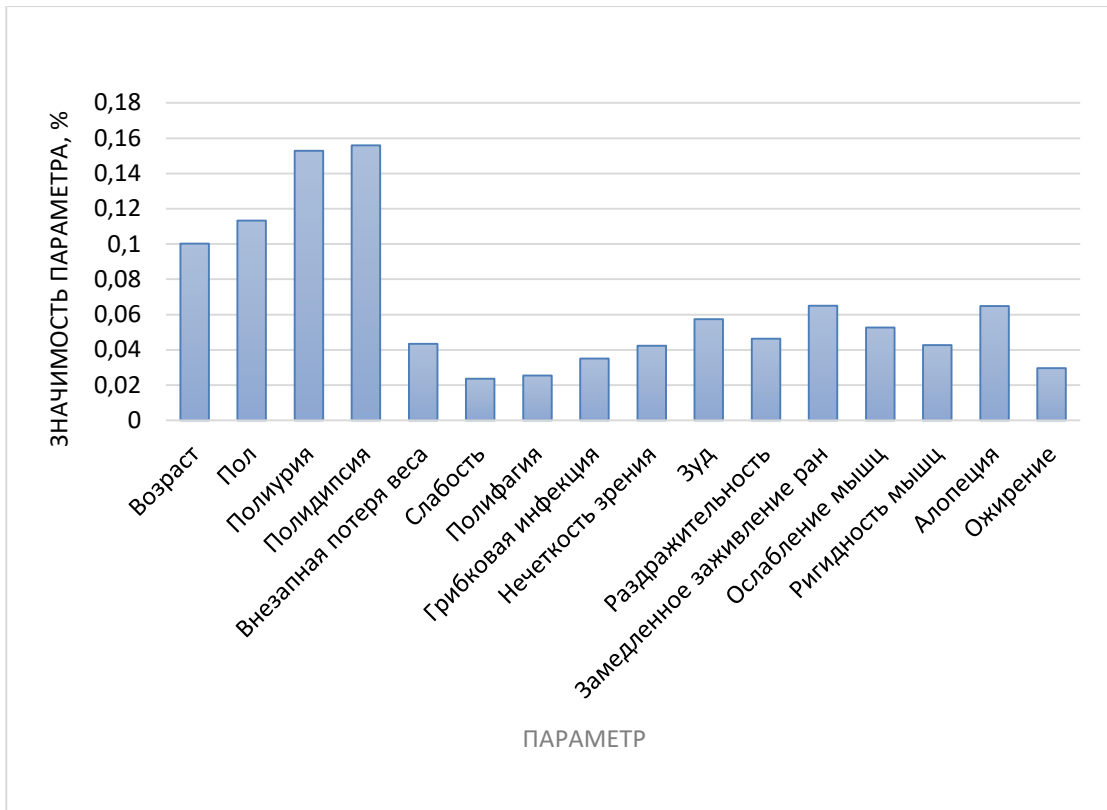
На основании данных, продемонстрированных на рисунке 3, наиболее значимыми параметрами нейросетевой модели являются X4 (**полидипсия – чрезмерная жажда**), X3 (**полиурия – чрезмерное мочеиспускание**), X2 (**пол**) и X1 (**возраст**).

Полученный результат подтверждается исследованием многих врачей: «главными признаками сахарного диабета являются сильная жажда, частое мочеиспускание с большим количеством мочи, иногда обезвоживание организма (дегидратация)» [7].

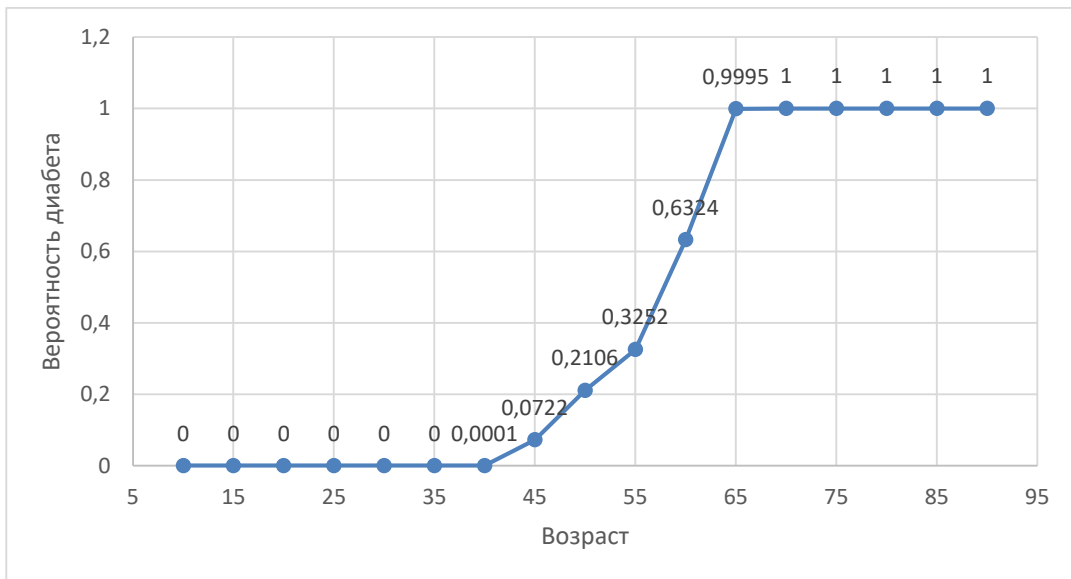
Далее было проведено исследование полученных зависимостей между входными параметрами и удовлетворенностью постояльцев отелей. Исследование проводилось с помощью метода «замораживания» [6], суть которого заключается в варьировании значения одного параметра и фиксации значений всех других параметров.

На рисунке 4 показан график влияния входного параметра «Возраст». Как следует из графика, вероятность диабета начинает расти в возрасте от сорока до шестидесяти пяти лет.

В качестве примера был взят пациент, у которого есть жалобы только на «слабость» (все параметры 0, параметр «слабость» – 1).



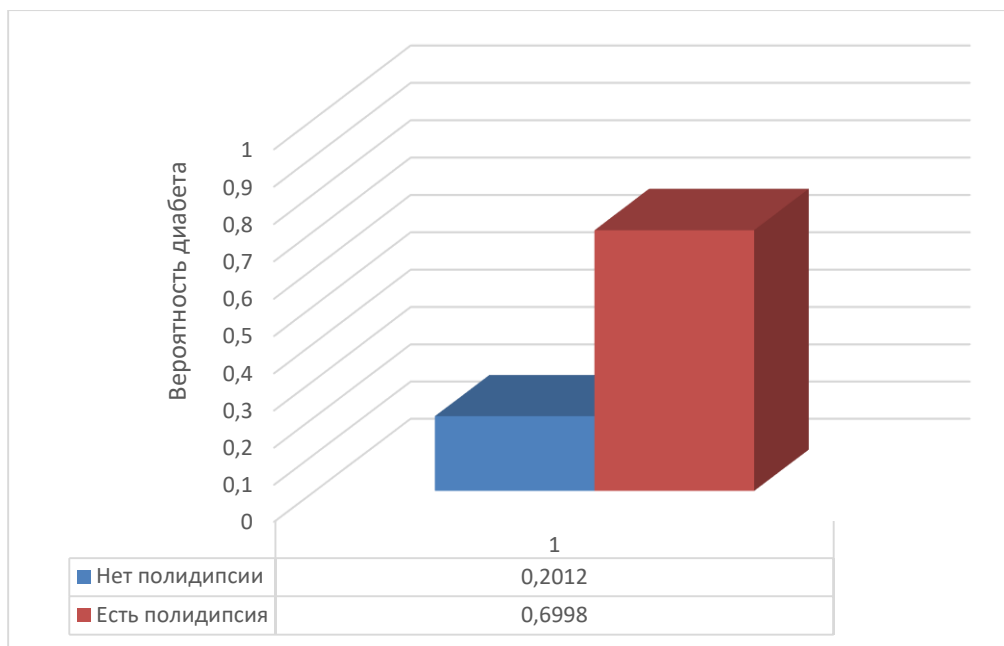
**Рисунок 3. Значимость параметров нейросетевой модели**



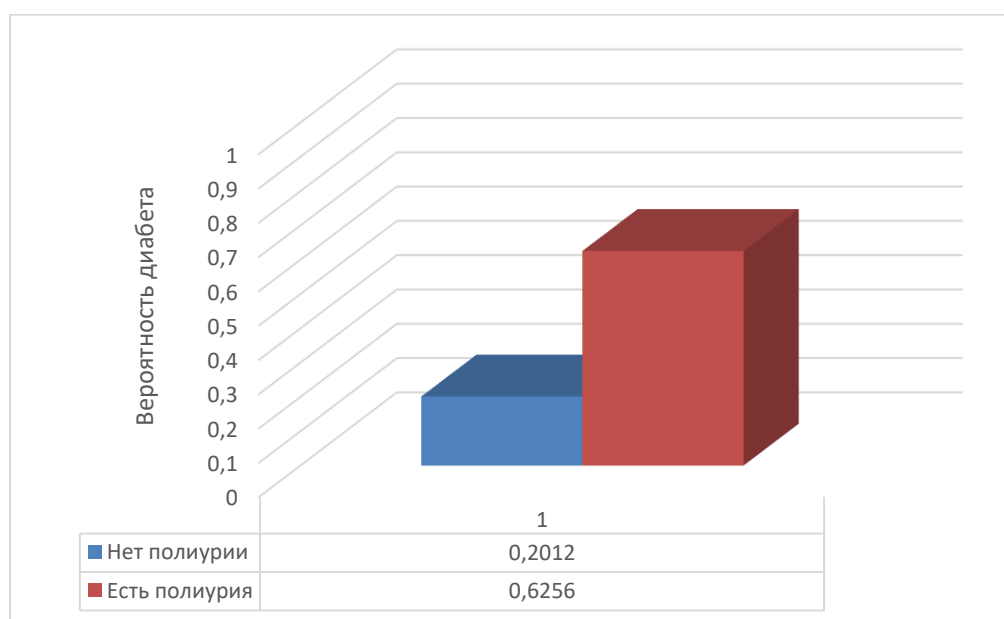
**Рисунок 4. Зависимость прогнозируемой вероятности диабета от возраста**

Для следующих двух примеров зависимостей, взят пример пациента: женщина 30ти лет, у которой жалобы только на «слабость» (все параметры 0, параметр «слабость» – 1).

На рисунке 5 продемонстрирована зависимость прогнозируемой вероятности диабета от полидипсии (чрезмерной жажды). По результатам данного исследования выяснилось, что вероятность того, что пациент болен диабетом резко возрастает, если у него присутствует симптом чрезмерной жажды.



**Рисунок 5. Зависимость прогнозируемой вероятности диабета от полидипсии (чрезмерной жажды)**



**Рисунок 6. Зависимость прогнозируемой вероятности диабета от полиурии (чрезмерного мочеиспускания)**

Следующим было проведено исследование параметра «полиурия», который по значимости не на много меньше предыдущего (см. рис. 3).

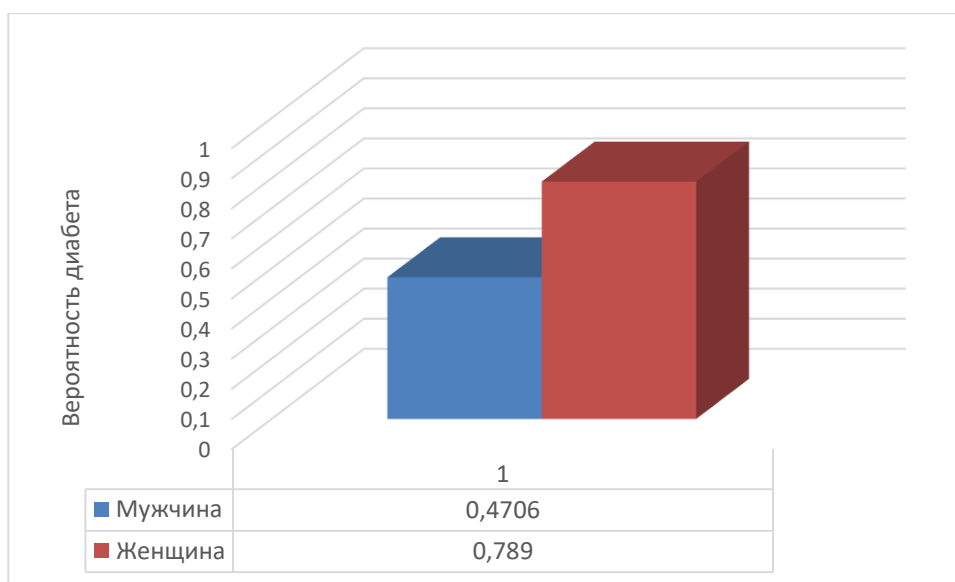
На рисунке 6 продемонстрирована зависимость прогнозируемой вероятности диабета от полиурии (чрезмерное мочеиспускание). Опять же прослеживается тенденция резкого роста вероятности, если у пациента присутствует симптом чрезмерного мочеиспускания.

Это говорит о том, что симптомы «полидипсия» и «полиурия» являются главными признаками сахарного диабете.

На рисунке 7 продемонстрирована зависимость прогнозируемой вероятности диабета от полиурии (чрезмерное мочеиспускание). Для данного исследования взят пример 30ти летнего пациента, у которого есть жалобы на усталость и чрезмерное выпадение волос (параметры «слабость» и «алопеция» равны, остальные параметры равны 1).

По результатам видно, что, при одинаковых параметрах, у женщин вероятность заболеваемости диабетом больше, чем у мужчин. И действительно, соотношение заболевших женщин и мужчин составляет 60 к 40. Женщины болеют чаще и имеют больше факторов риска развития заболевания [8].

Такие результаты могут натолкнуть на следующую мысль: вероятно, стоит организовывать какие-то групповые мероприятия именно для тех посетителей, которые приехали в отель не одни, а с друзьями, близкими, родственниками и т.п. Организация подобных мероприятий поможет привлечь больше людей, отдающих предпочтение путешествиям в компании близких людей, что в свою очередь положительно скажется на их удовлетворенности как от отеля, так и от поездки в целом, и, соответственно, поможет увеличить прибыль владельцу гостиничного комплекса.



**Рисунок 7. Зависимость прогнозируемого значения от рола пациента**

Полученные результаты исследований не противоречат реальности, что подтверждает, что спроектированную нейронную сеть можно считать пригодной для прогнозирования вероятности диабета.

**Заключение.** Построена система вероятности диабета. Спроектированная модель учитывает десять индивидуальных показателей, таких как: возраст, пол, полиурия (чрезмерное мочеиспускание), полидипсия (чрезмерная жажда), внезапная потеря веса, слабость, полифагия (нарушение пищевого поведения), грибковая инфекция, нечеткость зрения, зуд, раздражительность, замедленное заживление ран, ослабление мышц, ригидность (неподатливость) мышц, аллопатия (выпадение волос) и ожирение.



Проведены исследования работы данной модели в рамках предметной области, в которой эта модель работает корректно. Посредством прогнозирования были выявлены главные симптомы для определения человека в группе риска и своевременной профилактики.

### Библиографический список

1. Тарасенко Н. А. Сахарный диабет: действительность, прогнозы, профилактика // Современные проблемы науки и образования. – 2017. – № 6. – С. 34-34. URL: <https://s.science-education.ru/pdf/2017/6/27144.pdf>
2. Global report on diabetes. URL: <https://www.who.int/publications/i/item/9789241565257>
3. Early Diabetes Classification. Predict the risk of diabetes using 16 tabular features. URL: <https://www.kaggle.com/andrewmvd/early-diabetes-classification>
4. Kaggle – социальная сеть специалистов по обработке данных и машинному обучению. [Электронный ресурс]. Режим доступа: <https://www.kaggle.com>.
5. Черепанов Ф.М., Ясницкий Л.Н. Нейросимулятор 4.0. Свидетельство о регистрации программы для ЭВМ RUS 2014612546. Заявка № 2014610341 от 15.01.2014.
6. Ясницкий Л.Н., Сичинава З.И. Нейросетевые алгоритмы анализа поведения респондентов // Нейрокомпьютеры: разработка, применение. 2011. № 10. С. 59-64.
7. Сахарный диабет: признаки, симптомы, лечение, питание. URL: <https://www.mgb1-74.ru/novosti/154-saharnyj-diabet-priznaki-simptomu-lechenie-pitanie-dieta-pri-diabete.html>
8. Дедов И. И., Шестакова М. В., Викулова О. К., Железнякова А. В., Исаков М. А. Эпидемиологические характеристики сахарного диабета в российской федерации: клинико-статистический анализ по данным регистра сахарного диабета на 01.01.2021 // Сахарный диабет. 2021. №3. URL: <https://cyberleninka.ru/article/n/epidemiologicheskie-harakteristiki-saharnogo-diabeta-v-rossiyskoy-federatsii-kliniko-statisticheskiy-analiz-po-dannym-registra>

## НЕЙРОСЕТЕВАЯ СИСТЕМА ПРОГНОЗИРОВАНИЯ ПРОДОЛЖИТЕЛЬНОСТИ ЖИЗНИ ЧЕЛОВЕКА

*Дементьев Михаил Александрович*

Пермский государственный национальный исследовательский университет,  
ПМИ. 614990, Россия, г. Пермь, ул. Букирева, 15,  
mixail.dementev.2002@mail.ru

В статье представлено описание разработки нейросетевой системы для прогнозирования продолжительности жизни человека. Система позволяет с большой точностью предсказать продолжительность жизни человека на основании данных о разных людях. С помощью разработанной интеллектуальной системы проведено исследование предметной области, выявлены закономерности, имеющие практической значение.

**Ключевые слова:** искусственный интеллект, нейросетевые технологии, прогнозирование, продолжительность, жизни.

**Цель работы.** Создание компьютерной программы, основанной на параметрах – зависимостях, определяющих продолжительность жизни человека. Инструментом для выявления таких закономерностей и построение на их основе мат. моделей, являются методы искусственного интеллекта: нейросети [1, 2].

**Введение.** В последнее время все чаще возникают трудно формализуемые задачи, примером которой и является задача прогнозирования продолжительности человеческой жизни. Показатели смертности и ожидаемой продолжительности жизни населения являются основными критериями, характеризующими уровень общественного здоровья, естественного движения и качество жизни. Именно поэтому проблема расчета уровня смертности стоит очень остро.

Так в частности Эрикссон представил результаты уникального по своей масштабности исследования [3], получившего название «Н-70». Его участниками стали люди, родившиеся в 1901-1902 годах и наблюдавшиеся в возрасте от 70 до 90 лет.

Эрикссон исследовал влияние социального статуса пожилых людей на длительность их жизни и предложил четыре параметра, определяющих эту зависимость.

Первые два параметра обусловлены социальными обязательствами и соглашениями, которые усиливают сходство между пожилыми людьми. Третий реализуется в зависимости от того, как человек выстраивает и поддерживает уважение к самому себе, насколько успешно он реагирует на перемены. Четвертый включает в себя регулярное взаимодействие с другими людьми и общение. Они снижают уровень тревожности и дают ощущение поддержки

при принятии ежедневных решений, а также улучшают память и внимание и тренируют мозг.

Для прогнозирования продолжительности жизни Эрикссон использовал методы, лежащие в основе искусственных нейронных сетей (ИНС). Оказалось, что этот метод более применим для сложных ситуаций, где не срабатывают традиционные методики.

Также в [4] обучили нейронную сеть прогнозировать продолжительность жизни людей на основе 19 астрологических точек и пола человека. В качестве обучающего множества было использовано 2331 наблюдение из астрологического банка данных. Авторам статьи удалось достичь значения среднеквадратичной ошибки равной 10,8% на 100 тестовых наблюдениях, не входящих в обучающую выборку.

Становится ясно, нейросетевые технологии актуальны во многих сферах, в том числе и медицине, поэтому было решено разработать нейронную сеть, которая могла бы прогнозировать продолжительности жизни отдельно взятого человека. Данная программа помогла бы в анализе качества жизни и в вопросах её продления.

**Методика:** для прогнозирования были выбраны наиболее важные параметры, большинство из которых предполагают социальные и личностные проблемы, которые оказали влияние в данный момент или при развитии человека. **Параметры:** X1-Пол: 1- мужской, 2-женский; X2-Семейное положение: 1-женат/замужем, 2-холост, одинок, 3-вдовец/вдова, 4-разведен; X3-Вид деятельности (типология профессий Климова): 1-Человек, 2-Техника, 3-Художественный образ, 4-Природа, 5-Знак; X4-Состав семьи: 0-Полная, 1-Неполная; X5-Алкогольная зависимость: 0-отсутствует, 1-присутствует; X6-Наркотическая зависимость: 0-отсутствует, 1-присутствует; X7-Психические расстройства: 0-отсутствует, 1-присутствует; X8-Проблемы со здоровьем: 0-отсутствует, 1-присутствует. D-выходной сигнал, определяющий прогнозируемую длительность жизни человека(в годах)

Типология профессий Е.А.Климова (советский и российский психолог, доктор психологических наук)

Климов выделил пять объектов труда: **человек, техника, художественный образ, знак, природа.** В первой части названия типа профессии обозначен субъект труда, которым всегда является человек.

**Человек – человек** – все профессии, связанные с воспитанием, обслуживанием, обучением людей, общением с ними. К этой группе относятся все педагогические и медицинские профессии, профессии сферы услуги и другие.

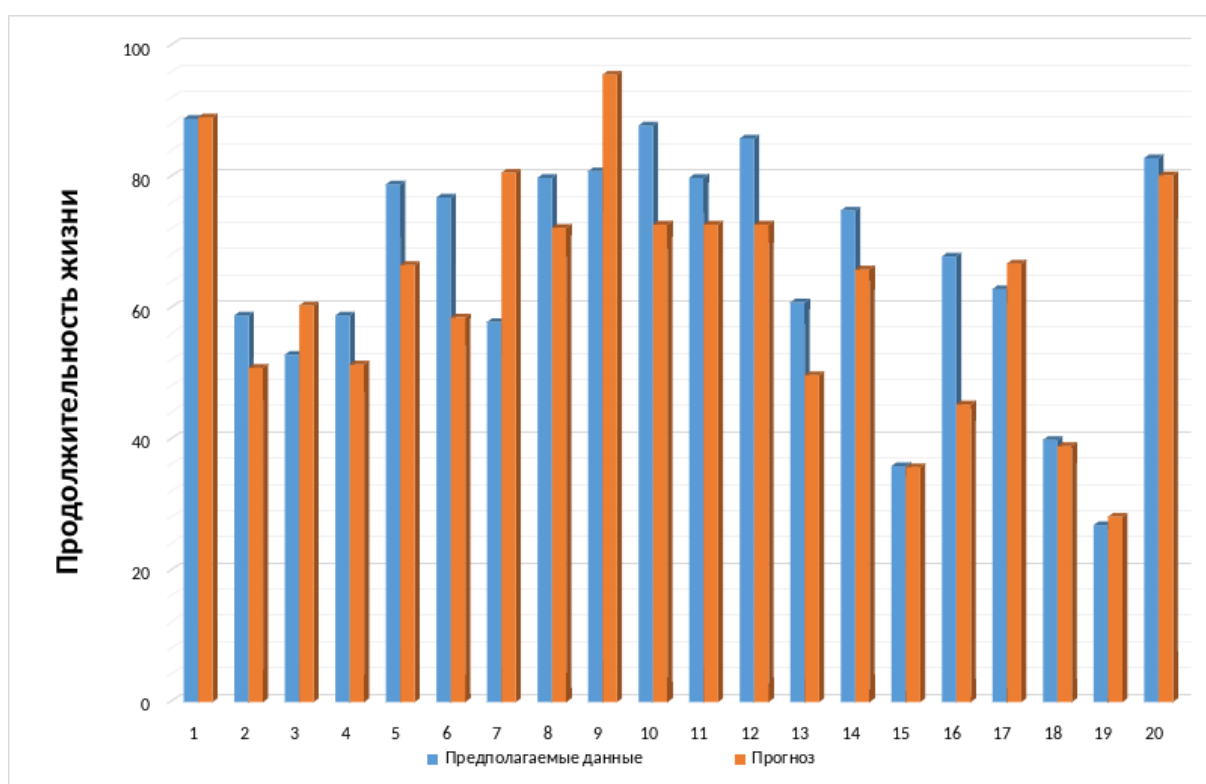
**Человек – техника** – все профессии, связанные с созданием, обслуживанием и эксплуатацией техники. Это такие профессии как: инженер-конструктор, автослесарь, системный администратор и другие.

**Человек – художественный образ** – все профессии, связанные с созданием, копированием, воспроизведением и изучением художественных образов. К этой группе относятся такие профессии как: художник, актер, певец, реставратор, искусствовед и другие.

**Человек – природа** – все профессии, связанные с изучением, охраной и преобразованием природы. К этой группе относятся такие профессии как: ветеринар, садовник, агроном, эколог и другие.

**Человек – знак** – все профессии, связанные с созданием и использованием знаковых систем (цифровых, буквенных, нотных). К этой группе относятся переводчики художественных и технических текстов, аналитики, финансисты и другие.

Для обучения нейронной сети было сформировано множество примеров, основанных на анализе биографии 121 публичных людей. Множество примеров было разбито на обучающее, состоящее из 106 примеров, использованное для обучения сети, и тестирующее множество, состоящее из 20-ти примеров, предназначенное для проверки ее прогностических свойств. Примеры тестирующего множества при обучении нейронной сети не использовались.



**Рисунок 1. Результат тестирования нейронной сети**

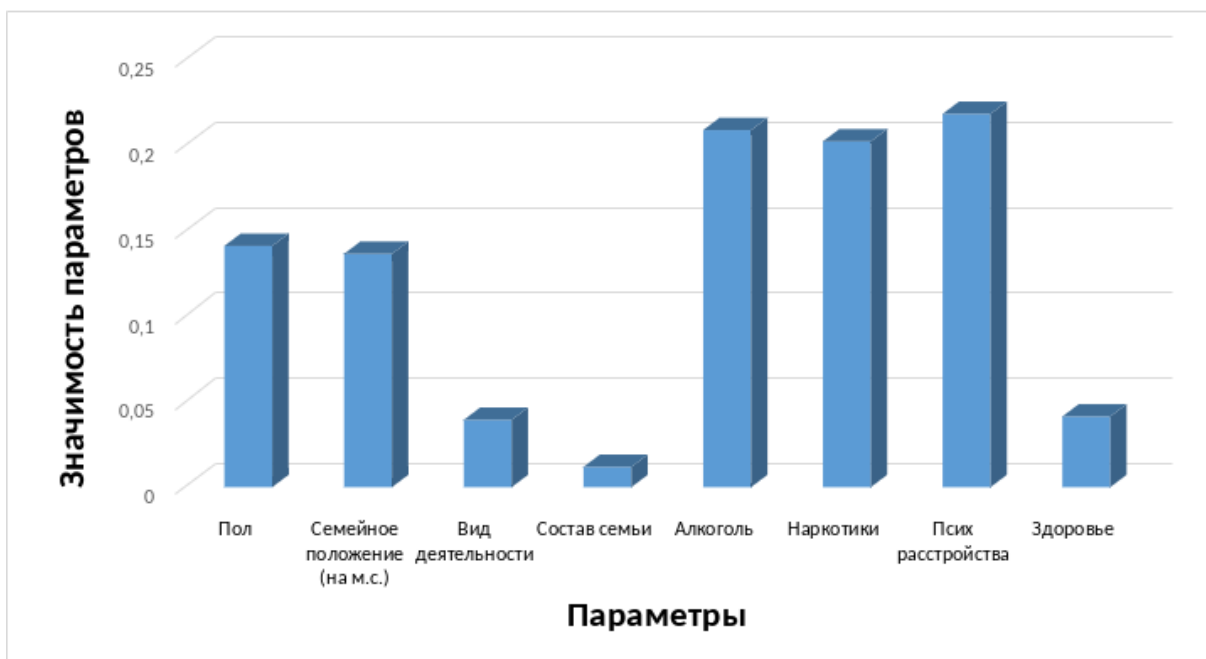
Проектирование, оптимизация, обучение, тестирование нейронной сети, также эксперименты над нейросетевой мат. моделью выполнялись с помощью программы «Нейросимулятор 5».

Из результатов, изображенных на рисунке 1, можно сделать вывод об адекватной работе спроектированной нейронной сети.

Оценка значимости входных параметров отображена на рисунке 2.

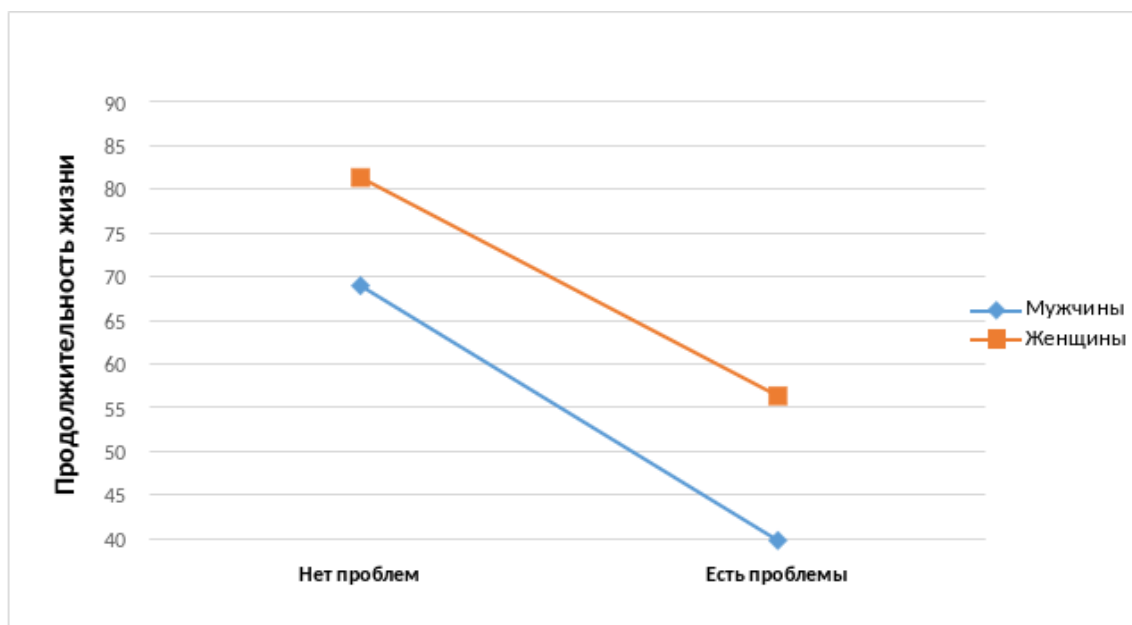
Как видно из рисунка 2, наиболее важные параметры, влияющие на длительность жизни – это психическое здоровье человека, пристрастие к алкоголю и наркотикам. С другой стороны, такой параметр как пол в нашей модели влияет на возраст с наименьшей силой: по-видимому, это результат

известного факта между различием средней продолжительности жизни у мужчин и женщин в пользу женского населения.



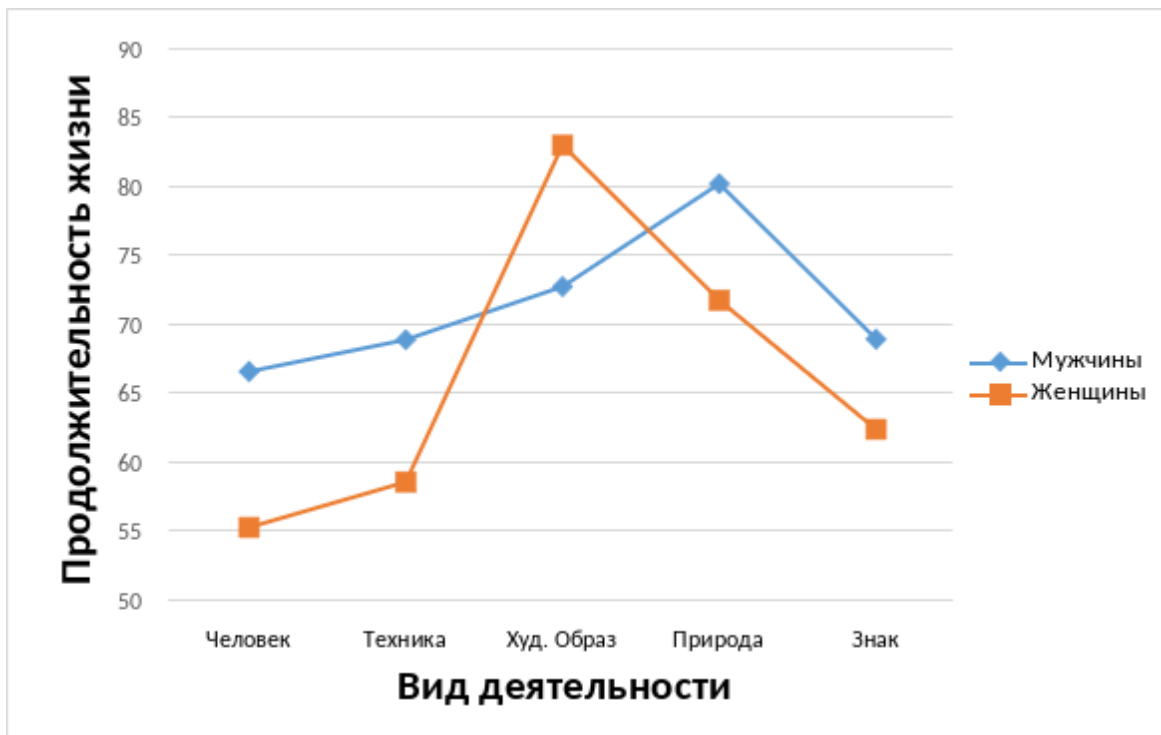
**Рисунок 2. Значимость входных параметров нейросетевой модели**

Влияние проблем с психическим здоровьем на холостого мужчину (женщину), работающего в сфере техники, можно увидеть на рисунке 3.



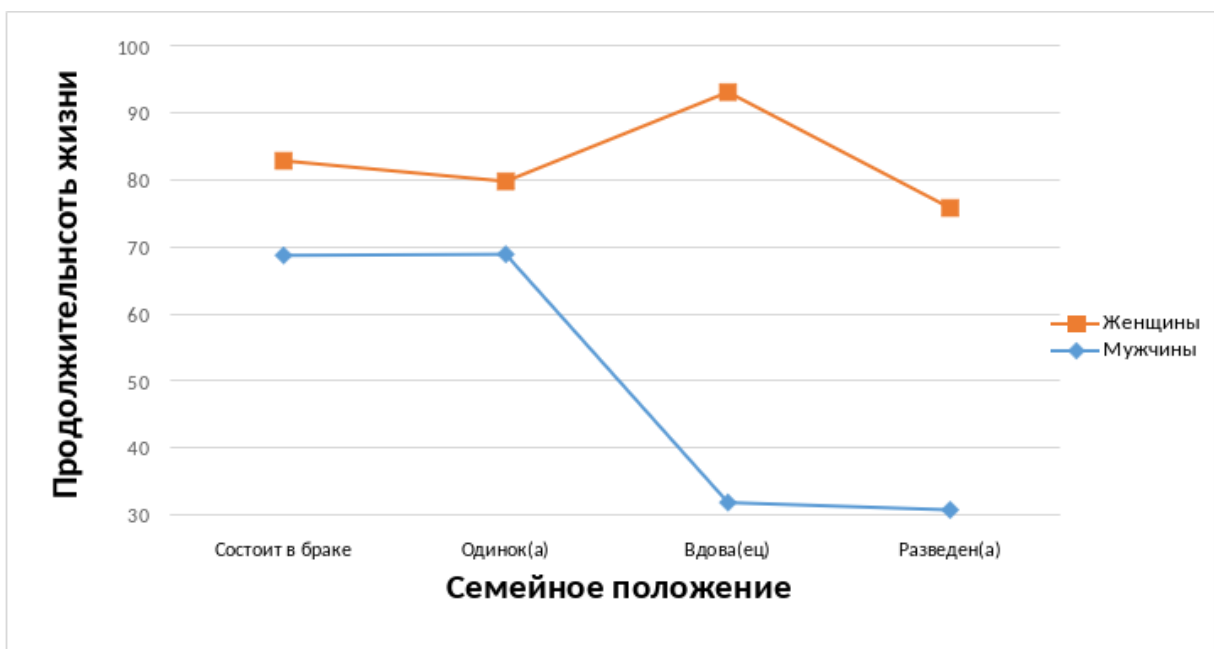
**Рисунок 3. Влияние психического здоровья на продолжительность жизни человека**

С другой стороны влияние выбранной профессии на долгожительство на женатого мужчину (замужнюю женщину) без вредных привычек и проблем со здоровьем.



**Рисунок 4. Влияние профессии на продолжительность жизни**

На рисунке 5 можно видеть, как влияет семейное положение на здоровую женщину (здорового мужчину) без вредных привычек, работающую(его) в творческой сфере:



**Рисунок 5. Влияние семейного положения**

Теперь попробуем спрогнозировать длительность жизни некоторых людей, погибших не своей смертью (несчастный случай, самоубийство и т.д.)

Персона	Пол	Семейное положение (на м.с.)	Вид деятельности	Состав семьи	Алкоголь	Наркотики	Псих расстройств	Здоровье	Возраст смерти
Курт Кобейн	1	4	3	1	0	1	1	1	27

Согласно официальной причине смерти – самоубийство: за три дня до того как обнаружили тело, он ввел себе дозу героина, не совместимую с жизнью, и выстрелил в рот из дробовика. Но существует и другая версия...

Предположим, что музыкант внезапно избавился от пристрастия к наркотикам:

Персона	Пол	Семейное положение (на м.с.)	Вид деятельности	Состав семьи	Алкоголь	Наркотики	Псих расстройств	Здоровье	Возраст смерти
Курт Кобейн	1	4	3	1	0	0	1	1	62,7379

Прогнозируемый возраст смерти увеличился в два раза, ведь употребление наркотиков сильно влияет на продолжительность жизни (рисунок 2).

Предположим, что музыкант избавился от психических расстройств:

Персона	Пол	Семейное положение (на м.с.)	Вид деятельности	Состав семьи	Алкоголь	Наркотики	Псих расстройств	Здоровье	Возраст смерти
Курт Кобейн	1	4	3	1	0	0	0	1	62,0979

Возьмем ещё одного человека, который погиб не своей смертью.

Персона	Пол	Семейное положение (на м.с.)	Вид деятельности	Состав семьи	Алкоголь	Наркотики	Псих расстройств	Здоровье	Возраст смерти
Сергей Есенин	1	4	3	0	1	0	1	0	30

Ниже приведены данные прогнозируемой продолжительности жизни (все входные данные остались прежними).

Персона	Пол	Семейное положение (на м.с.)	Вид деятельности	Состав семьи	Алкоголь	Наркотики	Псих расстройств	Здоровье	Возраст смерти
Сергей Есенин	1	4	3	0	1	0	1	0	31,1995

Таким образом, можно сделать вывод о том, что спроектированная нейронная сеть адекватно работает.

Предположим, что поэт избавился от пристрастия к алкоголю:

Персона	Пол	Семейное положение (на м.с.)	Вид деятельности	Состав семьи	Алкоголь	Наркотики	Псих расстройства	Здоровье	Возраст смерти
Сергей Есенин	1	4	3	0	0	0	1	0	27,9338

Можно сделать вывод, что поэт опять же мог совершить самоубийство. Предположим, что поэт избавился от психических расстройств:

Персона	Пол	Семейное положение (на м.с.)	Вид деятельности	Состав семьи	Алкоголь	Наркотики	Псих расстройства	Здоровье	Возраст смерти
Сергей Есенин	1	4	3	0	0	0	0	0	40,9473

В итоге, можно предположить, что поэт опять же умер от несчастного случая или же совершил самоубийство.

### Заключение

Разработана нейросетевая математическая модель, с помощью которой выполнено прогнозирование продолжительности жизни человека. Исследовано влияние некоторых параметров на этот показатель. Показана возможность применения нейросетевых технологий в сфере медицины.

### Литература

1. Ясницкий Л.Н. Введение в искусственный интеллект. М.: Издательский центр «Академия», 2005. 176 с.
2. Ясницкий Л.Н., Зайцева Н.В., Гусев А.Л., Шур П.З. Нейросетевая модель региона для выбора управляющих воздействий в области обеспечения гигиенической безопасности // Информатика и системы управления. 2011. № 3 (29). С. 51-59.
3. Bo G Eriksson. You're born a copy but die an original. Science Daily April 12, 2010, University of Gothenburg.
4. Ясницкий Л. Н., Гусев А. Л. Астрология под призмой нейронных сетей. МНЖ «Символ Науки», 2(3):17–24, 2017



## НЕЙРОСЕТЕВАЯ СИСТЕМА ПРОГНОЗИРОВАНИЯ РАКА ЛЁГКИХ

*Русаков Вадим Вячеславович*

Пермский государственный национальный исследовательский университет,  
ПМИ. 614990, Россия, г. Пермь, ул. Букирева, 15, Vadim.rusakov.02@mail.ru

В статье представлено описание разработки нейросетевой системы для прогнозирования рака легких среди граждан старшей возрастной категории, а также ход и результаты исследований, проведенных с её помощью.

**Ключевые слова:** искусственный интеллект, нейросетевые технологии, диагностирование, исследование, рак легких.

**Введение.** Рак легкого – это наиболее распространенное злокачественное образование, а также наиболее частая причина летальных исходов среди онкологических патологий, характеризуется достаточно скрытым течением и ранним появлением метастазов. Развивается из желез и слизистой оболочки легочной ткани и бронхов. Согласно статистики, эта онкология поражает мужчин в восемь раз чаще, чем женщин, причем было отмечено, что чем старше возраст, тем гораздо выше уровень заболеваемости.

Система прогнозирования рака позволит людям быстро и без лишних затрат узнать о риске развития рака лёгких и своевременно обратиться к врачу. Диагностирование производится на основе 14 параметров, которые согласно исследованиям [1-2], наиболее сильно влияют на шанс развития рака.

Для построения нейросетевой модели в качестве входных параметров было решено выбрать следующие:

X1 – Пол (1 – М, 2 – Ж)

X2 – Возраст

X3 – Курение (1 – Нет, 2 – Да)

X4 – Желтизна пальцев (1 – Нет, 2 – Да)

X5 – Тревожность (1 – Нет, 2 – Да)

X6 – Наличие хронических заболеваний (1 – Нет, 2 – Да)

X7 – Усталость (1 – Нет, 2 – Да)

X8 – Аллергия (1 – Нет, 2 – Да)

X9 – Свистящее дыхание (1 – Нет, 2 – Да)

X10 – Употребление алкоголя (1 – Нет, 2 – Да)

X11 – Кашель (1 – Нет, 2 – Да)

X12 – Отдышка (1 – Нет, 2 – Да)

X13 – Затруднённое дыхание (1 – Нет, 2 – Да)

X14 – Боли в груди (1 – Нет, 2 – Да)

Всего было собрано порядка 300 данных с ресурса [3] о пациентах, у большинства из которых было выявлено наличие рака легкого.

Всё множество примеров было разделено на обучающее, валидирующее и тестирующее в соотношении 70 : 20 : 10.

В процессе обучения нейросетевой модели была выполнена очистка исходного множества от выбросов [11].

Проектирование, обучение, тестирование нейронной сети выполнялись с помощью программы «Nsim5-10s» [8] по методике [9, 10].

Таким образом, спроектированная нейронная сеть представляла собой персептрон, который имеет 14 входных нейронов, 1 выходной и 1 скрытый слой с 4 нейронами.

В качестве активационных функций нейронов скрытого слоя и выходного слоя использовался тангенс гиперболический, а в качестве алгоритма обучения – алгоритм упругого распространения.

Для оценки корректной работы спроектированной нейронной сети использовалось тестирующее множество, состоящее из 30 примеров. Средняя относительная ошибка тестирования составила 6.5%, что является приемлемым результатом. На рисунке 1 представлена гистограмма, демонстрирующая разницу между фактическим и прогнозируемым нейросетью наличием рака лёгких.

Из результатов, изображенных на рисунке 1, можно сделать вывод об адекватной работе спроектированной нейронной сети.

Оценка значимости входных параметров была выполнена с помощью программы «Nsim5-10s» результат отображен на рисунке 2.

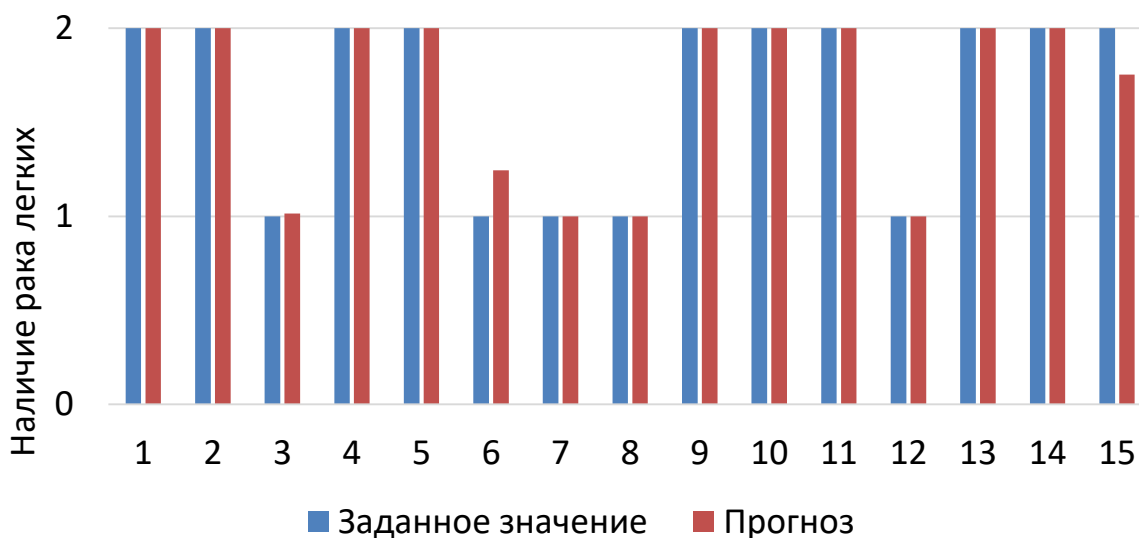
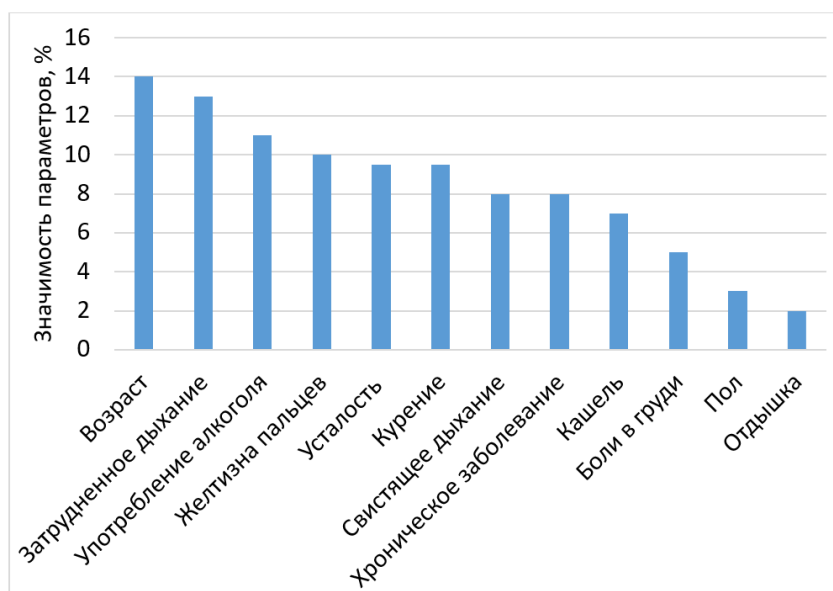


Рисунок 1. Результат тестирования нейронной сети

Из рисунка видно, что наиболее значимым параметром является возраст человека, что подтверждает похожая нейронная сеть [5]. Также, не менее важными параметрами являются: затрудненное дыхание, желтизна пальцев, употребление алкоголя и усталость, а вот курение находится лишь на 6 месте по значимости.



**Рисунок 2. Значимость входных параметров нейросетевой модели**

Далее были проведены исследования влияния некоторых входных параметров на риск развития рака легкого. Исследование производилось с помощью метода «замораживания» [6-7, 9], суть которого заключается в варьировании значения одного параметра и фиксировании значений всех других параметров. Для исследования были выбраны 2 пациента со схожими параметрами, но разных полов. Параметры пациентов приведены в таблице 1. В процессе исследования фиксировались значения, кроме анализируемого параметра и параметра, характеризующего возраст пациента, для определения зависимости влияния параметра на риск развития рака с возрастом.

**Таблица 1**

**Параметры исследуемого пациента**

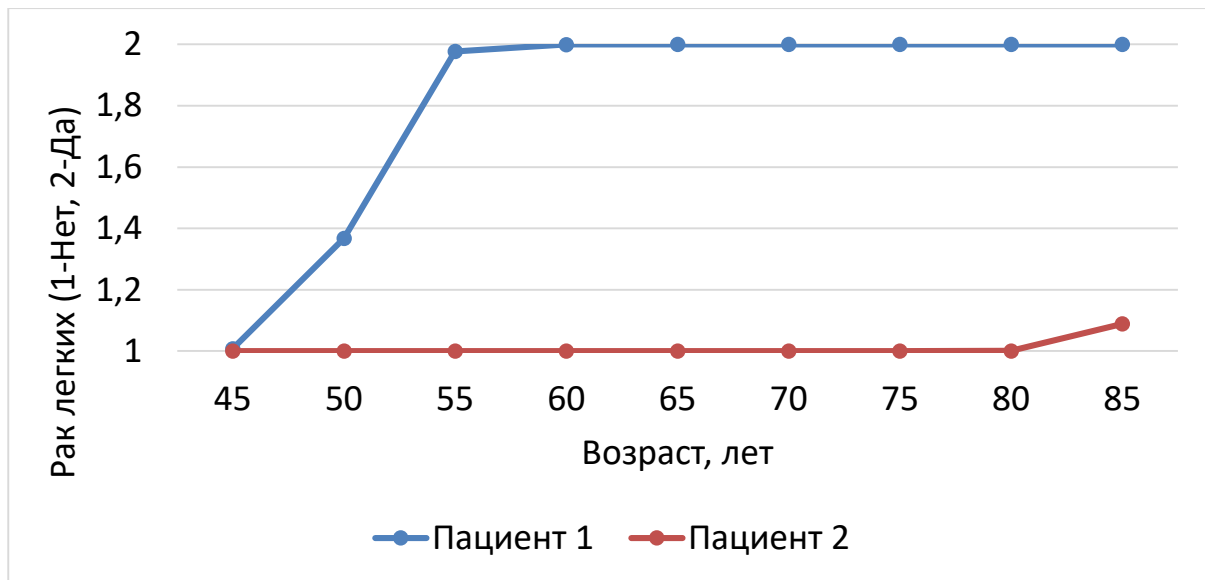
Пол	Возраст	Затрудненное дыхание (1 – Нет, 2 – Да)	Желтизна пальцев (1 – Нет, 2 – Да)	Употребление алкоголя (1 – Нет, 2 – Да)	Усталость (1 – Нет, 2 – Да)	Курение (1 – Нет, 2 – Да)
М	61	1	1	1	2	2
Ж	70	1	1	1	2	2

На рисунке 3 показан график влияния входного параметра “Возраст”. Исследование проводится на старшей возрастной категории, т.к. в основном она находится в зоне риска.

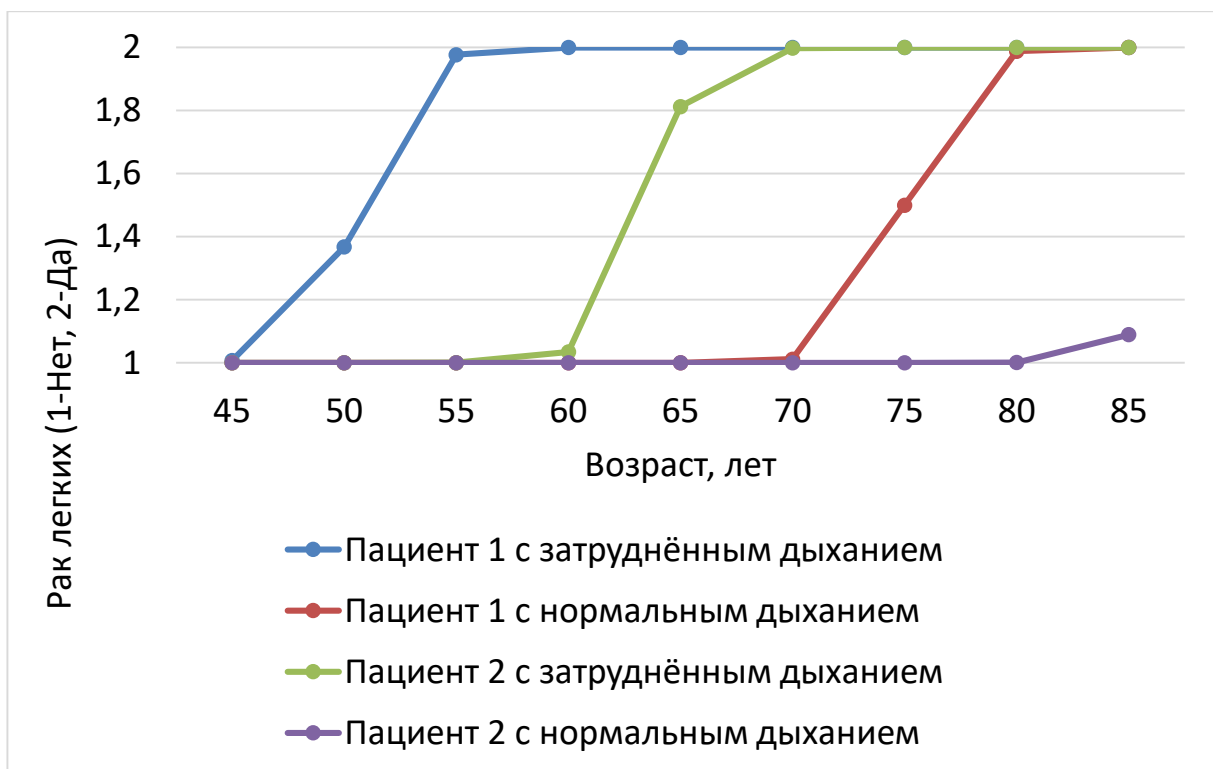
Как следует из графика, вероятность рака легких резко повышается в 55 лет.

Также, у мужчин шанс развития рака легких выше, чем у женщин.

Из рисунка 4 видно, что затрудненное дыхание значительно влияет на шанс развития рака легких. Такой же график получается, если исследовать параметр “Желтизна пальцев”



**Рисунок 3. Влияние возраста на развитие рака легких**



**Рисунок 4. Влияние затрудненного дыхания на развитие рака легких**

Рисунок 5 подтверждает высказывания ученых о существовании причинно-следственной связи влияния алкоголя на раковые клетки[3]. Аналогичные графики получаются при исследовании таких параметров, как “Усталость” и “Курение”.

**Заключение.** Построена система оценивания риска развития рака легких с помощью нейросети. Спроектированная нейросетевая модель обрабатывает параметры, характеризующие основные причины развития рака. Проведена оценка значимостей входных параметров, кроме того выполнена оценка ошибки обучения и тестирования, которая составила 6.5%. Также, из

проведённых исследований была установлена вероятность развития рака для двух пациентов разных полов.

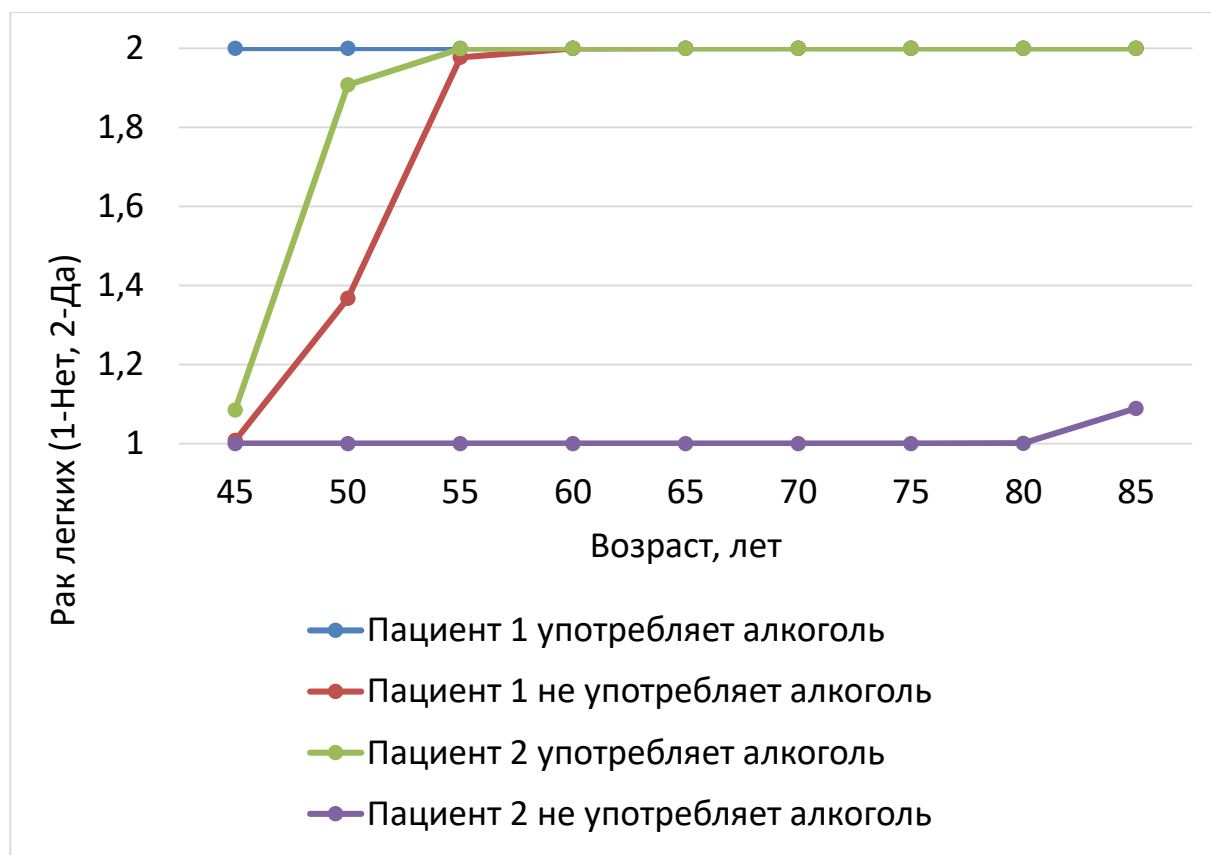


Рисунок 5. Влияние алкоголя на развитие рака легких

### Библиографический список

1. oncology.ru– всё про онкологию. [Электронный ресурс]. Режим доступа: [http://oncology.ru/specialist/library/lung\\_cancer/epidemiology\\_symptoms/](http://oncology.ru/specialist/library/lung_cancer/epidemiology_symptoms/)
2. nhs.uk– информация о раке легких. [Электронный ресурс]. Режим доступа: <https://www.nhs.uk/conditions/lung-cancer/>
3. kaggle.com – датасет на тему “Рак легкого”. [Электронный ресурс]. Режим доступа: <https://www.kaggle.com/datasets/mysarahmadbhat/lung-cancer>
4. ookod.ru – влияние алкоголя на риск развития рака. [Электронный ресурс]. Режим доступа: <https://ookod.ru/left/sanbilyuten/vliyanie-alkogolya-na-risk-razvitiya-raka>
5. <https://core.ac.uk/download/pdf/237182451.pdf> Nasser IM and Abu-Naser SS 2019 Lung Cancer Detection Using Artificial Neural Network.
6. Ясницкий Л.Н., Корнилова В.А. Нейросетевая система оценки риска развития инсульта // Искусственный интеллект в решении актуальных социальных и экономических проблем XXI века. 2021. С. 86.
7. Ясницкий Л.Н., Сичинава З.И. Нейросетевые алгоритмы анализа поведения респондентов // Нейрокомпьютеры: разработка, применение. 2011. № 10. С. 59-64.

8. Черепанов Ф.М., Ясницкий Л.Н. Нейросимулятор 4.0. Свидетельство о регистрации программы для ЭВМ RUS 2014612546. Заявка № 2014610341 от 15.01.2014.

9. Yasnitsky L.N., Dumler A., Cherepanov F.M. The capabilities of artificial intelligence to simulate the emergence and development of diseases, optimize prevention and treatment thereof, and identify new medical knowledge // Journal of Pharmaceutical Sciences and Research. 2018. Т. 10. № 9. С. 2192-2200.

10. Yasnitsky L.N., Dumler A.A., Cherepanov F.M. Dynamic artificial neural networks as basis for medicine revolution // Advances in Intelligent Systems and Computing. 2018. Т. 850. С. 351-358. 5.

11. Yasnitsky L.N. Algorithm for searching and analyzing abnormal observations of statistical information based on the Arnold – Kolmogorov – Hecht-Nielsen theorem // International Journal of Advanced Trends in Computer Science and Engineering. 2020. 9(2). Pp. 1814-1819. DOI: 10.30534/ijatcse/2020/139922020 <http://www.warse.org/IJATCSE/static/pdf/file/ijatcse139922020.pdf>

## **NEURONETWORK SYSTEM FOR PREDICTING LUNG CANCER**

*Rusakov Vadim V.*

Perm State University

Str. Bukireva, 15, Perm, Russia, 614990, Vadim.rusakov.02@mail.ru

The article presents a description of the development of a neural network system for the prediction of lung cancer among senior citizens, as well as the progress and results of research conducted with it's help.

**Keywords:** artificial intelligence, neural network technologies, diagnostics, research, lung cancer.

УДК 004.032.26

## **НЕЙРОСЕТЕВАЯ СИСТЕМА ОЦЕНКИ РИСКА ОБНАРУЖЕНИЯ ОПУХОЛИ МОЛОЧНОЙ ЖЕЛЕЗЫ**

*Соболева Дарья Александровна*

Пермский государственный национальный исследовательский университет, КМБ. 614990, Россия, г. Пермь, ул. Букирева, 15. daryasobolewa@yandex.ru

В статье представлено описание разработанной нейросетевой системы для прогнозирования вероятности обнаружения опухоли молочной железы. Система позволяет провести быструю диагностику данного заболевания на основе малого количества входных параметров.

**Ключевые слова:** нейросетевые технологии, опухоль молочной железы, прогнозирование, закономерности.

**Введение.** Новообразование (опухоль) молочной железы у женщин — это патологическое разрастание ткани, состоящей из качественно изменившихся клеток молочной железы. Отличительная особенность рака груди — быстрый агрессивный рост опухоли, проникновение в соседние ткани, метастазирование в отдалённые участки организма.

Причиной развития рака молочной железы чаще всего являются гормональные изменения в организме. Предрасполагают к развитию заболевания плохая наследственность, вредные факторы окружающей среды, а также перенесенные воспалительные заболевания груди.

На начальных стадиях развития онкологического заболевания при обследовании молочной железы прощупывается небольшое уплотнение. Других симптомов, как правило, не наблюдается. Из-за отсутствия симптомов на ранних стадиях, рак груди чаще всего выявляют в запущенной форме [1].

Предупреждение развития опухоли молочной железы представляется одной из важных клинических задач текущего времени, которое необходимо решать в условиях высокой загруженности специализированных стационаров и дефицита специалистов.

Целью данной работы является разработка модели, способной достаточно быстро и точно выявлять риск обнаружения опухоли молочной железы на основе малого количества входных параметров

**Методика прогнозирования.** В современной медицине принято считать причиной развития опухоли молочной железы являются гормональные изменения в организме. Предрасполагают к развитию заболевания плохая наследственность, вредные факторы окружающей среды, а также перенесенные воспалительные заболевания груди [1]. К основным факторам риска развития опухоли молочной железы, которые можно рассматривать относительно всех видов данного заболевания, причисляют возраст, не кормление грудью, неоднократные аборт. Именно эти факторы будут использоваться для отбора признаков при построении нейронной сети.

Набор данных для обучения нейронной сети был взят вручную с базы данных Пермского краевого онкологического диспансера.

Перед переходом к проектированию и обучению модели, была выполнена очистка исходного множества от выбросов по методике [7], а также произведено преобразование категориальных признаков в числовой формат. Было выявлено 15 «выбросов», все они были удалены из общего множества.

Таким образом, для проектирования системы были использованы следующие входные параметры, оценивающие состояние диагностируемого:

X1 – возраст, лет,

X2 – место проживания (1 – город, 0 – село),

X3 – род деятельности (1 – работа по найму, 2 – самозанятый),

X4 – число детей,

X5 – число беременностей,

X6 – число абортов,  
X7 – факт кормления грудью (0 – нет, 1 – да),  
X8 – факт наступления менопаузы (0 – нет, иначе – во сколько лет),  
X9 – рост, см,  
X10 – вес, кг,  
X11 – ИМТ, кг/м<sup>2</sup>,

Целевая переменная: D1 – обнаружена ли опухоль молочной железы, – принимает значения от 0 до 1, соответствующие вероятности развития опухоли молочной железы для диагностированного, представленной в десятичной форме.

Объем итогового множества для обучения включил в себя 96 примеров. Анализ распределения целевой переменной исходного множества – факт обнаружения опухоли молочной железы – показал, что исходное множество является сбалансированным: количество диагностируемых с опухолью молочной железы составляет около 50% от общего количества пациентов.

Для обучения модели множество было разделено на обучающее, тестовое и подтверждающее в соотношении 80%, 10% и 10% соответственно.

Проектирование, обучение, тестирование нейронной сети выполнялись с помощью программы «Нейросимулятор 5» [2] по методике [3]. В результате оптимизации спроектированная нейронная сеть представляла собой персептрон, который имеет одиннадцать входных нейронов, один выходной. В качестве активационной функции выходного слоя была применена биполярная.

Для оценки корректной работы спроектированной нейронной сети использовалось тестирующее множество, состоящее из 7 примеров. Средняя относительная ошибка тестирования составила 10,7%, что можно считать приемлемым результатом для медицинских исследований. На рисунке 1 показана гистограмма, демонстрирующая разницу между ожидаемым и полученным нейросетью результатом.

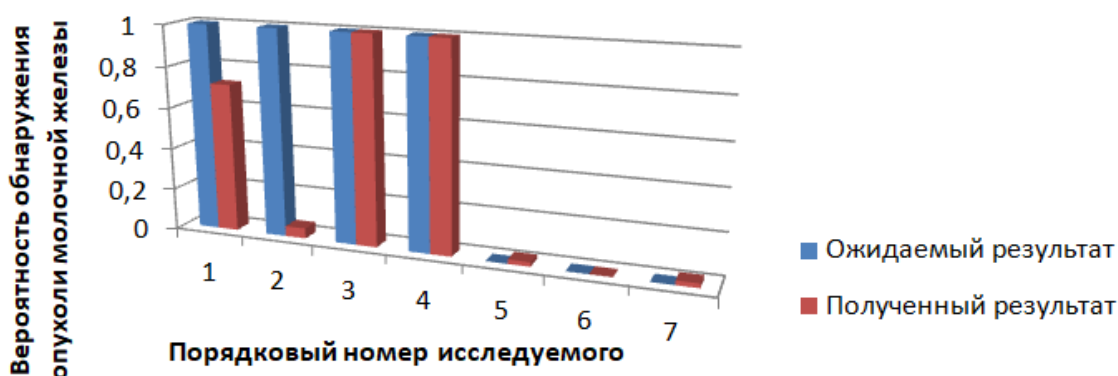
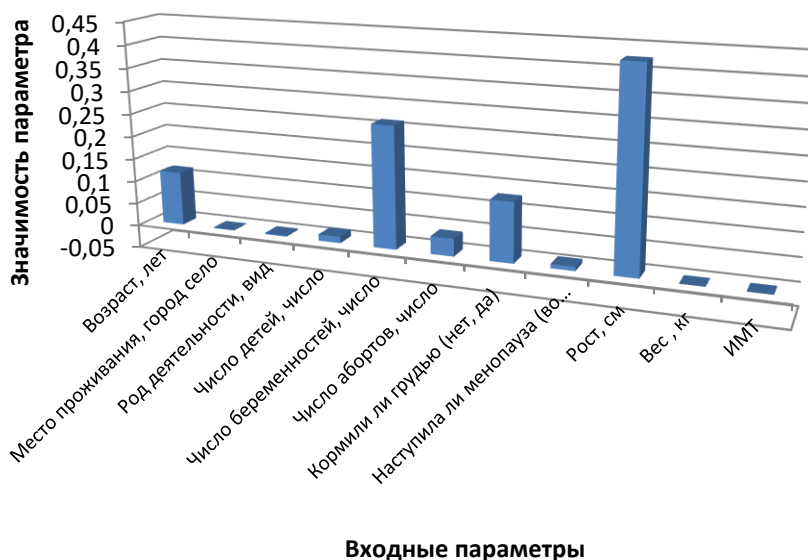


Рисунок 1. Результат тестирования нейронной сети

**Экспериментальная часть.** Оценка значимости входных параметров была рассчитана по методике [5]. Построенная данным способом гистограмма изображена на рисунке 2.





**Рисунок 2. Значимость входных параметров нейронной сети**

Как видно из рисунка 2, наиболее значимыми являются рост, число беременностей, факт кормления грудью и возраст.

Далее были проведены исследования влияния каждого входного параметра на риск развития опухоли молочной железы. Для этого был использован метод «замораживания» [5], суть которого заключается в варьировании значения одного параметра и фиксации значения всех остальных. Параметры пациента, взятого для исследований, приведены в таблице 1. В процессе исследования фиксировались значения, кроме анализируемого параметра и параметра. Результаты исследования отображены на рисунках 3 – 12.

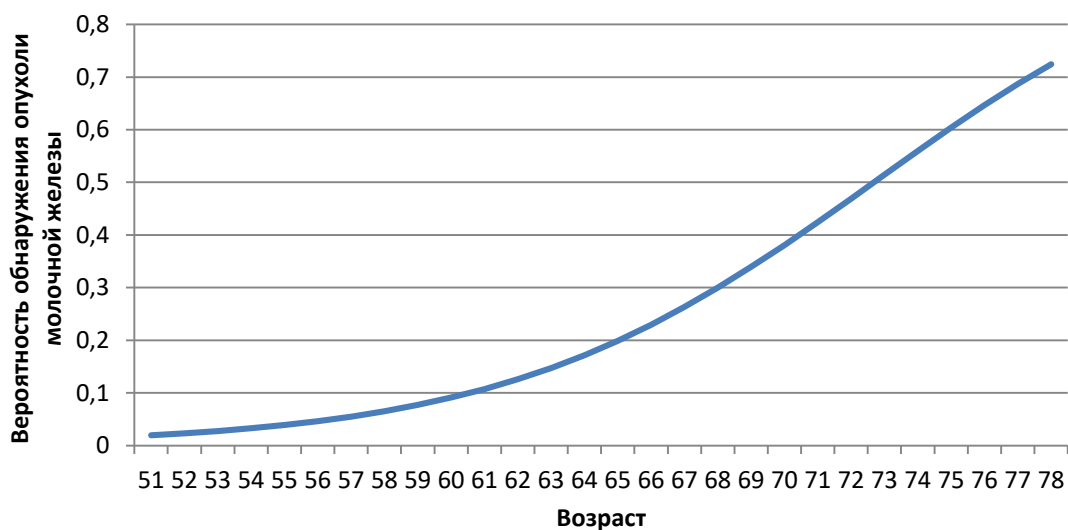
**Таблица 1**

**Параметры исследуемого пациента**

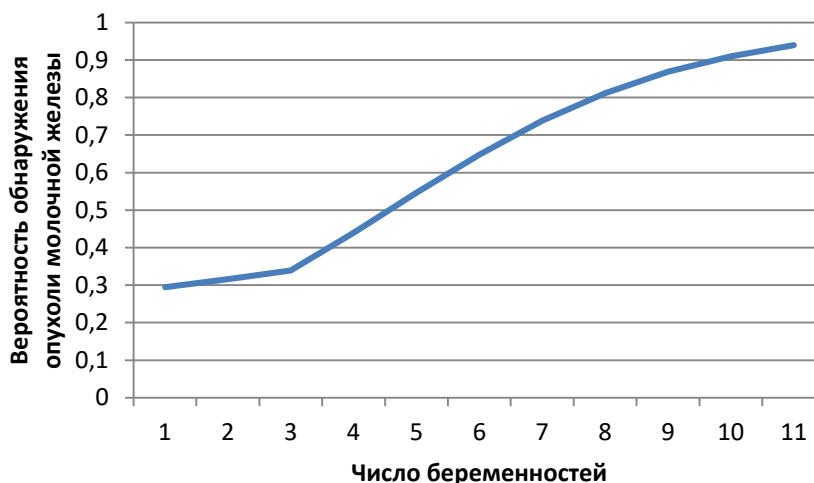
Возраст, лет	Место проживания	Род деятельности	Число детей	Число беременностей	Число абортов	Факт кормления грудью	Возраст наступления менопаузы	Рост	Вес	ИМТ
69	Село	Самозанятый	2	2	1	да	48	166	65	23,6

На рисунке 3 показан график влияния входного параметра «Возраст». Как следует из графика, вероятность обнаружения опухоли молочной железы с увеличением возраста растет.

Из рисунка 4 заметим, что при увеличении числа беременностей растет шанс обнаружения опухоли молочной железы, что можно объяснить стрессом для женского организма, так же как и факт аборта, поскольку он уже включен в этот параметр.

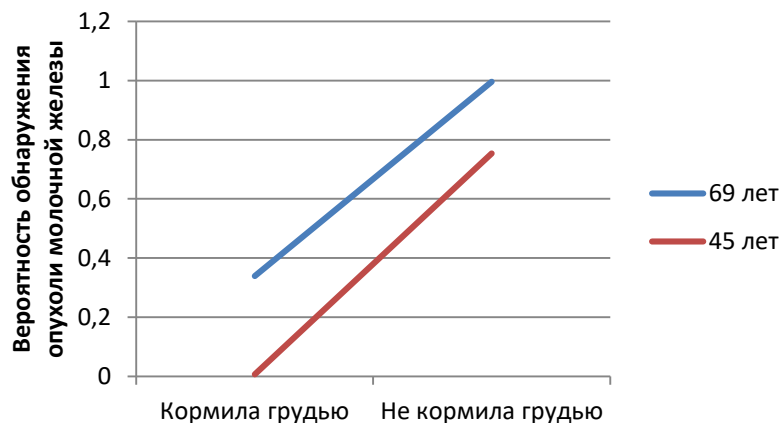


**Рисунок 3. Влияние возраста**



**Рисунок 4. Влияния числа беременностей**

Из рисунка 5 видно, что факт кормления грудью благотворно сказывается на женском организме. Было взято 2 возраста для наглядности этого факта.



**Рисунок 5. Влияния факта кормления грудью**

Из рисунка 6 видно, что люди низкого роста более подвержены опухоли молочной железы.

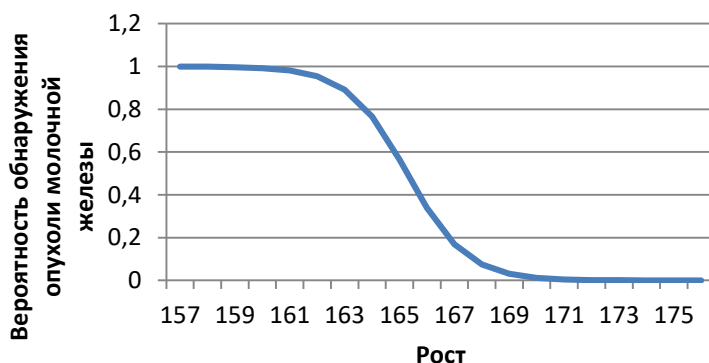


Рисунок 6. Влияния роста

**Заключение.** Построена система оценивания риска обнаружения опухоли молочной железы с помощью нейросети. Спроектированная нейросетевая модель обрабатывает параметры, характеризующие основные факторы риска наступления данной патологии. Проведены результаты оценки значимости входных параметров на возникновение опухоли молочной железы, выполнена оценка ошибки обучения и тестирования. Поскольку число входных данных является небольшим, а также не требует особых исследований, то набор входных данных может формироваться из сервиса «Гос. Услуги» [4] с уведомлением пользователя.

### Библиографический список

1. [https://www.dkb74.ru/vidy\\_pomoshhi/napravleniya/onkologiya/zabolevaniya/novoobrazovaniya\\_molochnoj\\_zhelezy](https://www.dkb74.ru/vidy_pomoshhi/napravleniya/onkologiya/zabolevaniya/novoobrazovaniya_molochnoj_zhelezy).
2. Черепанов Ф.М., Ясницкий Л.Н. Нейросимулятор 4.0. Свидетельство о регистрации программы для ЭВМ RUS 2014612546. Заявка № 2014610341 от 15.01.2014.
3. Ясницкий Л.Н. Введение в искусственный интеллект. М.: Издательский центр «Академия», 2005. – 176 с.
4. Портал государственных услуг Российской Федерации [Электронный ресурс] [Режим доступа: <https://www.gosuslugi.ru/>] [Проверено: 6.04.2021].
5. Ясницкий Л.Н. Интеллектуальные системы: учебник. М.: Лаборатория знаний, 2016. 221 с.
6. Yasnitsky L.N. Algorithm for searching and analyzing abnormal observations of statistical information based on the Arnold – Kolmogorov – Hecht-Nielsen theorem // International Journal of Advanced Trends in Computer Science and Engineering. 2020. 9(2). Pp. 1814-1819. DOI: 10.30534/ijatcse/2020/139922020

# NEURAL NETWORK SYSTEM FOR ASSESSING THE RISK OF DETECTING BREAST TUMOR

*Soboleva Daria A.*

Perm State University,  
614990, Russia, Perm, Bukireva str., 15

The article describes the developed neural network system for predicting the probability of detecting a breast tumor. The system allows a quick diagnosis of this disease based on a small number of input parameters.

**Keywords:** neural network technologies, breast tumor, prediction, patterns.

УДК 004.032.26

## НЕЙРОСЕТЕВАЯ СИСТЕМА ПРОГНОЗИРОВАНИЯ ПОКАЗАТЕЛЯ ОБЩЕГО КОЭФФИЦИЕНТА СМЕРТНОСТИ В ПЕРМСКОМ КРАЕ

*Филатова Полина Денисовна, Гусев Андрей Леонидович*

Пермский государственный национальный исследовательский университет,  
ПМИ. 614990, Россия, г. Пермь, ул. Букирева, 15

В статье представлено описание разработки нейросетевой системы для прогнозирования общего коэффициента смертности. Система позволяет с достаточной точностью предсказать общий коэффициент смертности на основании социальных и экономических данных по Пермскому краю. С помощью разработанной интеллектуальной системы проведено исследование предметной области, выявлены закономерности, имеющие практической значение.

**Ключевые слова:** искусственный интеллект, нейросетевые технологии, прогнозирование, общий коэффициент смертности.

**Введение.** В настоящее время прогнозирование все чаще применяют в различных сферах жизни. Благодаря прогнозам можно заранее узнать о вероятных событиях и попытаться их изменить в нужном направлении.

Показатели населения, а точнее – это данные о рождаемости и смертности, всегда были и остаются важным параметром оценки состояния жизни в некотором населенном пункте. Именно эти показатели являются ориентирами в дальнейших действиях по улучшению различных сфер. Актуальность данной темы в том, что данные параметров населения определяют качество жизни и служат критерием для будущих изменений в разных областях.

Основная цель настоящей работы заключается в сборе множества статистических данных о некоторых показателях социально – экономической сферы, а также создание и обучение нейросетевой модели на этих данных. Конечным результатом работы должна стать нейросетевая система, разработанная средствами Нейросимулятора [1], а также нейросетевая система, разработанная средствами библиотеки с открытым исходным кодом Keras [2] на языке программирования – Python.

Для создания нейросетевой системы были выбраны следующие параметры: X1 – численность детей в возрасте 5-18 лет в муниципальном образовании; X2 – численность воспитанников, посещающих организации, осуществляющие образовательную деятельность по образовательным программам дошкольного образования, присмотр и уход за детьми; X3 – дошкольные образовательные учреждения; X4 – число мест в организациях, осуществляющих образовательную деятельность по образовательным программам дошкольного образования, присмотр и уход за детьми; X5 – доходы местного бюджета, фактически исполненные; X6 общий объем всех продовольственных товаров, реализованных в границах муниципального района, в денежном выражении за финансовый год; X7 – среднегодовая численность постоянного населения; X8 – расходы местного бюджета, фактически исполненные; X9 – число муниципальных спортивных сооружений; X10 – число спортивных сооружений – всего; X11 – общая площадь жилых помещений; X12 – оценка численности населения на 1 января текущего года; X13 – доля налоговых и неналоговых доходов местного бюджета (за исключением поступлений налоговых доходов по дополнительным нормативам отчислений) в общем объеме собственных доходов бюджета муниципального образования; X14 – одиночное протяжение уличной водопроводной сети (до 2008 г. – км), метр. Выходной параметр – общий коэффициент смертности.

Для сбора данных использовался Паспорт муниципального образования с сайта Федеральной службы государственной статистики [3]. Таким образом, объем итогового множества включает в себя 126 примеров. Данное множество было разделено на обучающее и тестирующее в соотношении 80% и 20% соответственно.

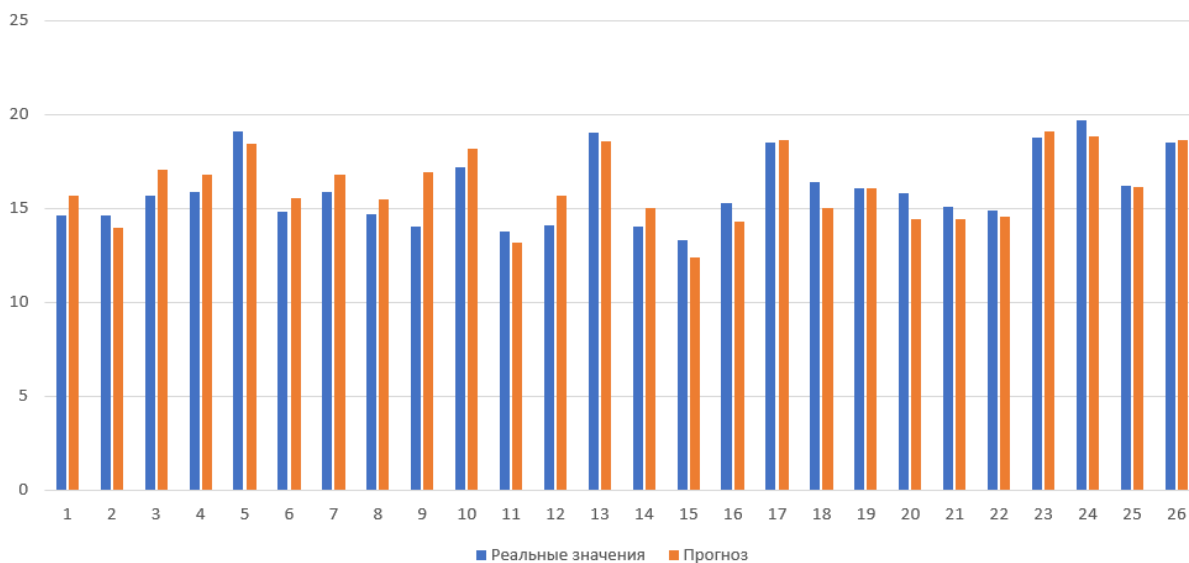
Проектирование, обучение, тестирование нейронной сети выполнялись с помощью программы «Нейросимулятор 5» [1] по методике [4]. В результате оптимизации спроектированная нейронная сеть представляла собой персептрон, который имеет 14 входных нейронов, один выходной параметр, один скрытый слой с 6 нейронами, функцией активации является гиперболический тангенс. При использовании библиотеки Keras, персептрон имеет 14 входных нейронов, один выходной нейрон, один скрытый слой с 8 нейронами, функция активации для входного и скрытого слоев – *selu*, для выходного слоя – линейная функция.

В случае использования Keras рассматривались следующие функции активации: *tanh* (гиперболический тангенс), *relu* (Rectified Linear Unit), *selu* (Scaled Exponential Linear Unit). В ходе построения модели самые лучшие результаты показывала функция *selu*. Также менялось количество итераций, оп-

тимальным оказалось количество в 200 итераций. Оптимизатором для нейросети был выбран алгоритм Адам, так как он показал наилучшие результаты.

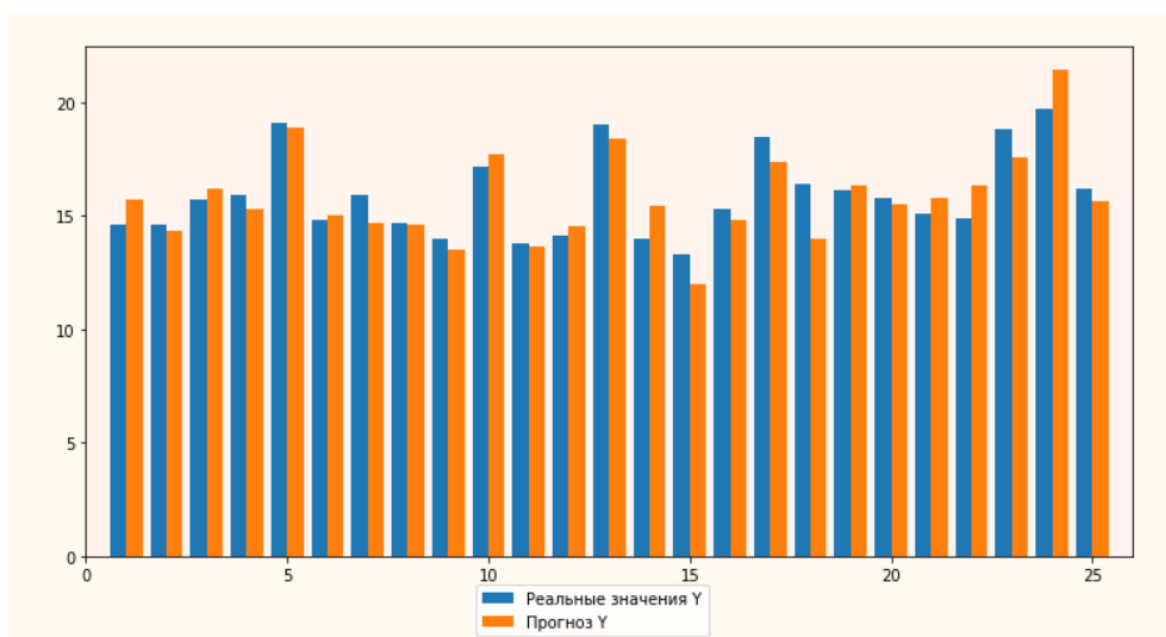
Для оценки корректной работы спроектированных нейронных сетей использовалось тестирующее множество, состоящее из 26 примеров. Коэффициент детерминации с Нейросимулятором – 0,73, с Keras – 0,7. Результаты можно считать приемлемыми.

На рисунке 1 представлена разница между реальными и прогнозными значениями выходного параметра, полученная с помощью Нейросимулятора.



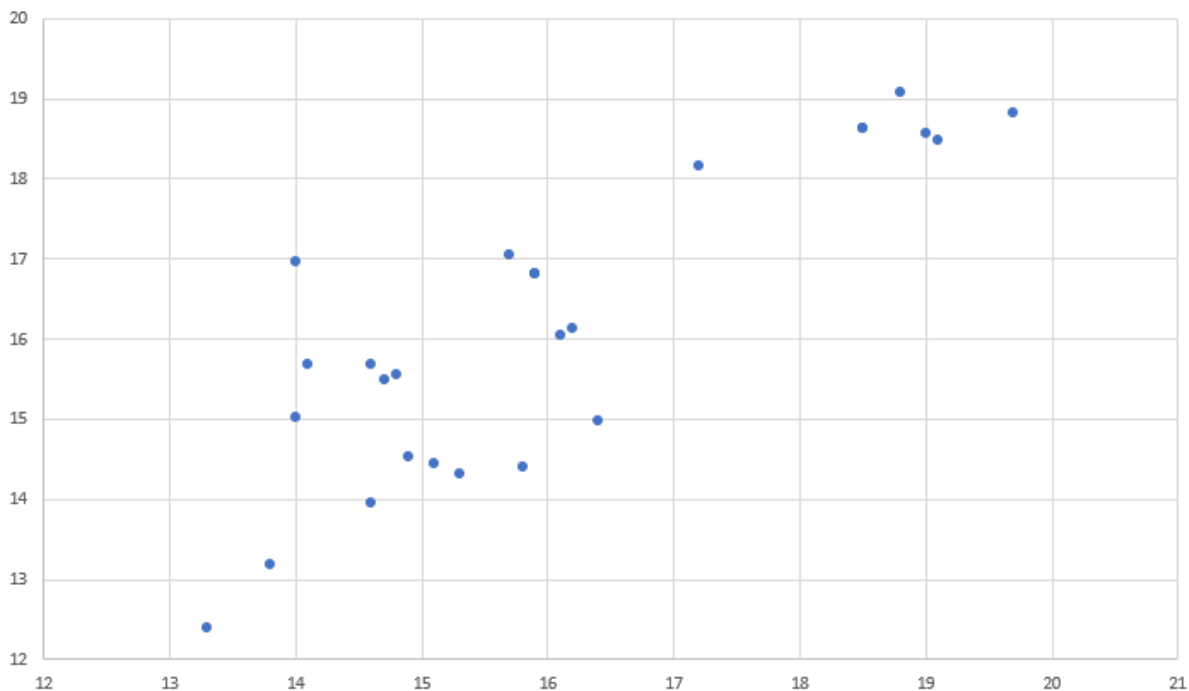
**Рисунок 1. Разница реальных и прогнозных значений (Нейросимулятор)**

На рисунке 2 представлена разница между реальными и прогнозными значениями выходного параметра, полученная с помощью Keras.



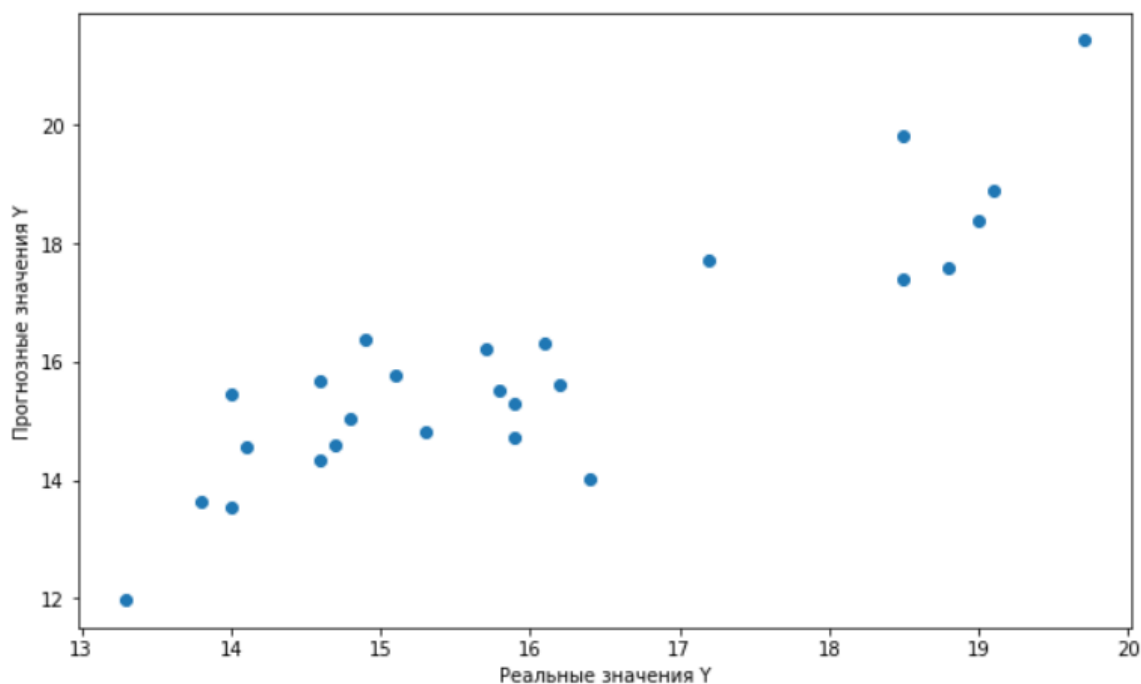
**Рисунок 2. Разница реальных и прогнозных значений (Keras)**

На рисунке 3 представлена диаграмма рассеяния между реальными и прогнозными значениями выходного параметра, полученные с помощью нейросетей, построенных средствами Нейросимулятора.



**Рисунок 3. Диаграмма рассеяния между реальными и прогнозными значениями выходного параметра (Нейросимулятор)**

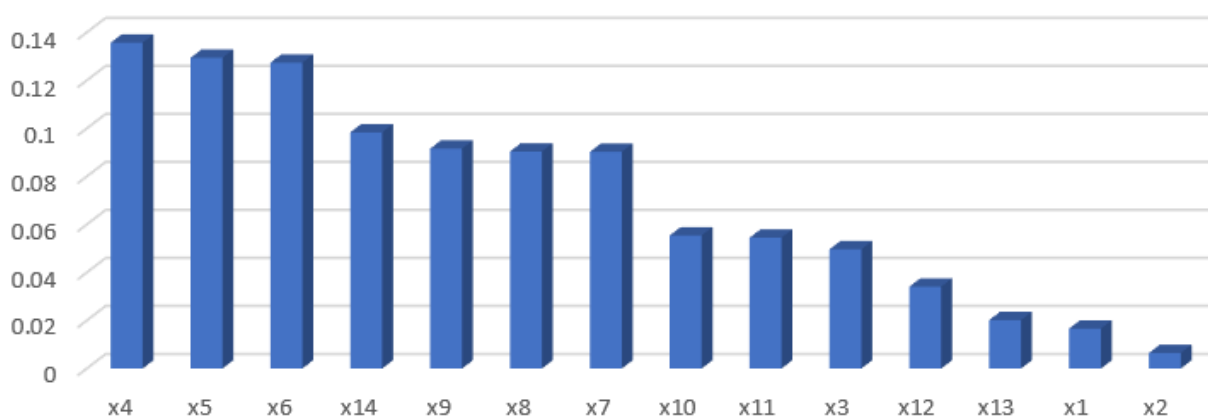
На рисунке 4 представлена диаграмма рассеяния между реальными и прогнозными значениями выходного параметра, полученные с помощью нейросетей, построенных средствами Keras.



**Рисунок 4. Диаграмма рассеяния между реальными и прогнозными значениями выходного параметра (Keras)**

Из результатов можно сделать вывод об адекватной работе этих двух нейронных сетей.

Оценка значимости входных параметров была выполнена с помощью программы «Нейросимулятор 5», результат отображен на рисунке 5.



**Рисунок 5. Значимость входных параметров нейросетевой модели**

Как видно из рисунка 5, наиболее значимыми являются число мест в организациях, осуществляющих образовательную деятельность по образовательным программам дошкольного образования, присмотр и уход за детьми. Также сильно значимыми являются параметры X5 – доходы местного бюджета, фактически исполненные, X6 – общий объем всех продовольственных товаров, реализованных в границах муниципального района, в денежном выражении за финансовый год.

Далее было проведено исследование полученных зависимостей между наиболее значимыми входными параметрами и результатом. Исследование производилось с помощью метода «замораживания», суть которого заключается в варьировании значения одного параметра и фиксации значений всех других параметров.

В таблице 1 представлено изменение значений общего коэффициента смертности от значений числа мест в организациях, осуществляющих образовательную деятельность по образовательным программам дошкольного образования, присмотр и уход за детьми (Y1 – общий коэффициент смертности, x4 – число мест в организациях).

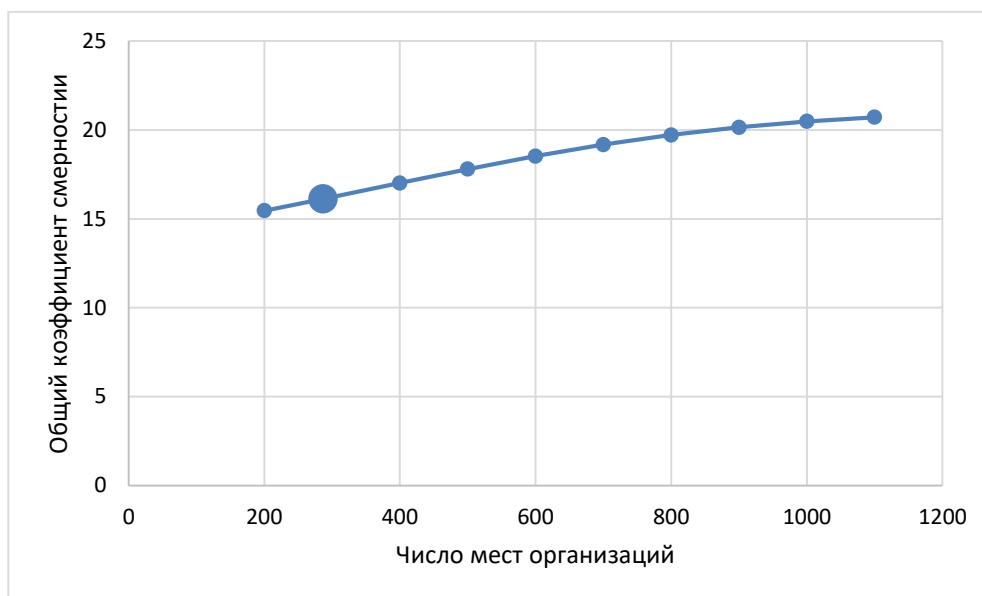
**Таблица 1**

**Изменение значений общего коэффициента смертности от числа мест в организациях**

X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12	X13	X14	Y1
12282	474531	473886	200	120833	650	772	897	12192	2407	19.9	479862.6	47	44	15.4553
12282	474531	473886	286.8	120833	650	772	897	12192	2407	19.9	479862.6	47	44	16.1197
12282	474531	473886	400	120833	650	772	897	12192	2407	19.9	479862.6	47	44	17.0135
12282	474531	473886	500	120833	650	772	897	12192	2407	19.9	479862.6	47	44	17.7948
12282	474531	473886	600	120833	650	772	897	12192	2407	19.9	479862.6	47	44	18.5278
12282	474531	473886	700	120833	650	772	897	12192	2407	19.9	479862.6	47	44	19.1773
12282	474531	473886	800	120833	650	772	897	12192	2407	19.9	479862.6	47	44	19.7207
12282	474531	473886	900	120833	650	772	897	12192	2407	19.9	479862.6	47	44	20.1518
12282	474531	473886	1000	120833	650	772	897	12192	2407	19.9	479862.6	47	44	20.4786
12282	474531	473886	1100	120833	650	772	897	12192	2407	19.9	479862.6	47	44	20.7174



На рисунке показан график зависимости общего коэффициента смертности от числа мест в организациях, осуществляющих образовательную деятельность по образовательным программам дошкольного образования, при-смотр и уход за детьми. При увеличении числа мест в организациях происходит рост коэффициента.



**Рисунок 6. Зависимость общего коэффициента смертности от числа мест организаций**

В таблице 2 представлено изменение значений общего коэффициента смертности от доходов местного бюджета, фактически исполненных ( $Y_1$  – общий коэффициент смертности,  $x_5$  – доходы местного бюджета).

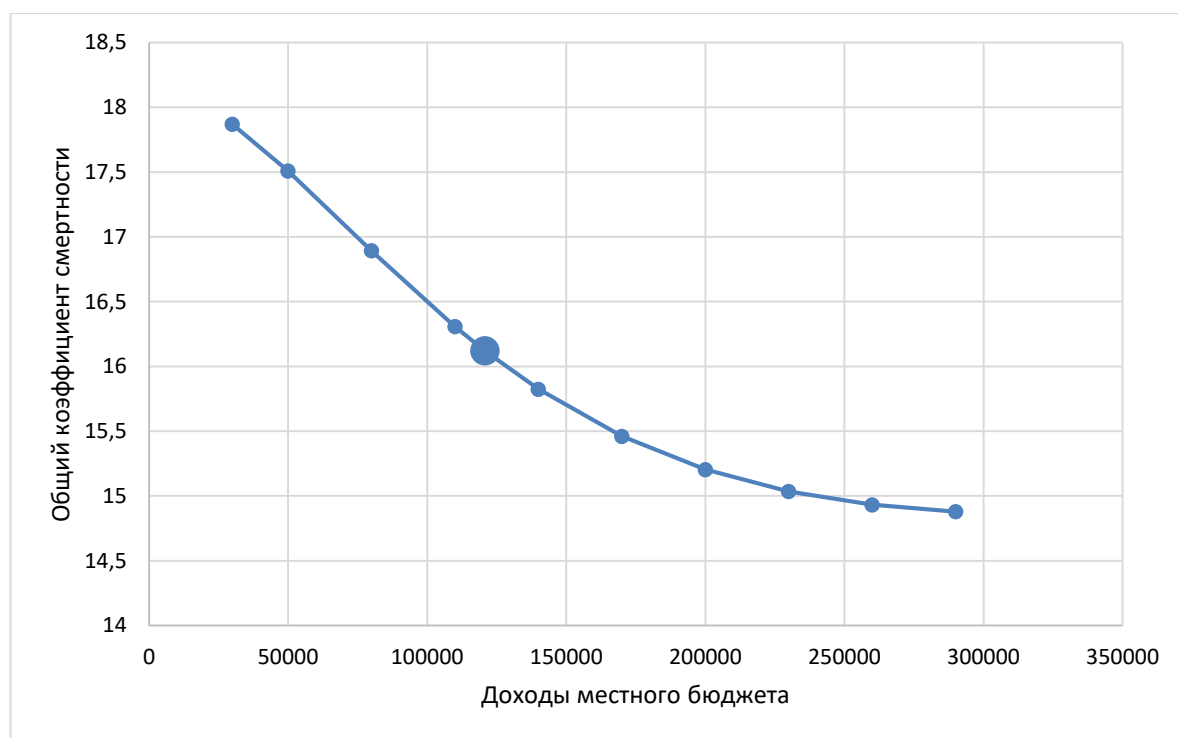
**Таблица 2**

**Изменение значений общего коэффициента смертности от доходов местного бюджета**

X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12	X13	X14	Y1
12282	474531	473886	286.8	30000	650	772	897	12192	2407	19.9	479862.6	47	44	17.8699
12282	474531	473886	286.8	50000	650	772	897	12192	2407	19.9	479862.6	47	44	17.5082
12282	474531	473886	286.8	80000	650	772	897	12192	2407	19.9	479862.6	47	44	16.8932
12282	474531	473886	286.8	110000	650	772	897	12192	2407	19.9	479862.6	47	44	16.3079
12282	474531	473886	286.8	120833	650	772	897	12192	2407	19.9	479862.6	47	44	16.1197
12282	474531	473886	286.8	140000	650	772	897	12192	2407	19.9	479862.6	47	44	15.8246
12282	474531	473886	286.8	170000	650	772	897	12192	2407	19.9	479862.6	47	44	15.4608
12282	474531	473886	286.8	200000	650	772	897	12192	2407	19.9	479862.6	47	44	15.2045
12282	474531	473886	286.8	230000	650	772	897	12192	2407	19.9	479862.6	47	44	15.0348
12282	474531	473886	286.8	260000	650	772	897	12192	2407	19.9	479862.6	47	44	14.9317
12282	474531	473886	286.8	290000	650	772	897	12192	2407	19.9	479862.6	47	44	14.8791

На рисунке 7 показан график зависимости общего коэффициента смертности от доходов местного бюджета, фактически исполненных. При увеличении доходов местного бюджета общий коэффициент убывает.

В таблице 3 представлено изменение значений общего коэффициента смертности от общего объема всех продовольственных товаров, реализованных в границах муниципального района, в денежном выражении за финансовый год ( $Y_1$  – общий коэффициент смертности,  $x_6$  – общий объем всех продовольственных товаров).



**Рисунок 7. Зависимость общего коэффициента смертности от доходов местного бюджета, фактически исполненных**

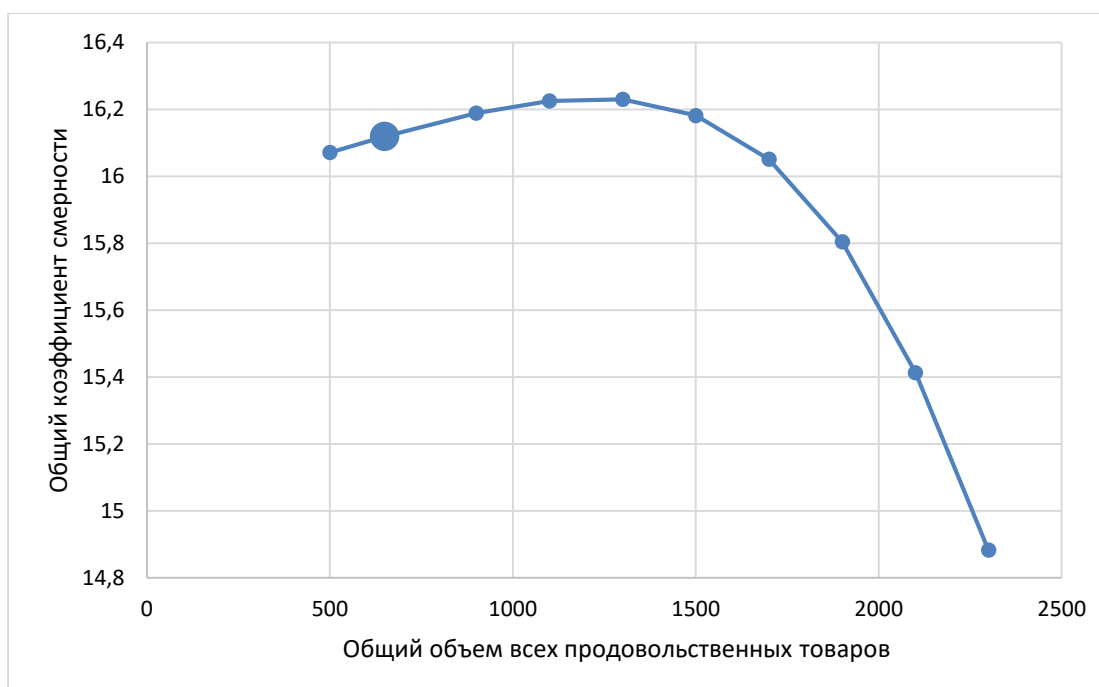
**Таблица 3**

**Изменение значений общего коэффициента смертности от общего объема всех продовольственных товаров**

X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12	X13	X14	Y1
12282	474531	473886	286.8	120833	500	772	897	12192	2407	19.9	479862.6	47	44	16.0715
12282	474531	473886	286.8	120833	650	772	897	12192	2407	19.9	479862.6	47	44	16.1197
12282	474531	473886	286.8	120833	900	772	897	12192	2407	19.9	479862.6	47	44	16.1895
12282	474531	473886	286.8	120833	1100	772	897	12192	2407	19.9	479862.6	47	44	16.226
12282	474531	473886	286.8	120833	1300	772	897	12192	2407	19.9	479862.6	47	44	16.2306
12282	474531	473886	286.8	120833	1500	772	897	12192	2407	19.9	479862.6	47	44	16.1823
12282	474531	473886	286.8	120833	1700	772	897	12192	2407	19.9	479862.6	47	44	16.0518
12282	474531	473886	286.8	120833	1900	772	897	12192	2407	19.9	479862.6	47	44	15.8048
12282	474531	473886	286.8	120833	2100	772	897	12192	2407	19.9	479862.6	47	44	15.4137
12282	474531	473886	286.8	120833	2300	772	897	12192	2407	19.9	479862.6	47	44	14.883

На рисунке 8 показан график зависимости общего коэффициента смертности от общего объема всех продовольственных товаров, реализованных в границах муниципального района, в денежном выражении за финансовый год. Сначала при увеличении общего объема всех продовольственных товаров общий коэффициент смертности увеличивается, но при дальнейшем увеличении происходит спад коэффициента.

**Заключение.** Построены две системы нейросетевого прогнозирования общего коэффициента смертности в Пермском крае. В системах учитываются 14 социально – экономических параметра. Методом сценарного прогнозирования [4] построены графики зависимостей прогнозируемого общего коэффициента смертности от изменения наиболее значимых входных параметров. Нейросетевая система, разработанная средствами Нейросимулятора [1] показала результаты чуть лучше, чем нейросетевая система, разработанная средствами библиотеки с открытым исходным кодом Keras [2].



**Рисунок 8. Зависимость общего коэффициента смертности от общего объема всех продовольственных товаров**

### **Библиографический список**

1. Черепанов Ф.М., Ясницкий Л.Н., Нейросимулятор 5.0 // Свидетельство о государственной регистрации программы для ЭВМ № 2014614649. Зарегистрировано в Реестре программ для ЭВМ 12 августа 2014 г.
2. Библиотека с открытым исходным кодом Keras для языка программирования Python [Электронный ресурс] // <https://keras.io/> (дата обращения 09.03.2022).
3. Паспорт муниципального образования [Электронный ресурс] // [https://www.gks.ru/scripts/db\\_inet2/passport/munr.aspx?base=munst5](https://www.gks.ru/scripts/db_inet2/passport/munr.aspx?base=munst5) (дата обращения 30.04.2021).
4. Ясницкий Л.Н., Зайцева Н.В., Гусев А.Л., Шур П.З. Нейросетевая модель региона для выбора управляющих воздействий в области обеспечения гигиенической безопасности // Информатика и системы управления. 2011. № 3 (29). С. 51-59.

## **NEURAL NETWORK SYSTEM FOR FORECASTING THE INDICATOR OF THE GENERAL DEATH RATE IN THE PERM KRAI**

*Filatova Polina D., Gusev Andrey L.*  
Perm State University  
Str. Bukireva, 15, Perm, Russia, 614990

The article presents a description of the development of a neural network system for predicting the total mortality rate. The system makes it possible to

predict the crude mortality rate with sufficient accuracy based on social and economic data for the Perm region. With the help of the developed intellectual system, a study of the subject area was carried out, patterns of practical importance were identified.

**Keywords:** artificial intelligence, neural network technologies, forecasting, crude mortality rate.

УДК 004.89

## НЕЙРОННЫЕ СЕТИ ДЛЯ КЛАССИФИКАЦИИ ВИРУСНЫХ ИНФЕКЦИЙ НА МОДИФИЦИРОВАННЫХ МОДЕЛЯХ ВОЙТА

*Стадниченко Никита Сергеевич, Новоселов Алексей Юрьевич,  
Пшеничный Александр Евгеньевич, Шаталова Ольга Владимировна*  
Юго-Западный государственный университет  
305040, Россия, г. Курск, ул. 50 лет Октября, 94, SHatOlg@mail.ru

Предложена модель импеданса биоматериала для классификаторов на основе обучаемых нейронных сетей. Для получения дескрипторов используется модифицированная модель Войта биоматериала.

**Ключевые слова:** модель биоимпеданса, многочастотное зондирование, обучаемый классификатор, итерационный алгоритм.

**Введение.** Исследование направлено на решение фундаментальной научной проблемы поиска новых, высокочувствительных, оперативных и неинвазивных методов оценки риска вирусных инфекций и их осложнений [1, 2]. Цель исследования – разработка метода формирования дескрипторов для классификаторов риска вирусных инфекций, позволяющих не только идентифицировать ранние проявления патологического очага, но и планировать профилактические и терапевтические мероприятия.

**Методы.** Дескрипторы для классификаторов риска вирусных инфекций формируются путем исследования модели пассивных электрических свойств биоматериалов (ПЭС) [3, 4]. Для построения модели ПЭС были использованы модифицированные звенья Войта. Звенья Войта были модифицированы с учетом того, что реальные конденсаторы имеют некоторое активное сопротивление помимо емкостного. На рисунке 1 представлена модель биоматериала с последовательной цепочкой модифицированных звеньев Войта.

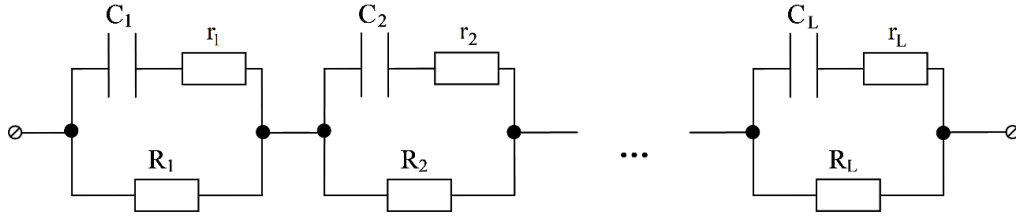


Рисунок 1. Структура модифицированной L-звенной модели Войта

Уравнения для расчета параметров модели имеют вид:

$$\left\{ \begin{array}{l}
 \frac{\omega_1^2 R_1 r_1 C_1 (r_1 C_1 + R_1 C_1) + R_1}{1 + \omega_1^2 (r_1 C_1 + R_1 C_1)^2} + \frac{\omega_1^2 R_2 r_2 C_2 (r_2 C_2 + R_2 C_2) + R_2}{1 + \omega_1^2 (r_2 C_2 + R_2 C_2)^2} + \dots \\
 \dots + \frac{\omega_1^2 R_L r_L C_L (r_L C_L + R_L C_L) + R_L}{1 + \omega_1^2 (r_L C_L + R_L C_L)^2} = a(\omega_1) \\
 \frac{-\omega_1 R_1^2 C_1}{1 + \omega_1^2 (r_1 C_1 + R_1 C_1)^2} + \frac{-\omega_1 R_2^2 C_2}{1 + \omega_1^2 (r_2 C_2 + R_2 C_2)^2} + \dots + \frac{-\omega_1 C_L R_L^2}{1 + \omega_1^2 (r_L C_L + R_L C_L)^2} = b(\omega_1) \\
 \frac{\omega_2^2 R_1 r_1 C_1 (r_1 C_1 + R_1 C_1) + R_1}{1 + \omega_2^2 (r_1 C_1 + R_1 C_1)^2} + \frac{\omega_2^2 R_2 r_2 C_2 (r_2 C_2 + R_2 C_2) + R_2}{1 + \omega_2^2 (r_2 C_2 + R_2 C_2)^2} + \dots \\
 \dots + \frac{\omega_2^2 R_L r_L C_L (r_L C_L + R_L C_L) + R_L}{1 + \omega_2^2 (r_L C_L + R_L C_L)^2} = a(\omega_2) \\
 \frac{-\omega_2 R_1^2 C_1}{1 + \omega_2^2 (r_1 C_1 + R_1 C_1)^2} + \frac{-\omega_2 R_2^2 C_2}{1 + \omega_2^2 (r_2 C_2 + R_2 C_2)^2} + \dots + \frac{-\omega_2 C_L R_L^2}{1 + \omega_2^2 (r_L C_L + R_L C_L)^2} = b(\omega_2) \quad (1) \\
 \vdots \\
 \frac{\omega_L^2 R_1 r_1 C_1 (r_1 C_1 + R_1 C_1) + R_1}{1 + \omega_L^2 (r_1 C_1 + R_1 C_1)^2} + \frac{\omega_L^2 R_2 r_2 C_2 (r_2 C_2 + R_2 C_2) + R_2}{1 + \omega_L^2 (r_2 C_2 + R_2 C_2)^2} + \dots \\
 \dots + \frac{\omega_L^2 R_L r_L C_L (r_L C_L + R_L C_L) + R_L}{1 + \omega_L^2 (r_L C_L + R_L C_L)^2} = a(\omega_L) \\
 \frac{-\omega_L R_1^2 C_1}{1 + \omega_L^2 (r_1 C_1 + R_1 C_1)^2} + \frac{-\omega_L R_2^2 C_2}{1 + \omega_L^2 (r_2 C_2 + R_2 C_2)^2} + \dots + \frac{-\omega_L C_L R_L^2}{1 + \omega_L^2 (r_L C_L + R_L C_L)^2} = b(\omega_L) \\
 \sum_{i=1}^L R_i = const
 \end{array} \right.$$

где  $a(\omega)$  и  $b(\omega)$  проекции частотных отсчетов графика Коула на действительную и мнимую оси, соответственно.

Поскольку каждое новое звено дает три новых неизвестных в систему уравнений (1), а каждый частотный отсчет на графике Коула дает 2 уравнения, то максимальное число звеньев в цепи, параметры которых возможно рассчитать согласно (1), составит 8 звеньев. Причем система уравнений (1) может быть либо переопределена, либо не доопределена.

Рассчитав модели групповым методом [2], получим график ошибок и импедансные диаграммы модели ПЭС (рисунок 2). Ошибки на каждой  $n$ -ой частоте импедансной диаграммы определяются как

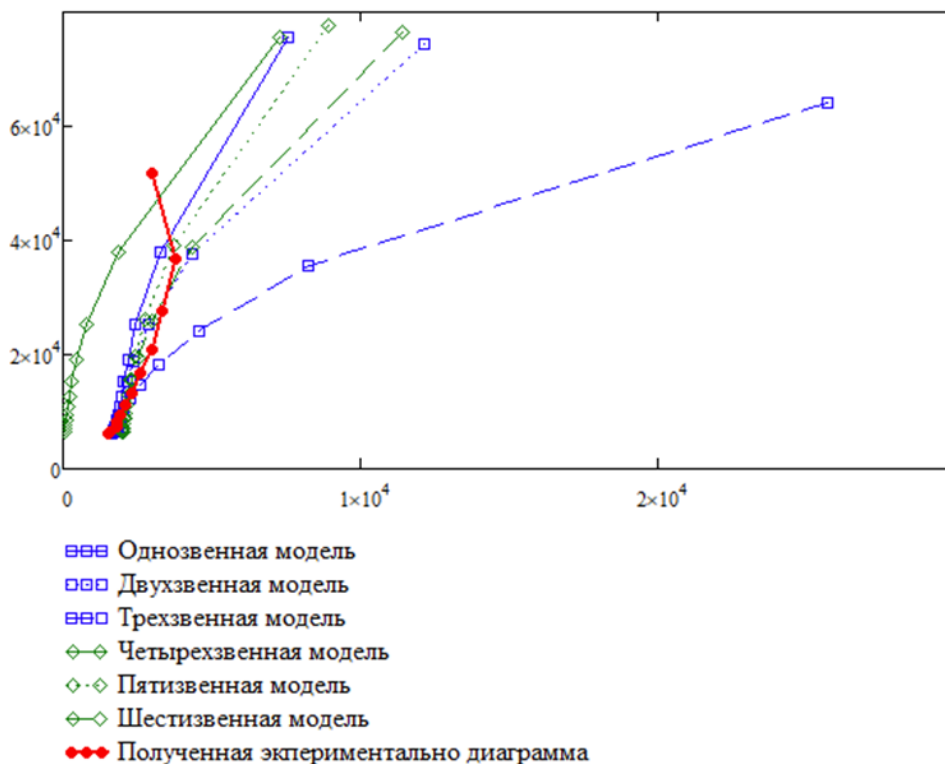
$$\Delta Re(n, \ell) = Re(n) - \overline{Re(n, \ell)}, \quad (2)$$

$$\Delta \text{Im}(n, \ell) = \text{Im}(n) - \overline{\text{Im}}(n, \ell), \quad (3)$$

где составляющие импеданса, полученные посредством модели Войта, имеют идентификаторы с верхней чертой.

Общая ошибка аппроксимации  $\ell$ -й модели Войта определяется как сумма векторов ошибок на каждом отсчете частот

$$\varepsilon(\ell) = \sum_{n=1}^N \sqrt{(\Delta \text{Re}(n, \ell))^2 + (\Delta \text{Im}(n, \ell))^2}. \quad (4)$$



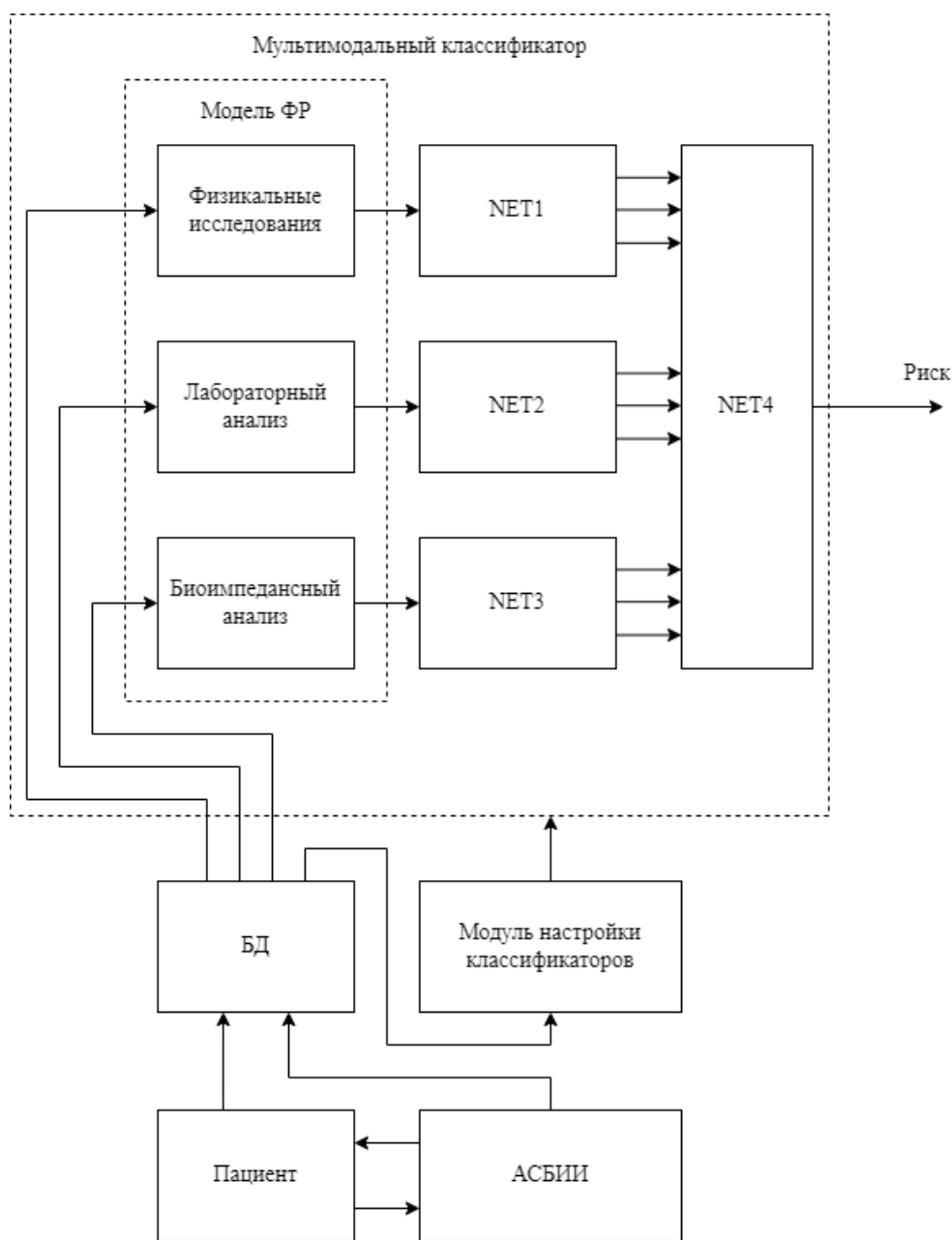
**Рисунок 2. Импедансные диаграммы моделей с наименьшими ошибками аппроксимации на частотах 1-12 кГц**

**Результаты.** В модели прогнозирования риска вирусных инфекций и их осложнений в пространстве ФР необходимо выделить сегменты для мультимодального классификатора (ММК) риска таким образом, чтобы посредством экспертной оценки можно было оценить влияние предикторов сегментов на этот вид МР. При этом ММК построен таким образом, чтобы решения классификаторов нижнего иерархического уровня использовались в качестве дескрипторов классификатора верхнего иерархического уровня.

В первый сегмент ФР включены физикальные показатели: лихорадка, постоянный кашель, утомляемость, одышка, диарея, делирий, пропуск приема пищи, боль в животе, боль в груди и охриплость голоса, а также возраст, пол и индекс массы тела (ИМТ). Во втором сегменте использовать показатели исследования крови. В качестве третьего сегмента ФР взяты дескрипторы, полученные на основе БА. Классификаторы, построенные на их основе, рассмотрены в разделе 3 настоящей диссертации. В связи с этим в качестве ба-

зовой системы прогнозирования риска вирусной инфекции и ее осложнений была взята диагностическая модель ML, представленная рисунке 3.

Ядром мультиагентной интеллектуальной системы является мультимодальный классификатор (ММК) риска вирусной инфекции и ее осложнений, включающий три автономных интеллектуальных агента (АИА): NET1, NET2 и NET3, а также АИА агрегатор – NET4. Модели факторов риска для этих АИА формируются на основе физикальных, лабораторных и биоимпедансных исследований. Данные о пациенте, а также дата сет для обучения АИА, хранятся в базе данных (БД). Для биоимпедансных исследований пациента используется автоматизированная система биоимпедансных исследований (АСБИИ). Для настройки нейронных сетей ММК используется модуль настройки классификаторов, ПО которого разработано в среде Matlab 2018b.



**Рисунок 3. Структурная схема мультиагентной интеллектуальной системы для прогнозирования риска вирусной инфекции и ее осложнений**

Для получения «сырых» данных биоимпедансного анализа на грудную клетку пациента надевался электродный пояс и определялось множество графиков Коула, соответствующих определенному сочетанию электродов. Для формирования матричного пространства дескрипторов, исходя из графика зависимости ошибки аппроксимации импедансной диаграммы от числа звеньев в модели Войта, выбиралось число звеньев  $\ell$  в модели Войта. Таким образом, путем решения систем нелинейных алгебраических уравнений, полученным по экспериментальным графикам Коула, вычисляем матрицу дескрипторов с размерностью  $\ell \times I$ , где  $I$  – число электродов в электродном поясе.

Для определения качества классификации использовались известные показатели качества: диагностическая чувствительность (ДЧ), диагностическая специфичность (ДС) и диагностическая эффективность (ДЭ). Показатели качества диагностики предлагаемого классификатора сравнивались, как с прототипом, с показателями качества рентгеновских исследований на той же контрольной выборке. Показатели качества различных моделей классификаторов достигали 78% и не опускались ниже 62%, что сопоставимо с показателями качества методов рентгеновской диагностики.

**Заключение.** В результате исследования получены принципиально новые результаты, позволяющие создавать интеллектуальные системы поддержки принятия решений для диагностики кардиореспираторных заболеваний. Создана модель биоимпедансного анализа на основе многочастотного измерения биоимпеданса, позволяющая разложить импеданс биоматериала на структурные элементы.

На основе многочастотного зондирования и моделей Войта с использованием нейросетевых моделей могут быть получены алгоритмы дифференциального контроля импеданса тканей и импеданса биожидкости, что позволит получить новые правила принятия решений для диагностики патологических состояний организма (сердечно-сосудистых, инфекционных и онкологических заболеваний).

### Библиографический список

1. Мирошников, А.В. Модели импеданса биоматериала для формирования дескрипторов в интеллектуальных системах диагностики инфекционных заболеваний / А.В. Мирошников, Н.С. Стадниченко, О.В. Шаталова, С.А. Филист // Моделирование, оптимизация и информационные технологии. – 2020. – Т. 8, № 4 (31). – С. 1-14. – URL: <https://moitvvt.ru/ru/journal/pdf?id=864>. – DOI: 10.26102/2310-6018/2020.31.4.018.
2. Мирошников, А.В. Алгоритм оптимизации модели Войта в классификаторах функционального состояния живых систем / А.В. Мирошников, О.В. Шаталова, М.А. Ефремов, Н.С. Стадниченко, А.Ю. Новоселов, А.В. Павленко // Известия Юго-Западного государственного университета. Серия: Управление, вычислительная техника, информатика. Медицинское приборостроение. – 2022. – Т.12, № 2. – С. 59-75.
3. Филист, С.А. Биотехническая система для контроля импеданса биоматериалов в экспериментах *in vivo* / С.А. Филист, А.А. Кузьмин, М.Н. Кузьмина // Биомедицинская радиоэлектроника – 2014. – №9 – С. 38-42.



4. Мирошников, А.В. Классификации биологических объектов на основе многомерного биоимпедансного анализа// А.В. Мирошников, О.В. Шаталова, Н.С. Стадниченко, Л.В. Шульга //Известия Юго-Западного государственного университета. Серия: Управление, вычислительная техника, информатика. Медицинское приборостроение. – 2020. – Т. 10, № 3/4. – С. 29-49.
5. Филист, С.А. Модель формирования функциональных систем с учетом менеджмента адаптационного потенциала / Филист С.А., Шуткин А.Н., Шкатова Е.С., Дегтярев С.В., Савинов Д.Ю.// Биотехносфера 1(55) 2018. С.32-37.
6. Попечителей, Е.П. Способы и модели идентификации биоматериалов на основе анализа многочастотного импеданса/ Е.П. Попечителей, С.А. Филист // Известия Юго-Западного государственного университета. Серия Управление, вычислительная техника. Медицинское приборостроение №1, 2011. С.74-80.
7. Шаталова, О.В. Итерационная многопараметрическая модель биоимпеданса в экспериментах *in vivo* / О.В. Шаталова // Известия Юго-Западного государственного университета. Серия: Управление, вычислительная техника, информатика. Медицинское приборостроение – 2019. – Т. 9, № 1 (30). – С. 26-38. – URL: [https://swsu.ru/izvestiya/seriesivt/archiv/1\\_2019.pdf](https://swsu.ru/izvestiya/seriesivt/archiv/1_2019.pdf).
8. Кассим, К.Д.А. Параметрические модели биоимпеданса для идентификации функционального состояния живой системы / К.Д.А. Кассим, И.А. Ключиков, О.В. Шаталова, З.Д. Яа // Биомедицинская радиоэлектроника. – 2012. – № 4. – С. 50-56.
9. Филист, С.А. Гибридные информационные технологии по экспресс-диагностике инфекционных заболеваний на основе многочастотного анализа пассивных свойств биотканей/С.А. Филист, В.А. Алексенко, Кабус Кассим // Известия ЮФУ. Технические науки. Тематический выпуск. «Медицинские информационные системы» / Таганрог, 2010. №8(109). С.12-17.
10. Филист, С.А. Гибридная нейронная сеть с макрослоями для медицинских приложений / С.А. Филист, О.В. Шаталова, М.А. Ефремов // Нейрокомпьютеры. Разработка и применение. – 2014. – №6. – С. 35-39.

## NEURAL NETWORKS FOR THE CLASSIFICATION OF VIRAL INFECTIONS ON MODIFIED VOIT MODELS

*Stadnichenko Nikita S., Novoselov Alexey Yu.,  
Pshenichny Alexander E., Shatalova Olga V.*

Southwestern State University

Str. 50 years of October, 94, Kursk, Russia, 30540, shatolg@mail.ru

A biomaterial impedance model for classifiers based on trained neural networks is proposed. To obtain descriptors, a modified Voigt model of the biomaterial is used.

**Keywords:** bioimpedance model, multifrequency sounding, trainable classifier, iterative algorithm.

## ИНТЕЛЛЕКТУАЛЬНАЯ СИСТЕМА ИДЕНТИФИКАЦИИ НЕМАРКИРОВАННЫХ ЭЛЕМЕНТОВ ПИТАНИЯ

*Коротышева Анна Андреевна<sup>1</sup>, Милов Владимир Ростиславович<sup>2</sup>,  
Егоров Юрий Сергеевич<sup>2</sup>, Кербенева Анна Юрьевна<sup>2</sup>*

<sup>1</sup> Нижегородский государственный университет им. Н.И. Лобачевского  
603022, г. Нижний Новгород, пр. Гагарина, 23, ania.korotishewa@yandex.ru

<sup>2</sup> Нижегородский государственный технический университет им. Р.Е. Алексева  
603950, г. Нижний Новгород, ул. Минина, 24, ckar@list.ru

В статье представлено описание разработки нейросетевой системы для идентификации немаркированных элементов питания. Предложенная система позволяет детектировать с высокой точностью элементы питания, а также классифицировать их тип на основании цветных изображений и рентген-снимков.

**Ключевые слова:** искусственный интеллект, нейронная сеть, элементы питания.

**Введение.** В условиях изменения мировой экологической политики, все более актуальной становится проблема переработки твердых бытовых отходов (ТБО), а также таких опасных отходов, как электрохимические источники электрического тока, к которым относятся гальванические элементы питания (батарейки) и электрические аккумуляторы.

Несмотря на то, что элементы питания (ЭП) должны утилизироваться отдельно от остальных бытовых отходов, удаленность пунктов сбора вынуждает большинство жителей не накапливать отработавшие батарейки в домашних условиях, а выбрасывать их вместе с обычным мусором.

Попадая на полигон, металлический корпус батарейки начинает разлагаться, этикетка, как правило, стирается. В течение нескольких лет под воздействием осадков происходит процесс выщелачивания металлов (цинк, магний, кадмий, олово, свинец), которые попадают в почву и проникают в водоемы, а оттуда уже в систему водоснабжения, что оказывает сильное негативное воздействие на окружающую среду и человека.

Необходимость сортировки опасных отходов предопределяет актуальность исследования. В настоящее время на отечественном рынке отсутствуют универсальные решения для автоматической идентификации немаркированных элементов в процессе их сортировки. Указанная проблема не позволяет автоматизировать и масштабировать переработку ТБО, а также сдерживает создание автоматизированных сортировочных линий, обладающих значительно более высоким потенциалом, чем ручная сортировка элементов питания.

Цель работы – разработка прототипа интеллектуальной системы идентификации немаркированных ЭП с использованием методов машинного обучения (ИСИНЭП).

**Практическая часть.** Функциональное моделирование основных процессов разрабатываемой интеллектуальной системы идентификации немаркированных элементов питания выполнено в нотации IDEF0. На рисунке 1 представлена контекстная диаграмма с указанием входных и выходных воздействий, пользователя и управляющих воздействий.



**Рисунок 1. Контекстная диаграмма А-0 «Интеллектуальная система идентификации немаркированных элементов питания»**

Прототип ИСИНЭП состоит из трех основных модулей.

1. Модуль ввода данных – предназначен для загрузки изображений, ввода данных об элементах питания и команд пользователя.

2. Модуль интеллектуальной обработки данных – предназначен для детектирования объектов на изображении и их идентификации с использованием методов машинного обучения.

3. Модуль визуализации – предназначен для графической визуализации информации для пользователя, а также выгрузки результатов обработки данных.

На вход модуля ввода данных подаются системные пути до RGB изображения и рентген-снимка с элементами питания, а также команды пользователя. Под командами понимаются пороговые значения оценок вероятностей детектирования элементов питания и классификации рентген снимков нейросетевыми алгоритмами.

В процессе интеллектуальной обработки данных происходит детектирование элементов питания нейронной сетью на входном изображении с вы-

числением координат рамок, ограничивающих каждый из элементов, сопоставление полученных координат ЭП с входным рентген-снимком, определение типа ЭП на рентген-снимке с помощью нейросетевого классификатора.

На завершающем этапе в модуле визуализации найденные ограничивающие рамки ЭП накладываются на исходное изображение и отображаются пользователю. Типы элементов питания вместе с координатами сохраняются в базе данных и итоговый табличный файл с аннотацией.

Датасеты для нейросетевых моделей были собраны с использованием видеокамеры и рентгеновского аппарата и поделены на обучающие и тестовые выборки.

Примеры изображений ЭП (фотография и рентген-снимок) из собранных датасетов представлены на рисунке 2.



**Рисунок 2.** Примеры изображений ЭП из собранных датасетов

Для предотвращения переобучения использована процедура перекрестной проверки, основанная на методе повторной выборки M-fold CV (в работе использована 4-fold CV) [1]. Детектор в интеллектуальной системе идентификации немаркированных элементов питания построен на основе архитектуры YOLOv5 [2], а классификатор на основе нейронной сети MobileNetV2 [3].

Рекомендации по использованию созданного прототипа ИСИНЭП заключаются в разработке на его основе программно-аппаратного комплекса для автоматизации сортировки твердых бытовых отходов в части обнаружения и выделения (сортировки) элементов питания из общего объема ТБО.

**Заключение.** Построена система с нейросетевым детектором и классификатором для идентификации немаркированных элементов питания [4], основанная на комплексировании обработки цветных изображений и рентген-снимков. Разработана архитектура и алгоритмы работы системы, реализован программный код, проведено тестирование, по результатам которого подтверждена возможность внедрения на мусороперерабатывающем предприятии для автоматизированной идентификации и сортировки ЭП в составе ТБО.

Работа выполнена при поддержке Фонда содействия развитию малых форм предприятий в научно-технической сфере (договор № 57ГС1ИИС12-D7/72200 от 21.12.2021 г.).

### Библиографический список

1. Шунина Ю.С. Влияние способа формирования обучающей и тестовой выборки на качество классификации // Вестник Ульяновского государственного технического университета. 2015. № 2 (70). С. 43-46.
2. Redmon J., Divvala S., Girshick R., Farhadi A. You Only Look Once: Unified, Real-Time Object Detection // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016. Pp. 779-788.
3. Sandler M., Howard A.G., Zhu M., Zhmoginov A., Chen L.-C. MobileNetV2: Inverted Residuals and Linear Bottlenecks // Computer Vision and Pattern Recognition. 2018. Pp. 4510-4520.
4. Блатов Р.И., Вострякова Е.А., Москвин А.С., Чупров Д. А., Егоров Ю.С., Коротышева А.А., Милов В.Р., Дубов М. С., Кербенева А.Ю. Свидетельство о регистрации программы для ЭВМ RUS 2022663863. Заявка № 2022662975 от 11.07.2022.

## INTELLIGENT SYSTEM FOR IDENTIFYING UNLABELED BATTERIES

*Anna A. Korotysheva<sup>1</sup>, Milov Vladimir R.<sup>2</sup>,  
Egorov Yuri S.<sup>2</sup>, Kerbeneva Anna Y.<sup>2</sup>*

<sup>1</sup> Lobachevsky State University of Nizhny Novgorod  
603022, Nizhny Novgorod, Gagarin Ave, 23, ania.korotishewa@yandex.ru

<sup>2</sup> Nizhny Novgorod State Technical University n.a. R.E. Alekseev  
603950, Nizhny Novgorod, Minin Street, 24, ckar@list.ru

The article describes the development of an intelligent system for the identification of unlabeled batteries. The system allows to detect with great accuracy the elements and classify their type based on RGB images and X-rays images.

**Keywords:** artificial intelligence, neural network, batteries.

# **ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ В ПСИХОЛОГИИ И СОЦИОЛОГИИ**

## НЕЙРОСЕТЕВАЯ СИСТЕМА ПРОГНОЗИРОВАНИЯ УРОВНЯ СТРЕССА ПОСЛЕ ФИЗИЧЕСКОЙ АКТИВНОСТИ

*Болотная Татьяна Евгеньевна*

Национальный исследовательский университет  
«Высшая школа экономики». 614990, Россия, г. Пермь,  
ул. Бульвар Гагарина, 37, tebolotnaya@edu.hse.ru

В статье представлено описание разработки нейросетевой системы для прогнозирования уровня стресса. Система позволяет с большой точностью определить уровень стресса человека после физической активности, основываясь на следующих показателях: влажность тела, температура тела, количество шагов. С помощью разработанной интеллектуальной системы проведено исследование предметной области, выявлены закономерности, имеющие практическое значение.

**Ключевые слова:** искусственный интеллект, нейросетевые технологии, прогнозирование, определение уровня стресса.

**Введение.** Стресс — это ответная реакция организма человека на перенапряжение, негативные и позитивные эмоции. Во время стресса организм человека вырабатывает гормон адреналин, который заставляет искать выход. Стресс в небольших количествах нужен всем, так как он заставляет думать, искать выход из проблемы, и в этом случае он имеет положительное значение. Но, с другой стороны, если стрессов становится слишком много, организм слабеет, теряет силы, способность решать проблемы и может вызвать серьезные заболевания. Психологический стресс влияет на физиологические параметры человека. Длительное воздействие стресса может иметь пагубные последствия, которые могут потребовать дорогостоящего лечения.

В последнее время специалисты исследуют множество путей преодоления стресса и взаимосвязи с разными сферами человеческой жизни, например -взаимосвязь физической активности и уровня стресса человека. Также исследователи делают обозначают, что очень важно мониторить уровень стресса [6], дабы избежать летальных последствий. На сегодняшний день существует очень мало исследований на тему прогнозирования медицинских показателей [1, 2, 4, 5], в основном исследователи обращаются к качественному анализу для формирования гипотез. Однако на просторе Интернета всё-таки можно найти парочку исследований, например, исследование Laavana Rachakonda [7]. В данном исследовании разрабатывается система для прогнозирования уровня стресса, которую в дальнейшем можно будет использовать в медицинских учреждениях.

Основная цель настоящей работы заключается в сборе множества публичных данных о состоянии человеческого тела, а также создание и обучение

нейросетевой модели на основе имеющихся данных. В результате исследования должна работать нейросетевая система, способная прогнозировать уровень стресса после активности.

Для создания нейросетевой системы выбраны следующие входные параметры:

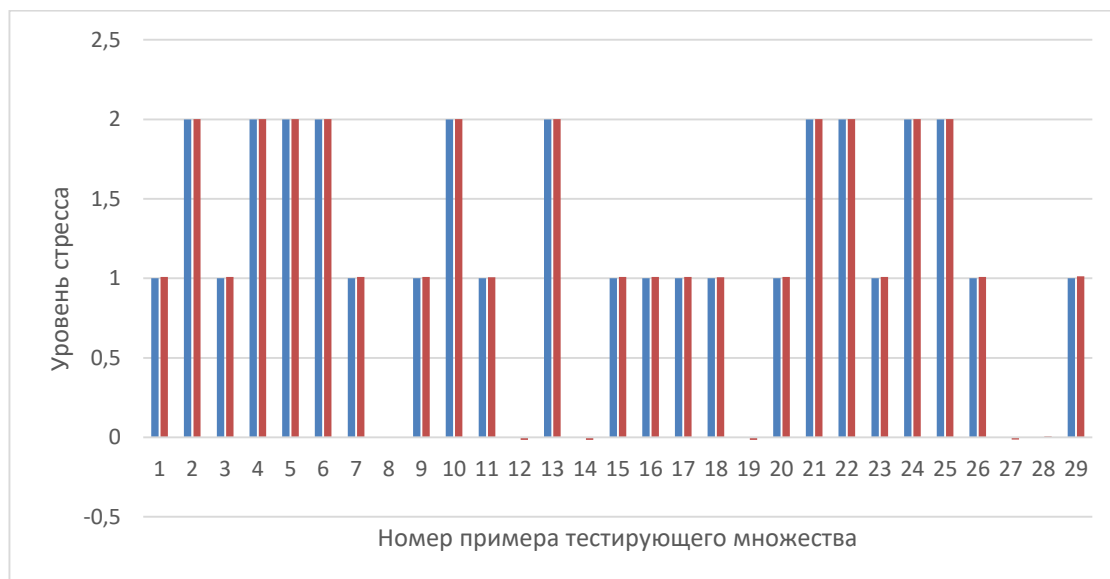
- X1 – влажность тела после физической активности,
- X2 – температура тела,
- X3 – количество шагов.

За выходной параметр принимается уровень стресса (от 0 до 2). Исходное множество собрано с помощью интернет-ресурса организации конкурсов по исследованию данных. Объем исследовательского множества составляет – 2001 примеров. Данное множество разделено на обучающее, валидирующее и тестовое в соотношении 70%:10%:20%:

- обучающее – 1361 пример (~70%);
- валидирующее – 240 примеров (~10%);
- тестовое – 400 примеров (~20%).

Проектирование, обучение, тестирование нейронной сети выполнялись с помощью программы «Нейросимулятор 5» [3]. В результате оптимизации спроектированная нейронная сеть представляла собой модель (персептрон), который имеет 3 входных нейрона, один выходной и один скрытый слой с двумя нейронами.

После обучения нейросимулятора и для оценки корректной работы спроектированной нейронной сети использовалось тестирующее множество, которое состоит из 400 примеров. Средняя относительная ошибка тестирования составила 0,9199%, что можно считать приемлемым результатом.

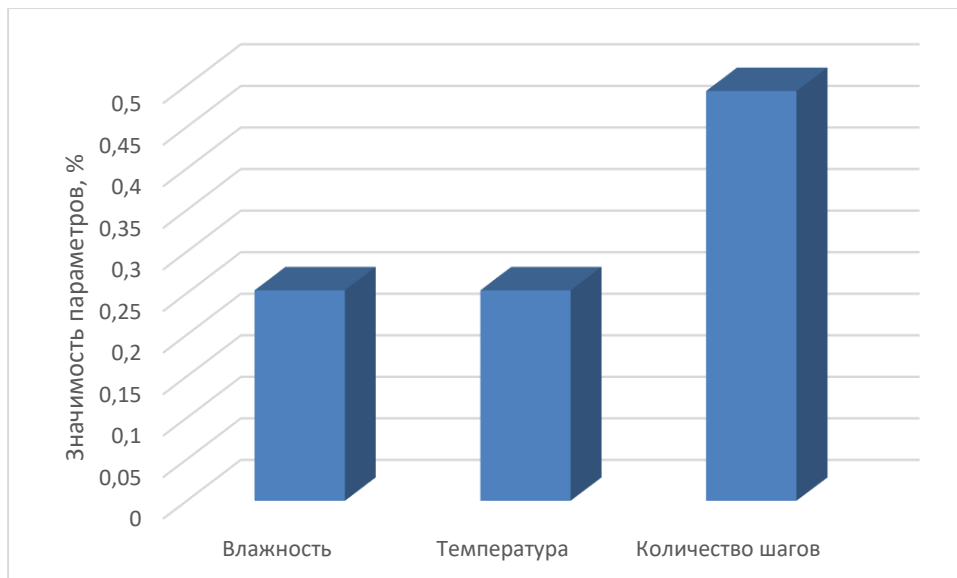


**Рисунок 1. Результат тестирования нейронной сети**

Исходя из результатов, которые отображены на рисунке 1, можно сделать вывод об адекватной работе спроектированной нейронной сети.

Далее рассчитана оценка значимости признаков, диаграмму можно посмотреть на рисунке 2.

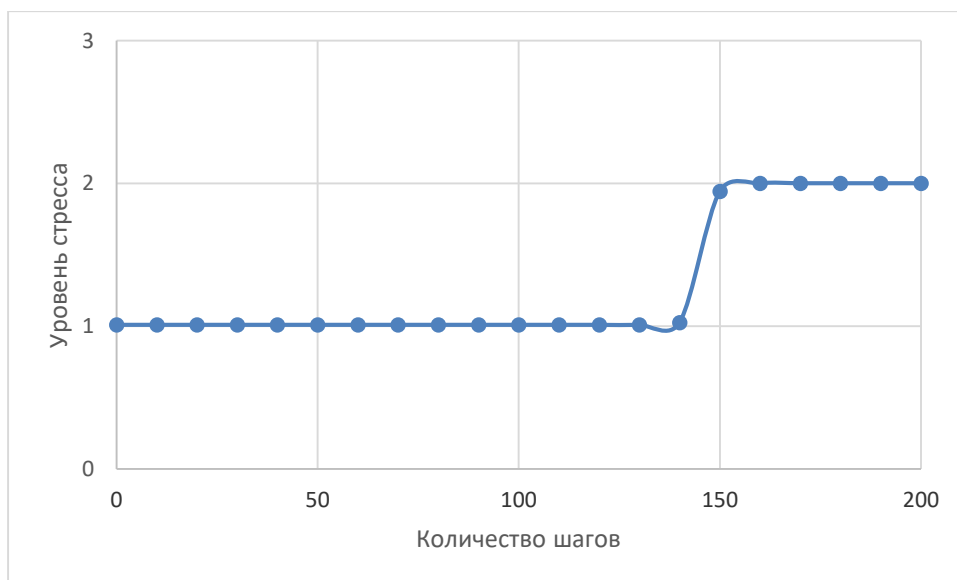




**Рисунок 2. Значимость входных параметров нейросетевой модели**

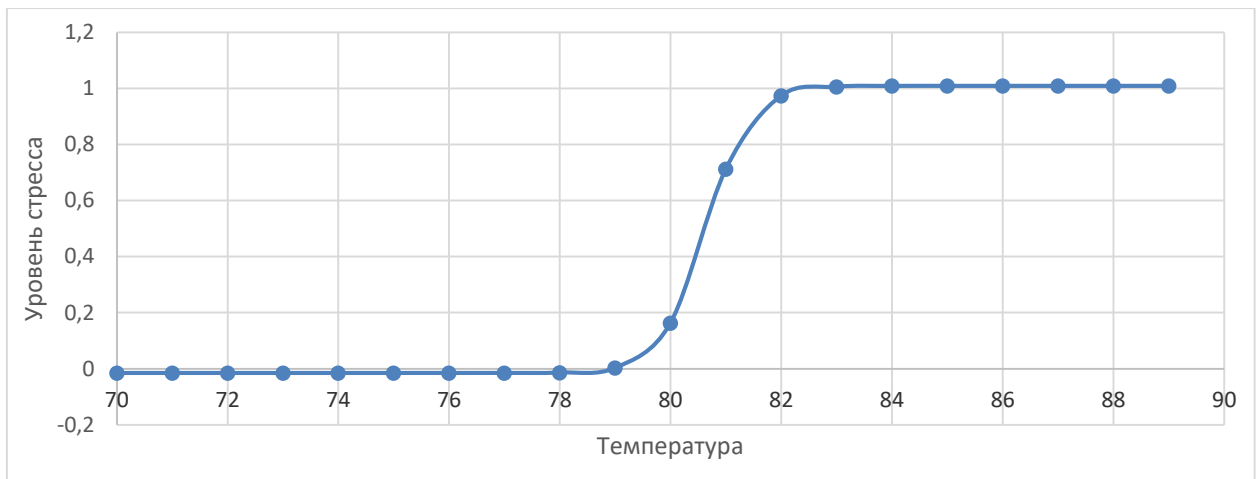
На диаграмме видно, что наиболее значимым входным параметром является количество шагов, он же и самый влиятельный, так как от того, сколько человек прошел, зависит состояние его тела.

Далее было проведено исследование полученных зависимостей между входными параметрами и уровнем стресса. Для этого исследование для начала случайным образом был отобран пример, по которому нейросеть не сможет определить точный уровень стресса. Зависимость уровня стресса от количества шагов:



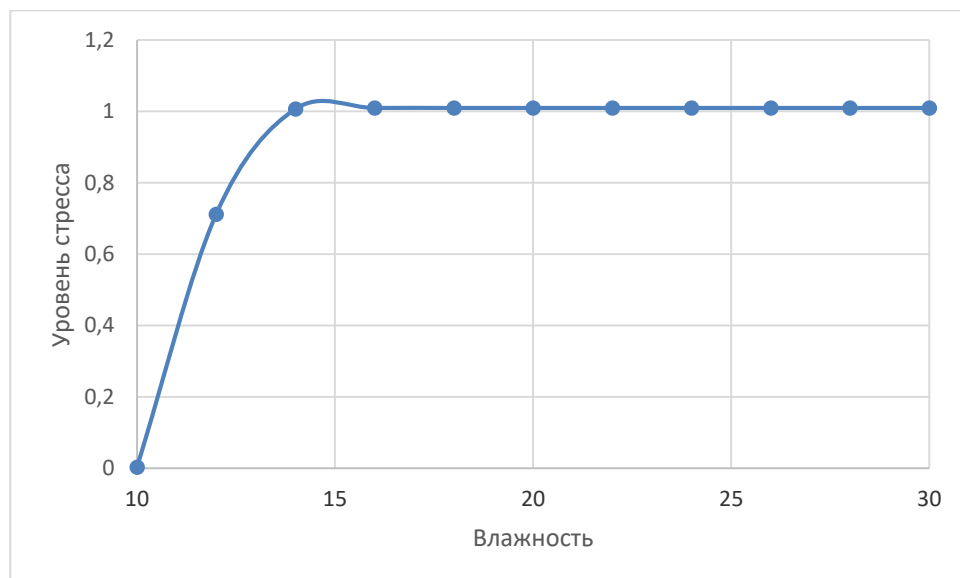
**Рисунок 3. Зависимость уровня стресса от количества шагов**

На рисунке 3 показан график зависимости прогнозируемого уровня стресса от количества шагов – чем больше количество шагов, тем больше уровень стресса.



**Рисунок 4. Зависимость уровня стресса от температуры тела**

На рисунке 4 продемонстрирована зависимость прогнозируемого уровня стресса от температуры тела после спортивных активностей. Уровень стресса увеличивается, если повышается температура.



**Рисунок 5. Зависимость уровня стресса от влажности тела**

На рисунке 5 изображен график зависимости уровня стресса от влажности тела. Как видно из графика, с повышением уровня пота, повышается уровень стресса.

Полученные результаты исследований не противоречат реальности, что подтверждает, что спроектированную нейронную сеть можно считать пригодной для прогнозирования уровня стресса, однако, многие исследователи рекомендуют быть аккуратнее и внимательнее при прогнозировании медицинских показателей.

**Заключение.** Построена система нейросетевого прогнозирования уровня стресса на основе показателей после спортивных активностей человека. Спроектированная нейросетевая модель учитывает 3 параметра: уровень пота (влажность тела) после спортивных активностей, температура тела, ко-

личество шагов. Методом сценарного прогнозирования построены графики зависимостей прогнозируемого уровня стресса от изменения входных параметров. Применение такого набора параметров в модели позволяет с высокой точностью прогнозировать уровень стресса.

### **Библиографический список**

1. Ясницкий Л.Н., Грацилёв В.И., Куляшова Ю.С., Черепанов Ф.М. Возможности моделирования предрасположенности к наркозависимости методами искусственного интеллекта // Вестник Пермского университета. Философия. Психология. Социология. 2015. № 1 (21). С. 61-73.

2. Ясницкий Л.Н., Павлов И.В., Черепанов Ф.М. Прогнозирование результатов олимпийских игр 2014 года в неофициальном командном зачете методами искусственного интеллекта // Современные проблемы науки и образования. 2013. № 6. С. 985.

3. Ясницкий Л.Н. Черепанов Ф.М. Нейросимулятор 4.0. Свидетельство о регистрации программы для ЭВМ RUS 2014612546. Заявка № 2014610341 от 15.01.2014.

4. Ясницкий Л.Н., Сичинава З.И. Нейросетевые алгоритмы анализа поведения респондентов // Нейрокомпьютеры: разработка, применение. 2011. № 10. С. 59-64.

5. Yasnitsky L.N., Dumler A., Cherepanov F.M. The capabilities of artificial intelligence to simulate the emergence and development of diseases, optimize prevention and treatment thereof, and identify new medical knowledge // Journal of Pharmaceutical Sciences and Research. 2018. Т. 10. № 9. С. 2192-2200.

6. H. Thapliyal, V. Khalus, and C. Labrado, "Stress Detection and Management: A Survey of Wearable Smart Health Devices," IEEE Consum. Electron. Mag., vol. 6, no. 4, pp. 64–69, Oct 2017.

7. Laavanya Rachakonda. Stress-Lysis: A DNN-Integrated Edge Device for Stress Level Detection in the IoMT, IEEE Consum. Electron. Mag., vol. 6, no. 4, pp. 64–69, Oct 2017.

## **NEURAL NETWORK SYSTEM FOR PREDICTING THE STRESS LEVEL AFTER PHYSICAL ACTIVITY**

*Bolotnaya Tatiana E.*

National Research University Higher School of Economics,  
Studencheskaya st., 38, Perm, Russia, 614990, tebolotnaya@edu.hse.ru

The article describes the development of a neural network system for predicting the stress level. The system allows you to predict with great accuracy stress level after physical activity.

**Keywords:** artificial intelligence, neural network technologies, forecasting, stress level determination.

## НЕЙРОСЕТЕВАЯ СИСТЕМА ПРОГНОЗИРОВАНИЯ СКЛОННОСТИ ЧЕЛОВЕКА К ДЕПРЕССИИ

*Красных Роман Сергеевич*

Пермский государственный национальный исследовательский университет,  
ПМИ. 614990, Россия, г. Пермь, ул. Букирева, 15, ro-  
man.krasnykh02@gmail.com

В статье представлено описание разработки нейросетевой системы для прогнозирования склонности человека к депрессии. Система позволяет склонности человека к депрессии на основе психологического теста и общей информации о человеке.

**Ключевые слова:** искусственный интеллект, нейросетевые технологии, прогнозирование, депрессия.

**Введение.** Депрессия – это вид психического расстройства, из-за которого человек находится в подавленном состоянии. Такое заболевание часто провоцирует снижение самооценки, апатию и ощущение ненужности.

Основные симптомы депрессии:

Подавленное настроение, потеря удовольствия, усталость, потеря уверенности в себе или самоуважения, чрезмерная самокритика или необоснованное чувство вины, повторяющиеся мысли о смерти или самоубийстве, либо попытки к нему, чувство нерешительности, потеря концентрации, психомоторные заторможенность или возбуждение, нарушения сна, изменение аппетита и веса, нестабильность эмоций, тревожность, беспричинные страхи, потеря интереса ко всему.

Инструментом для выяснения скрытых закономерностей и построением на их основе математической модели являются методы искусственного интеллекта – нейросетевые технологии.

В качестве входных параметров были использованы:

X1 Меня расстраивали довольно тривиальные вещи.

X2 Мне показалось, что я вообще не испытывал никаких положительных эмоций.

X3 Я просто не могу начать.

X4 Я был склонен чрезмерно реагировать на ситуации.

X5 Мне было трудно расслабиться.

X6 Я попадал в ситуации, которые меня так тревожили, когда они закончились, я испытал огромное облегчение.

X7 Я чувствовал, что мне нечего ждать.

X8 Я довольно легко расстраивался.

X9 Я чувствовал, что трачу много нервной энергии.

X10 Я чувствовал себя грустным и подавленным.

X11 Я терял терпение, когда меня задерживали каким-либо образом (например, лифты, светофоры, меня заставляли ждать).

X12 У меня было чувство дурноты.

X13 Я чувствовал, что потерял интерес практически ко всему.

X14 Я чувствовал, что ничего не стою как человек.

X15 Я чувствовал себя довольно обидчивым.

X16 Я заметно потел (например, потеют руки) при отсутствии высоких температур или физических нагрузок.

X17 Я испугался без уважительной причины.

X18 Я чувствовал, что жизнь бесполезна.

X19 Мне было трудно успокоиться.

X20 Мне казалось, что я не получаю никакого удовольствия от того, чем занимаюсь.

X21 Я осознавал работу своего сердца при отсутствии физических нагрузок (например, ощущение учащения пульса, отсутствие биения сердца).

X22 Я чувствовал себя подавленным и грустным.

X23 Я чувствовал, что близок к панике.

X24 Мне было трудно успокоиться после того, как что-то меня расстроило.

X25 Я боялся, что меня "выкинут" какой-то банальной, но незнакомой задачей.

X26 Я не мог ни в чем проявлять энтузиазм.

X27 Мне было трудно терпеть отвлечения от того, что я делал.

X28 Я находился в состоянии нервного напряжения.

X29 Я чувствовал себя бесполезным.

X30 Я был нетерпим ко всему, что мешало мне продолжать то, что я делал.

X31 Я испугался.

X32 Я не видел ничего, на что можно было надеяться в будущем.

X33 Я чувствовал, что жизнь бессмысленна.

X34 Я начал волноваться.

X35 Меня беспокоили ситуации, в которых я мог запаниковать и выставить себя дураком.

X36 Я почувствовал дрожь (например, в руках).

X37 Мне было трудно развить инициативу, чтобы что-то сделать.

Каждый элемент был представлен по одному в случайном порядке для каждого нового участника вместе с 4-балльной рейтинговой шкалой, в которой пользователя просили указать, как часто это относилось к ним за последнюю неделю, где

1 = Не относился ко мне вообще

2 = Относится ко мне в некоторой степени или иногда

3 = Применяется ко мне в значительной степени или большую часть времени

4 = Применялся ко мне очень часто или большую часть времени

X38 Пол:

1 – мужской

2 – женский

X39 Возраст

X40 Семейное положение

1 = никогда не был в браке

2 = в настоящее время женат

3 = был ранее женат

Первые 37 параметров – это ответы на психологический тест, 38-40 параметры – общие данные.

Для обучения нейронной сети была взята часть множества с сайта kaggle.com.

Множество разбито на два подмножества:

1. Обучающее, содержащее 2005 примеров

2. Тестирующее, содержащее 390 примеров

Проектирование, обучение и тестирование происходило в программе «Нейросимулятор 5» [5] по методике [6,7]. Оптимальная функция активации сети представляет собой гиперболический тангенс. Нейросеть состоит из 40 входных нейронов, 7 нейронов на скрытом слое и 1 выходного нейрона.

Чтобы проверить нейросеть на корректность, использовалось тестирующее множество, состоящее из 390 примеров. Средняя относительная ошибка тестирования составила 5%, что можно считать приемлемым результатом. Ниже представлена разница между ожидаемыми и прогнозируемыми нейросетью значениями.

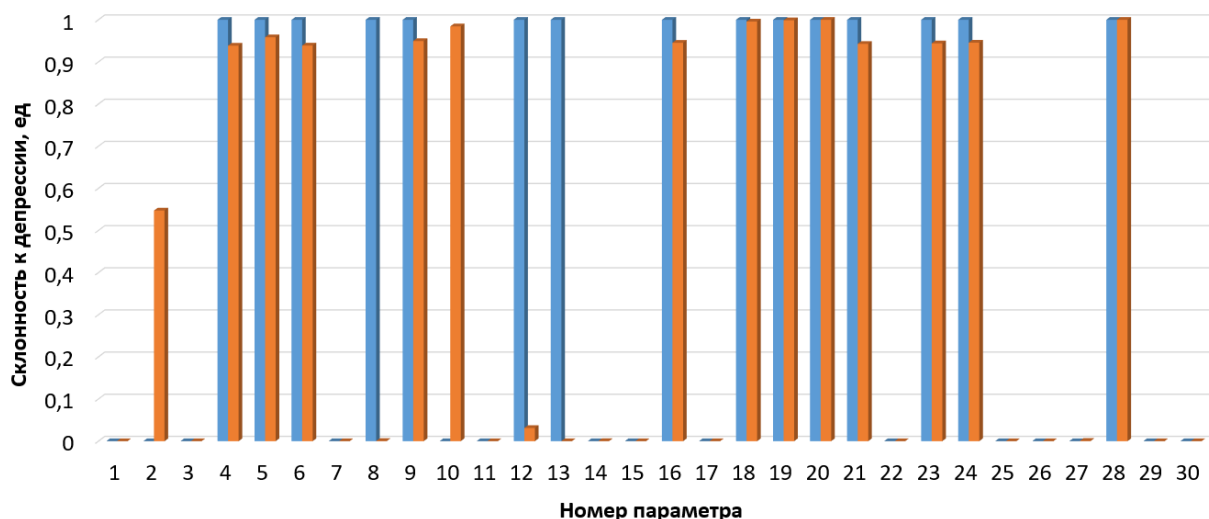


Рисунок 1. Результаты тестирования.

Оценка значимости параметров изображена ниже.

Как видно на графике, самыми значимыми параметрами являются ответы на вопросы:

15: Я чувствовал себя довольно обидчивым

12: У меня было чувство дурноты

16: Я заметно потел при отсутствии высоких температур или физических нагрузок

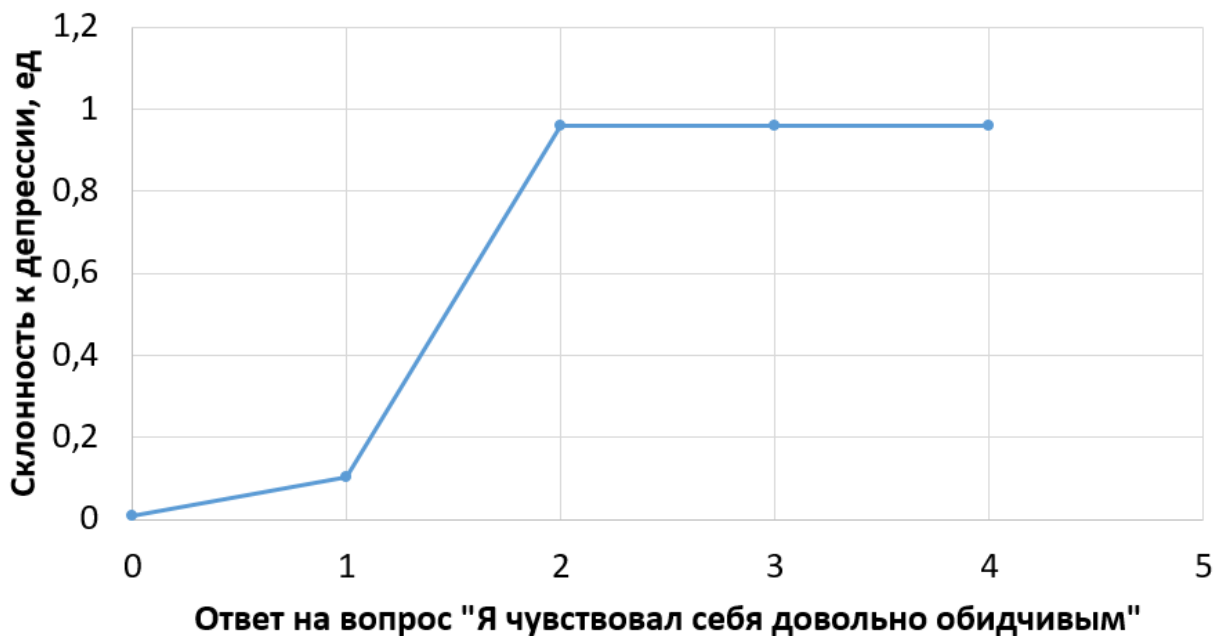
40: Семейное положение



**Рисунок 2. Значимость параметров**

Из этого можно сделать вывод, что депрессия как заболевание психики, может иметь различные симптомы и причины. Из этого следует, что существует большое количество вариантов депрессии, в зависимости от проявляемых симптомов.

Ниже представлена зависимость склонности к депрессии, от ответа на вопрос: «Я чувствовал себя довольно обидчивым».



**Рисунок 3. Склонность к депрессии в зависимости от обидчивости**

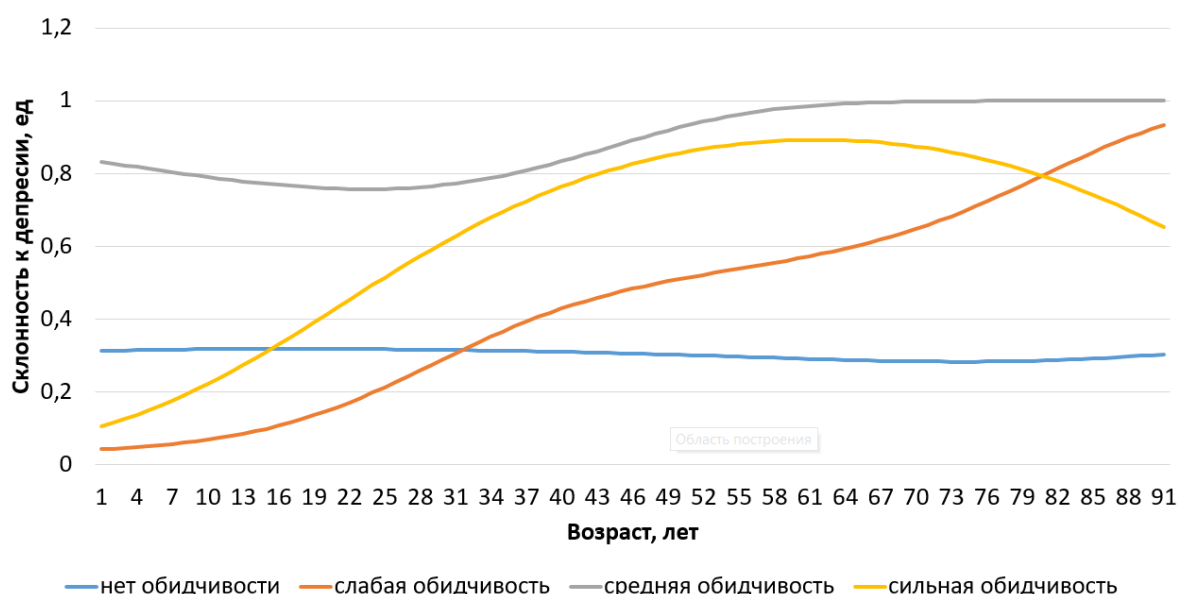
Как можно видеть на графике, при выборе варианта ответа «1», то есть отсутствия обидчивости, человек почти не склонен к депрессии. Но если у человека есть хотя бы небольшое проявление обидчивости (варианты ответа «2», «3», «4»), то склонность к депрессии довольно высока. Из этого можно

сделать вывод, что обидчивость является одним из ключевых симптомов депрессии, при ее наличии стоит обратиться к специалисту.

В качестве метода исследования зависимостей был выбран метод замораживания. Исследуемые параметры менялись в некотором диапазоне, остальные не изменялись. Ниже приведены входные данные с наиболее значимыми параметрами.

X1	X7	X10	X12	X13	X15	X16	X17	X19	X21	X30	X35	X37	X39	X40
3	4	4	2	3	4	2	2	2	3	2	2	2	20	1

Ниже представлен график влияния проявления обидчивости на склонность к депрессии.



**Рисунок 4. Влияние обидчивости на склонность к депрессии с возрастом**

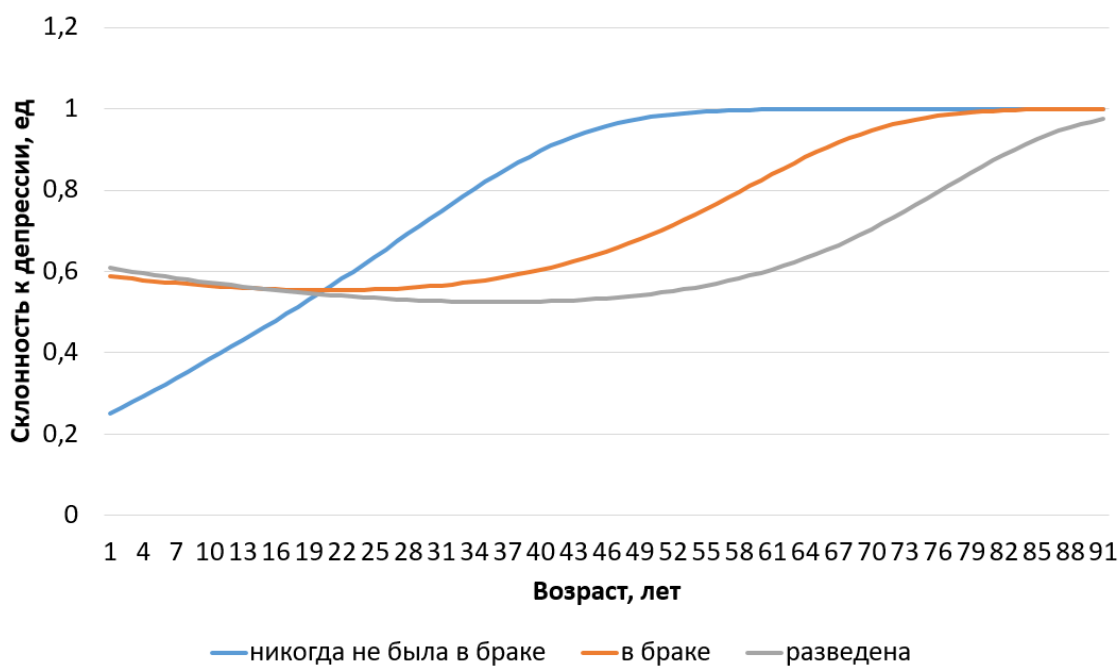
Как видно из графика, отсутствие обидчивости не влияет на склонность к депрессии. В случаях слабой, средней и сильной обидчивости с возрастом наблюдается значительный рост депрессии, особенно это проявляется при средней и сильной обидчивости.

Немаловажную роль так же играют возраст и семейное положение человека. Ниже представлены различные зависимости возраста и семейного положения к склонности к депрессии.

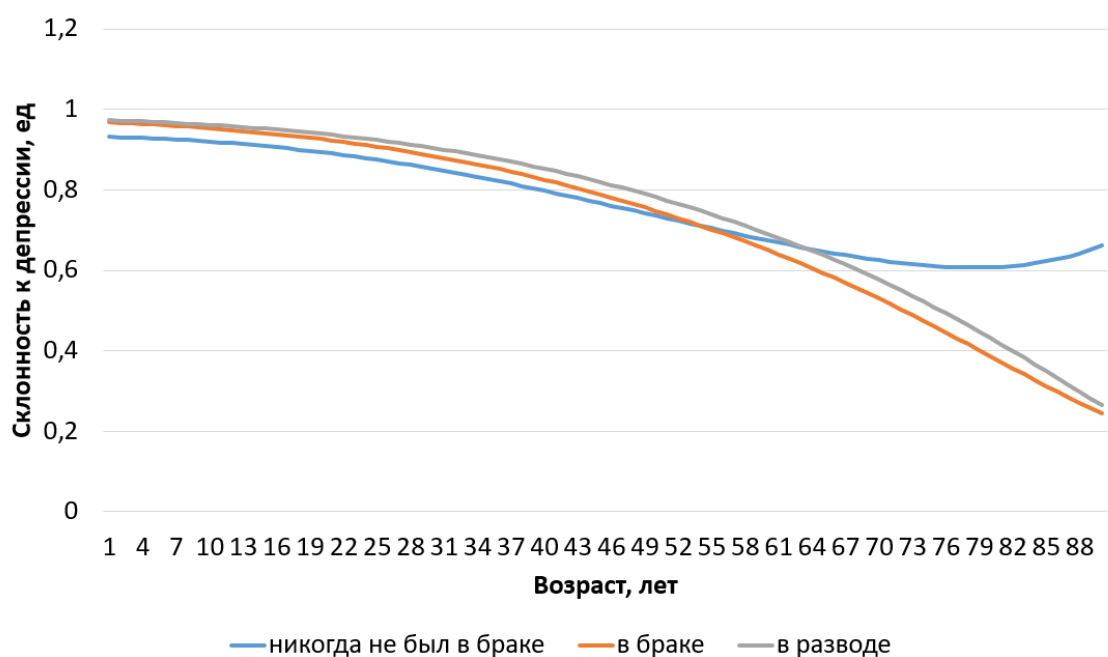
Как можно видеть на графике, наиболее склонны к депрессии те женщины, которые никогда не были в браке. В случаях, когда женщина находится в браке или, когда женщина разведена, наблюдается некоторое уменьшение склонности депрессии к 25-28 и 40-46 годам соответственно, но далее с возрастом наступает лишь рост склонности к депрессии. К 91 году у всех женщин высокая склонность к депрессии.

Как видно на графике, склонность мужчин к депрессии, относительно возраста и семейного положения, уменьшается одинаково до 52 лет. Далее склонность к депрессии мужчин, которые находятся в браке или в разводе, падает, а у мужчин, которые никогда не были в браке растет.





**Рисунок 5. Склонность женщин различного семейного положения и возраста к депрессии**



**Рисунок 6. Склонность мужчин различного семейного положения и возраста к депрессии**

Исходя из рисунка 4 и рисунка 5, можно сделать следующие выводы:

- 1) И мужчины, и женщины склонны к депрессии, если никогда не были в браке. Мужчины склонны в меньшей степени
- 2) Склонность к депрессии у женщин в браке с возрастом растет, у мужчин, наоборот, падает
- 3) Склонность к депрессии женщин в разводе с возрастом растет, у мужчин падает

**Заключение.** Построена система нейросетевого прогнозирования склонности человека к депрессии. Спроецированная нейросетевая модель учитывает 40 входных параметров, а именно 37 ответов на вопросы психологического теста, возраст, пол и семейное положение. Методом сценарного прогнозирования построены графики зависимостей склонности к депрессии от возраста, семейного положения и пола. Построены график важности входных параметров и график влияния наиболее значимого параметра – обидчивости. Применение такого набора параметров позволяет с высокой точностью определять склонность человека к депрессии.

### **Библиографический список**

1. Ясницкий Л.Н., Левченко Е. В., Митрофанов И. А.. Нейросетевое моделирование феномена депрессии //Наука и глобальные вызовы XXI века. 2018. С. 139.
2. Депрессия: распространённость, классификация. [Электронный ресурс]. Режим доступа: <https://neuroprofi.ru/articles/depressiya-kak-vylechit/>
3. Депрессия. [Электронный ресурс]. Режим доступа: <https://www.medicina.ru/patsientam/zabolevanija/depressiya/>
4. Прогнозирование депрессии, беспокойства и стресса – EDA. [Электронный ресурс]. Режим доступа: <https://www.kaggle.com/yamqwe/predicting-depression-anxiety-and-stress-eda/notebook>
5. Черепанов Ф.М., Ясницкий Л.Н. Нейросимулятор 4.0. Свидетельство о регистрации программы для ЭВМ RUS 2014612546. Заявка № 2014610341 от 15.01.2014.
6. Ясницкий Л.Н. Введение в искусственный интеллект. М.: Издательский центр «Академия», 2005. – 176 с.
7. Ясницкий Л.Н., Грацилёв В.И., Куляшова Ю.С., Черепанов Ф.М. Возможности моделирования предрасположенности к наркозависимости методами искусственного интеллекта // Вестник Пермского университета. Философия. Психология. Социология. 2015. № 1 (21). С. 61-73.

## **NEURAL NETWORK SYSTEM FOR PREDICTION OF A HUMAN'S TENDENCY TO DEPRESSION**

*Krasnykh Roman S.*

Perm State University, Str. Bukireva, 15, Perm, Russia,  
614990, roman.krasnykh02@gmail.com

The article presents a description of the development of a neural network system for predicting a person's tendency to depression. The system allows a person's susceptibility to depression based on a psychological test and general information about the person.

**Keywords:** artificial intelligence, neural network technologies, forecasting, depression.

## ПРЕДСКАЗАНИЕ СИМПТОМОВ ДЕПРЕССИИ НА ОСНОВЕ СОЦИО-ДЕМОГРАФИЧЕСКИХ И ПОВЕДЕНЧЕСКИХ ПОКАЗАТЕЛЕЙ С ИСПОЛЬЗОВАНИЕМ НЕЙРОННЫХ СЕТЕЙ

*Фаизова Лилия Тимуровна*

Национальный исследовательский университет  
«Высшая школа экономики» – Санкт-Петербург,  
194100, г. Санкт-Петербург, ул. Кантемировская, 3, корп.1, лит. А,  
lfaizova@edu.hse.ru

В статье представлено описание разработки нейросетевой системы для предсказания симптомов депрессии у пациентов. Система позволяет с большой точностью предсказать наличие вышеперечисленных симптомов у респондента на основании данных, собранных количественным методом ранее в опросе. С помощью разработанной интеллектуальной системы проведено исследование предметной области, выявлены закономерности, имеющие практического значения.

**Ключевые слова:** депрессия, тревожность, прогнозирование, искусственный интеллект, нейросетевые технологии.

**Введение.** Депрессия считается одной из наиболее распространенных проблем психического здоровья, затрагивающих все население планеты (Weissman, 1996). Депрессия определяется как состояние пониженного настроения, связанное с плохой самооценкой, чувством неполноценности, одиночества и иногда суицидальными мыслями. Термин «большое депрессивное расстройство» (БДР) впервые был введен учеными в США в 1970-х годах, однако в качестве диагноза был введен только в 1980 году, когда и был включен в DSM-3. Текущее издание диагностического руководства — DSM-5, и оно является одним из основных инструментов, используемых для диагностики депрессивных расстройств. Тревожность же является проявлением нервного поведения, вызванное дискомфортом от определенной ситуации, реакцией на которое может стать приступ паники или бессмысленное однотипное повторение одного действия. Генерализованное тревожное расстройство (ГТР) также появилось как диагностическая категория в DSM-3 в 1980 году, когда тревожный невроз был разделен на ГТР и паническое расстройство. Оба этих диагноза самостоятельны в лечении и симптоматике, однако часто могут накладываться друг на друга по схожим характеристикам: в лечении обоих есть когнитивно-поведенческой терапия и схожие препараты (Kendall & Watson, 1989).

В данной работе представлена нейронная сеть, которая может предсказывать наличие симптомов депрессии у пациентов на основе их социо-

демографических и поведенческих показателях. В качестве обучающего множества была выбрана база данных, собранная ранее с помощью сервиса проведения опросов SurveyMonkey и опубликованная на сайте Kaggle.com [3]. Объем выборки составил 334 респондента, самостоятельно заполнивших онлайн анкету (примечательным является то, что диагноз респонденты давали себе самостоятельно, так что для использования работы в медицинских целях необходимо будет дополнить множество данными с профессиональными диагнозами). Из этого множества 80 человек имели одно из психическим заболеваний (24%), что немного больше показателя распространения психических заболеваний в мире, составляющим 22 % (Charlson et.al, 2019). Это дает основание полагать, что выборка приближена к реальной ситуации в мире.

Для моделирования нейронной сети были выбраны 18 показателей:

- X1 – возраст (18-29 – 1; 30-44 – 2; 45-60 – 3; >60 – 4),
- X2 – пол (0 – мужской, 1 – женский),
- X3 – самоидентификация на наличие психического заболевания (0 – нет, 1 – да),
- X4 – наличие собственного компьютера (0 – нет, 1 – да),
- X5 – регулярный доступ к Интернету (0 – нет, 1 – да),
- X6 – ранее был госпитализирован из-за психического расстройства (0 – нет, 1 – да),
- X7 – количество дней госпитализации (число),
- X8 – наличие постоянной работы (0 – нет, 1 – да),
- X9 – инвалидность (0 – нет, 1 – да),
- X10 – проживает ли человек с родителями (0 – нет, 1 – да),
- X11 – наличие пропусков в резюме (т.е. являлся безработным какое-то время) (0 – нет, 1 – да),
- X12 – количество пропусков в резюме (в месяцах),
- X13 – годовой доход (0-25 – 1; 25-50 – 2; 50-75 – 3; 75-100 – 4),
- X14 – читает ли человек литературу, не связанную с учебой или работой (0 – нет, 1 – да),
- X15 – есть ли доход из социальной поддержки (0 – нет, 1 – да),
- X16 – количество госпитализаций по причине психических заболеваний (число),
- X17 – отсутствие концентрации (0 – нет, 1 – да),
- X18 – усталость (0 – нет, 1 – да).

Все показатели X1-X18 приводят к выводу:

- D1 – наличие симптомов депрессии (0 – нет, 1 – да).

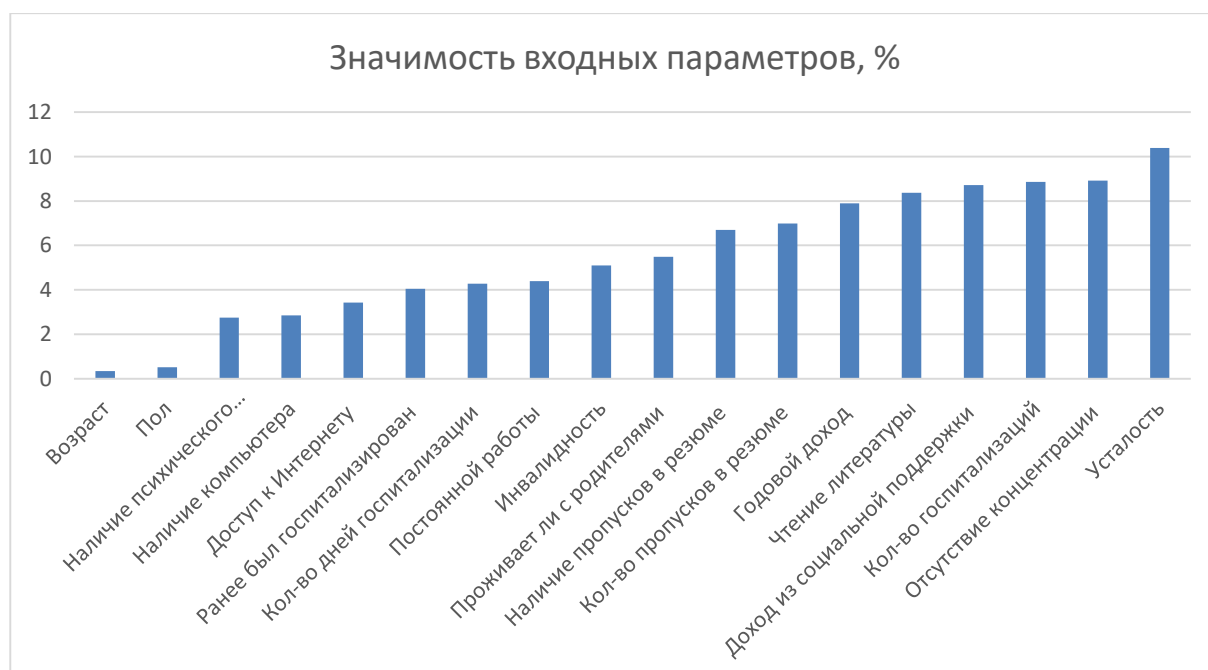
Проектирование, обучение и тестирование нейронной сети происходило в программе «Нейросимулятор 5» [6] по методике [7]. В результате оптимизации спроектированная нейронная сеть представляла собой перцептрон, состоящий из восемнадцати входных нейронов, двумя выходными и одним скрытым слоем с шестью нейронами. В качестве активационных функций нейронов скрытого слоя и выходного слоя использовался тангенс гиперболический, а в качестве алгоритма обучения – алгоритм упругого обратного рас-

пространения. Средняя квадратичная относительная ошибка тестирования составила 17,96%.



**Рисунок 1. Результат тестирования нейронной сети**

По результатам тестирования нейронной сети, изображенном на рис. 1, можем сделать вывод, что нейросетевая модель работает. Заметим, что в случаях с 4, 15, 19 и 20 респондентами модель выдала противоположный правильному результат, в остальных же случаях с небольшими погрешностями модель выдала правильный результат.



**Рисунок 2. Значимость входных параметров нейросетевой модели**

Оценка значимости входных параметров была посчитана с помощью программы «Нейросимулятор 5», результат показан на рис. 2. Как видно из

диаграммы, наиболее значимым параметром для прогнозирования депрессии и тревожности является усталость (10,4%), что неудивительно, ведь последствия выгорания и хронической усталости включают снижение трудоемкости и общее снижение удовлетворенности работой и личной жизнью (Messias, E., Flynn, V., 2018). Далее по убыванию идут такие параметры, как отсутствие концентрации (8,9%), количество госпитализаций из-за психических заболеваний (8,9%), наличие дохода из социальной поддержки (8,7%) и чтение литературы, не связанной с работой или обучением (8,4%). Отдельно отмечу маленькую значимость таких параметров, как пол (0,5%), и возраст (0,3%), откуда можно сделать вывод, что демографические показатели не оказывают влияния на проявление симптомов депрессии и тревожности у пациентов.

Далее были проведены исследования каждого входного параметра с симптомами депрессии. Для этих целей был использован метод «замораживания» (Ясницкий, Черепанов, 2012), который заключается в варьировании значения одного параметра при неизменных остальных. Далее для эксперимента было отобрано 5 пациентов, данные которых приведены ниже в таблице 1.

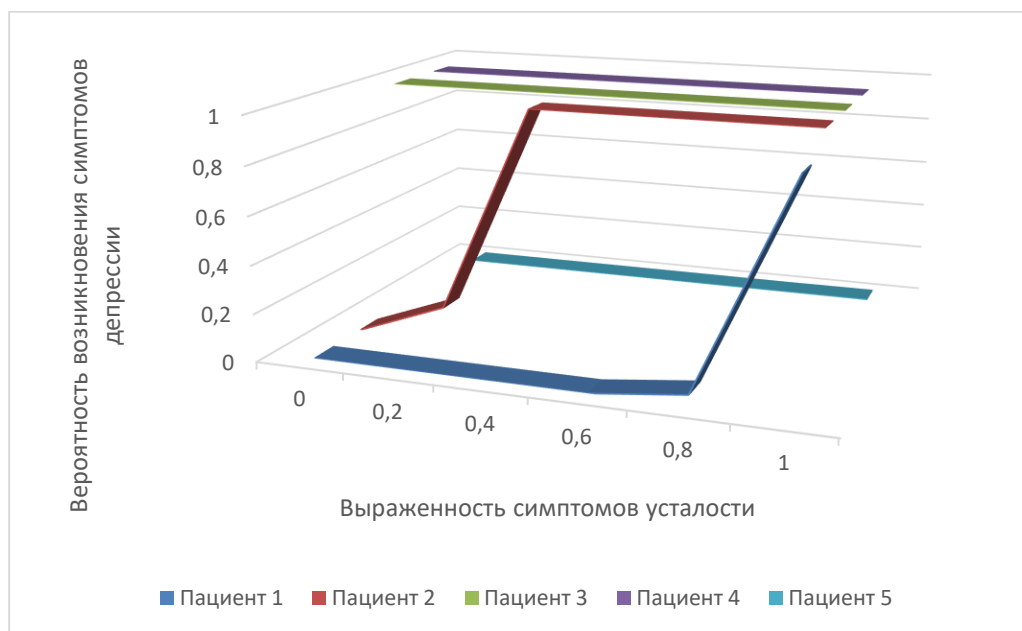
**Таблица 1**

**Характеристики пациентов, выбранных для исследования**

	Пациент 1	Пациент 2	Пациент 3	Пациент 4	Пациент 5
X1	1	1	2	1	3
X2	0	1	1	1	1
X3	0	0	1	1	0
X4	1	0	1	1	1
X5	1	1	1	1	1
X6	0	0	1	0	0
X7	0	0	15	0	0
X8	1	1	1	1	0
X9	0	0	0	0	0
X10	0	1	0	0	0
X11	0	1	0	1	0
X12	0	11	0	6	0
X13	3	3	1	2	1
X14	1	1	1	1	1
X15	0	0	0	0	0
X16	0	0	1	0	0
X17	1	0	1	1	0
X18	1	1	1	1	1
D1	0,9079	1	1	1	0

На рисунке 3 изображена зависимость возникновения симптомов депрессии от выраженности симптомов усталости. Основываясь на этой информации, можно сделать вывод, что степень выраженности симптомов усталости имеет большое влияние на вероятность возникновения депрессии. Пациенты 3 и 4, однако, подвержены депрессии вне зависимости от выраженности симптомов усталости, это объясняется тем, что пациент 3 имеет

другие психические заболевания и был госпитализирован в психиатрические лечебные заведения, а у пациента 4 четко выражены другие значимые для депрессии признаки, как отсутствие концентрации и большое количество пропусков в резюме. Пациент 5, напротив, наименее подвержен депрессии, и при повышении выраженности симптомов усталости вероятность возникновения депрессии у него повышается всего на 0,2%. В случаях же пациентов 1 и 2 можно заметить прямую зависимость вероятности депрессии от выраженности симптомов усталости: разными темпами оба показателя неукоснительно поднимаются вверх.

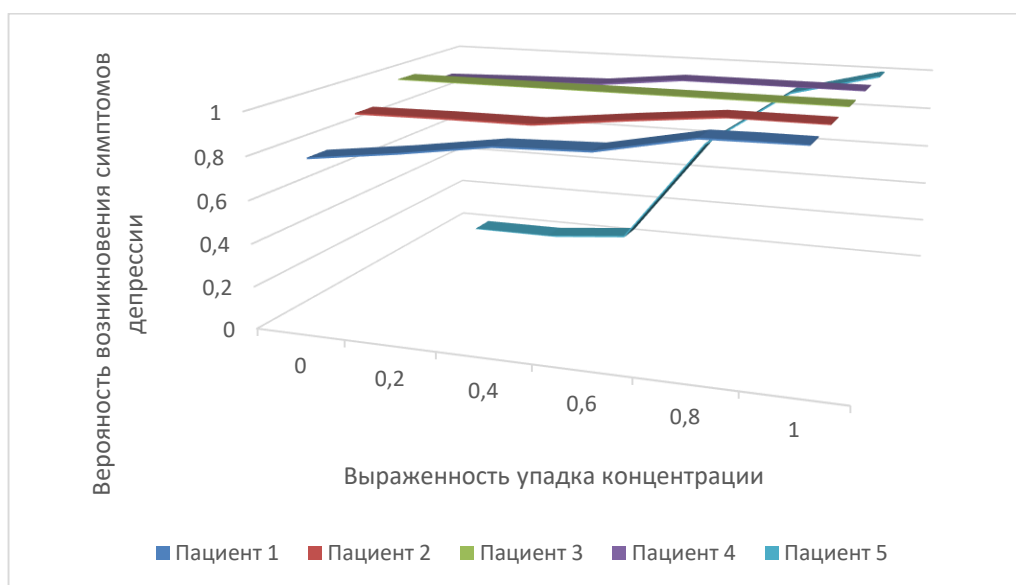


**Рисунок 3. Зависимость возникновения симптомов депрессии от симптомов усталости**

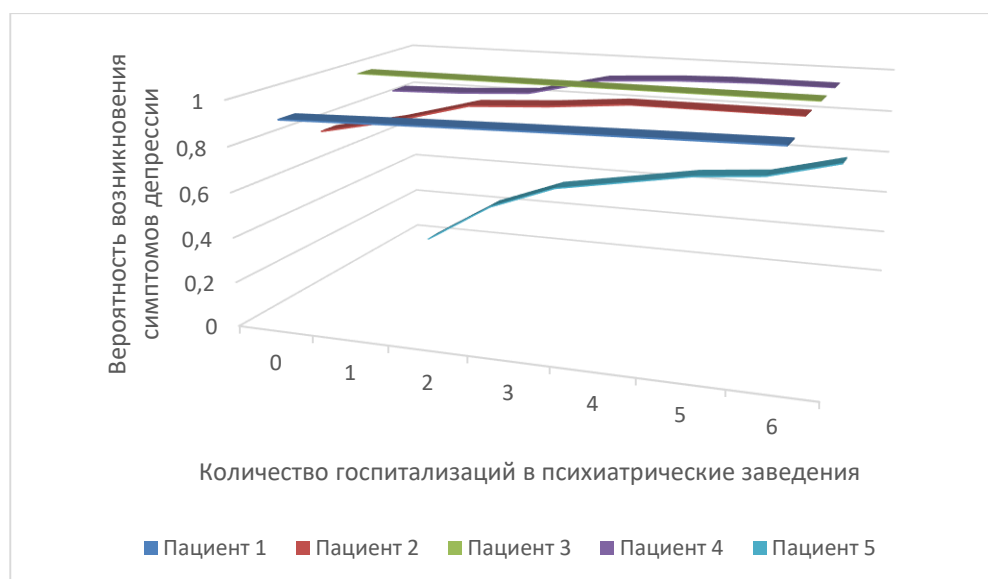
Рисунок 4 демонстрирует зависимость возникновения симптомов депрессии от выраженности упадка концентрации, обозначенном на горизонтальной оси. Здесь также можно наблюдать прямую зависимость этих параметров, так как, судя по графику, чем менее концентрированным чувствует себя человек, тем более вероятность у него возникновения симптомов депрессии. Опять же, это не касается пациента 3, так как он был ранее госпитализирован в лечебное психиатрическое заведение, а также имеет другие заболевания, связанные с психикой, – его вероятность депрессии при любом уровне концентрации составляет 100%. Самый большой подъем вероятности депрессии наблюдается у пациента 5, что может говорить о его изначальной малой подверженности этому заболеванию и показателях, которые свидетельствуют об отсутствии проблем, связанных с психикой. Пациенты же 1, 2 и 4 демонстрируют медленное повышение вероятности появления депрессии с уменьшением уровня концентрации.

На рисунке 5 изображена зависимость возникновения депрессии от количества госпитализаций, связанных с психическими заболеваниями. Стоит отметить, что для изменения параметра X16 (количество госпитализаций в

психиатрические заведения) было необходимо изменить и параметр X6 (был ли ранее пациент госпитализирован в психиатрические заведения) с 0 («нет») на 1 («да») у всех пациентов, кроме 3, так как он уже был ранее госпитализирован в психиатрическую больницу. В остальном график показывает, что чем больше количество госпитализаций, тем более вероятна депрессия. На горизонтальной оси обозначено количество госпитализаций, и можно заметить, что при условии 6 госпитализаций в психиатрические лечебницы 4 из 5 пациентов с почти 100% вероятностью имеют симптомы депрессии. Исключением является лишь Пациент 5, вероятность депрессии которого на 6-ю госпитализацию составляет лишь 58%.



**Рисунок 4. Зависимость возникновения симптомов депрессии от упадка концентрации**



**Рисунок 5. Зависимость возникновения симптомов депрессии от количества госпитализаций в лечебные психиатрические заведения**



Полученные в данном исследовании результаты не противоречат исследованиям, посвященным симптоматике депрессии, и медицинским знаниям. Таким образом, спроектированную нейронную сеть можно использовать для прогнозирования депрессии у людей, имеющих риск развития этой болезни. Однако, как уже было ранее отмечено, для более точного функционирования нейросетевой модели необходима выборка с диагнозами от медицинских учреждений.

**Заключение.** Разработана нейросетевая математическая модель, использованная для прогнозирования вероятности возникновения депрессии у пациентов. Эта нейросетевая модель дает представление о влиянии не только демографических и генетических параметров на возникновение заболевания, но также и поведенческих, социальных, а также культурных показателей, которые были обозначены в выборке. На примере представленных опросом данных была обучена нейронная сеть, а также были получены графики, доказывающие ее исправность и предсказательную способность. Эти знания могут быть использованы медицинскими или образовательными учреждениями для определения предрасположенности пациентов к такому психическому заболеванию, как депрессия.

### Библиографический список

1. Charlson, F., van Ommeren, M., Flaxman, A., Cornett, J., Whiteford, H., & Saxena, S. (2019). New WHO prevalence estimates of mental disorders in conflict settings: a systematic review and meta-analysis. *The Lancet*, 394(10194), 240-248. Weissman, M. M., Bland, R. C., Canino, G. J.,
2. Faravelli, C., Greenwald, S., Hwu, H. G., ... & Yeh, E. K. (1996). Cross-national epidemiology of major depression and bipolar disorder. *Jama*, 276(4), 293-299.
3. Kaggle [Электронный ресурс]. – URL: <https://www.kaggle.com/michaelacorley/unemployment-and-mental-illness-survey> (дата обращения: 03.03.2021)
4. Kendall, P. C., & Watson, D. E. (1989). *Anxiety and depression: Distinctive and overlapping features*. Academic Press.
5. Messias, E., & Flynn, V. (2018). The tired, retired, and recovered physician: professional burnout versus major depressive disorder. *American Journal of Psychiatry*, 175(8), 716-719.
6. Черепанов Ф.М., Ясницкий Л.Н. Нейросимулятор 4.0. Свидетельство о регистрации программы для ЭВМ RUS 2014612546. Заявка № 2014610341 от 15.01.2014.
7. Ясницкий Л.Н., Грацилёв В.И., Куляшова Ю.С., Черепанов Ф.М. Возможности моделирования предрасположенности к наркозависимости методами искусственного интеллекта // Вестник Пермского университета. Философия. Психология. Социология. 2015. № 1 (21). С. 61-73.

# PREDICTION OF DEPRESSION SYMPTOMS ON THE BASIS OF SOCIO-DEMOGRAPHIC AND BEHAVIORAL INDICATORS USING NEURAL NETWORKS

*Faizova Liliia*

National Research University "Higher School of Economics"  
194100, St. Petersburg, st. Kantemirovskaya 3, building 1, lit. A,  
lfaizova@edu.hse.ru

The article describes the development of a neural network system aimed for predicting symptoms of depression in patients. The system makes it possible to predict the presence of the symptoms mentioned above based on respondent data collected by a quantitative method earlier in the survey. With the help of the developed intellectual system, a study of the subject area was carried out, and patterns of practical importance were identified.

**Keywords:** depression, prediction, artificial intelligence, neural network technologies.

УДК 004.032.26

## НЕЙРОСЕТЕВАЯ СИСТЕМА ОПРЕДЕЛЕНИЯ ЗАВИСИМОСТИ ОТ СЕРИАЛОВ

*Шиляев Вадим Викторович*

Пермский государственный национальный исследовательский университет,  
ПМИ. 614990, Россия, г. Пермь, ул. Букирева, 15, yasn@psu.ru

В статье представлено описание разработки нейросетевой системы для определения и прогнозирования зависимости от сериалов. Система позволяет определить зависимость от сериалов у человека с учетом его пола, успеваемости, способности планировать свое время, времени нахождения на свежем воздухе и степени одиночества.

**Ключевые слова:** искусственный интеллект, нейросетевые технологии, прогнозирование, прогнозирование зависимости от сериалов, сериаломания.

**Введение.** В современном мире многие люди увлекаются просмотром сериалов. Но зачастую люди не осознают, что стали зависимыми от сериалов, до тех пор, пока у них не появляются проблемы с учебой или с работой. Например, у студентов может снизиться успеваемость из-за того, что они стали зависимыми от сериалов. Таким образом, очень важно понять есть ли у вас зависимость от сериалов.

Постараемся по максимуму исключить человеческий фактор из оценки зависимости от сериалов, для этого воспользуемся интеллектуальными системами [1].

Цель данной работы заключается в создании и обучении нейросетевой модели [2] на основе собранных данных. Результатом исследования является система оценки зависимости от сериалов.

В результате анализа литературных источников [3-5] был выбран следующий набор показателей: X1 – пол, X2 – успеваемость, X3 – способность планировать свое время, X4 – степень одиночества, X5 – время нахождения на свежем воздухе, D1 – процент зависимости от сериалов.

Параметр X1 будет закодирован: 1 – мужской, 2 – женский. Параметр X2 изменяется от 2 до 5, где 2 – имеет долги, 3 – средний балл от 3.0 до 3.5, 4 – средний балл от 3.5 до 4.5 и 5 – средний балл 4.5 и выше. Параметр X3 будет закодирован: 1 – планирует свое время, 2 – не планирует свое время. Параметр X4 будет закодирован: 1 – полностью одинок, 2 – проводит время с семьей, 3 – проводит время с друзьями, 4 – проводит время с любимым человеком, 5 – никогда не чувствует себя одиноким. Параметр X5 будет закодирован: 1 – проводит на свежем воздухе менее 30 минут, 2 – проводит на свежем воздухе 30 минут и более.

Выходной параметр – процент зависимости от сериалов изменяется в пределах от 0 до 100.

Для получения обучающего и тестирующего множества были опрошены более 80 человек. После обработки данных и исключения противоречивых примеров, выбросов и дубликатов получилось 62 примера для обучающего множества и 8 примеров для тестирующего множества.

Проектирование, обучение, тестирование нейронной сети выполнялись с помощью программы «Нейросимулятор 5» [6] по методике [8]. В результате оптимизации спроектированная нейронная сеть представляла собой персептрон, который имеет пять входных нейронов, один выходной и один скрытый слой с шестью нейронами.

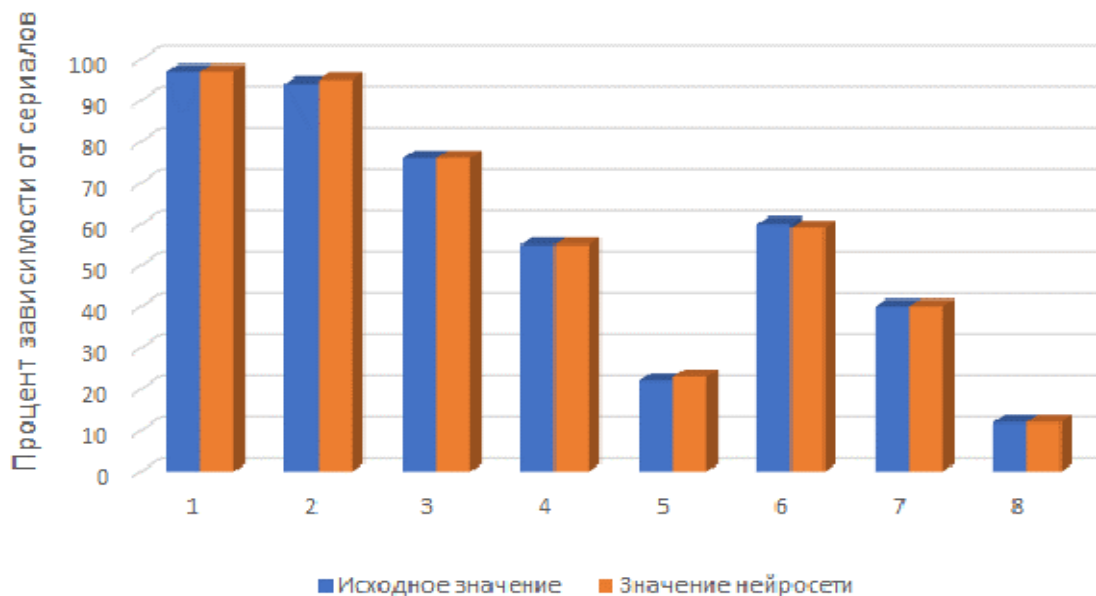


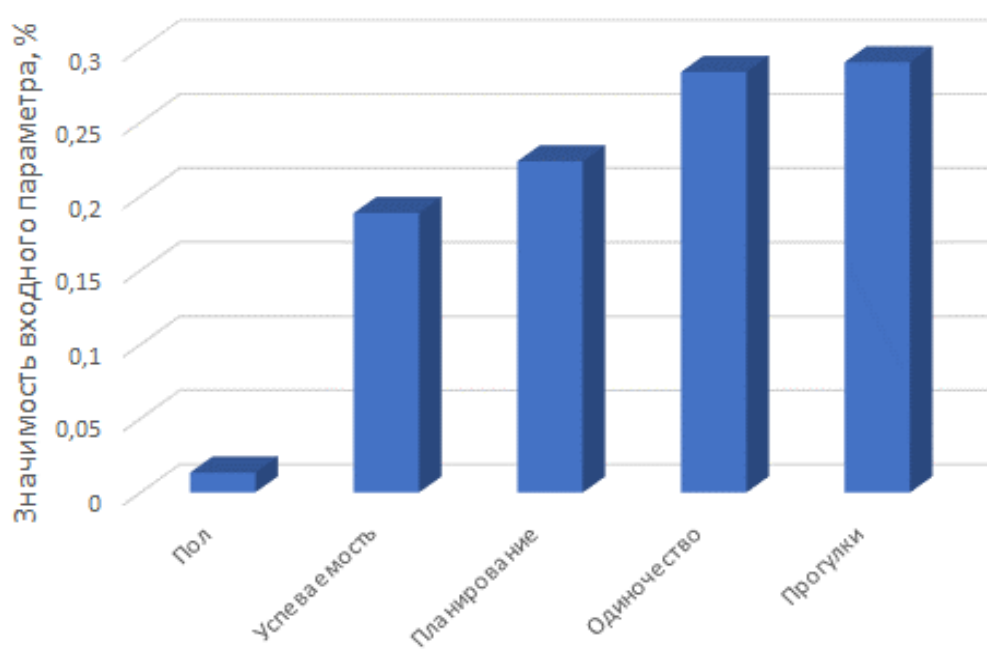
Рисунок 1. Результат тестирования нейронной сети

Для оценки корректной работы спроектированной нейронной сети использовалось тестирующее множество, состоящее из 8 примеров. Средняя ошибка тестирования составила 0,4%. На рисунке 1 представлена гистограмма, демонстрирующая разницу между фактическим и прогнозируемым нейросетью процентом зависимости от сериалов.

Из результатов, изображенных на рисунке 1, можно сделать вывод об адекватной работе спроектированной нейронной сети.

Оценка значимости входных параметров была выполнена с помощью программы «Нейросимулятор 5» [6], результат отображен на рисунке 2.

Как видно из рисунка 2, наиболее значимыми параметрами являются Прогулки, Одиночество, Планирование и Успеваемость. Также на графике есть параметр, который имеет около нулевую значимость это параметр пол.



**Рисунок 2. Значимость входных параметров нейросетевой модели**

Для исследования значимости параметра пол был взят следующий пример:

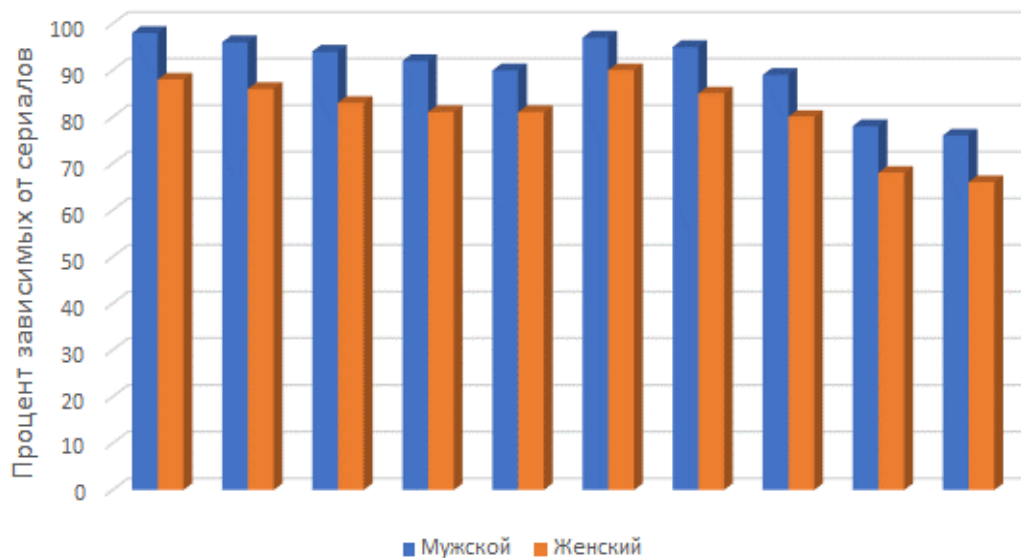
**Таблица 1**

**Исследование значимости параметра пол**

№ студента	X1	X2	X3	X4	X5
1	1	2	2	1	1
2	2	2	2	1	1
3	1	2	2	2	1
4	2	2	2	2	1
5	1	2	2	4	1
6	2	2	2	4	1
7	1	2	2	5	1
8	2	2	2	5	1
9	1	2	2	2	2
10	2	2	2	2	2

№ студента	X1	X2	X3	X4	X5
11	1	2	2	3	2
12	2	2	2	3	2
13	1	2	2	4	2
14	2	2	2	4	2
15	1	2	2	5	2
16	2	2	2	5	2
17	1	3	2	1	1
18	2	3	2	1	1
19	1	3	2	2	1
20	2	3	2	2	1

На рисунке 3 показан график зависимости процента зависимости от сериалов от пола студента. Как следует из графика, женский пол меньше подвержен зависимости от сериалов. Результат этого соответствует приведенным материалам [3].



**Рисунок 3. Зависимость процента зависимости от компьютерных игр от пола студента**

Для исследования значимости параметра успеваемость был взят следующий пример:

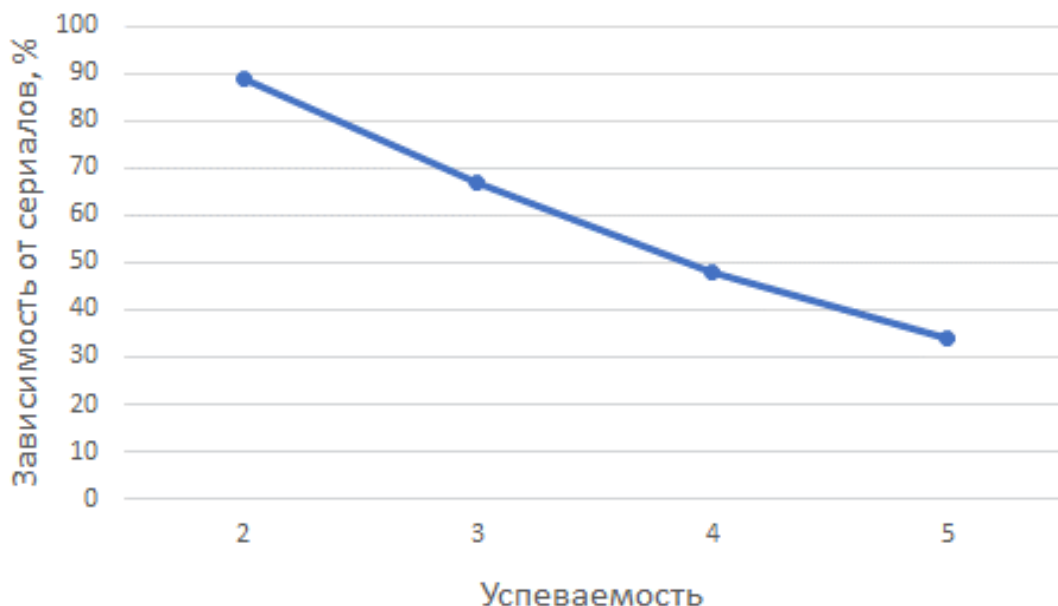
**Таблица 2**

**Исследование значимости параметра успеваемость**

№ студента	X1	X2	X3	X4	X5
1	1	2	2	5	2
2	1	3	2	5	2
3	1	4	2	5	2
4	1	5	2	5	2

На рисунке **Ошибка! Источник ссылки не найден.** продемонстрирована зависимость процента зависимости от сериалов от успеваемости. На

данном графике мы видим обратную зависимость, чем выше у человека успеваемость, тем меньше он зависит от сериалов. Данное утверждение соответствует реальности, т. к. люди, которые тратят много времени на учебу не могут тратить много времени на сериалы, а значит не могут быть зависимыми от сериалов.



**Рисунок 4. Зависимость процента зависимости от сериалов от успеваемости**

Для исследования значимости планирования был взят следующий пример:

**Таблица 3**

**Исследование значимости параметра планирования**

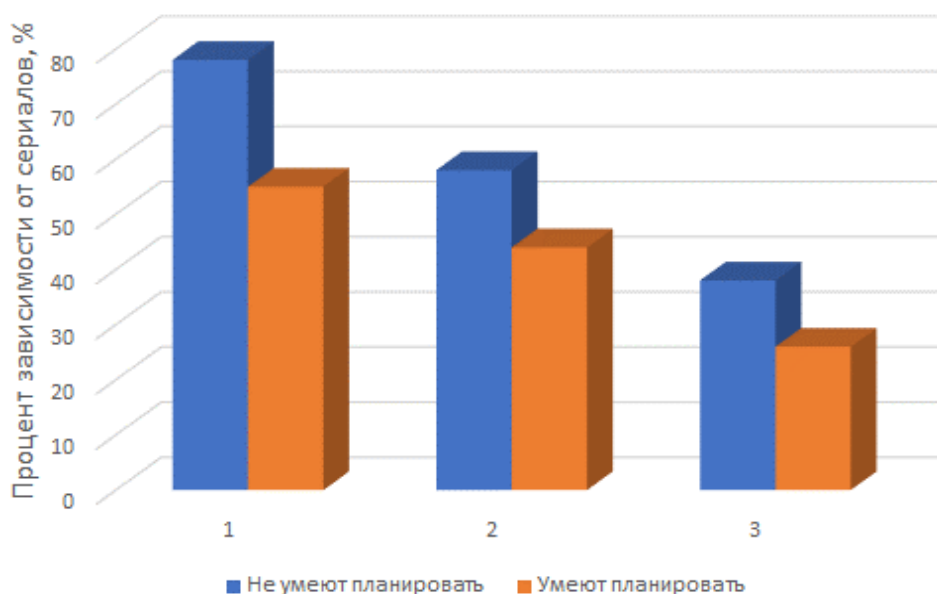
№ студента	X1	X2	X3	X4	X5
1	1	3	2	1	1
2	1	3	1	1	1
3	1	4	2	1	1
4	1	4	1	1	1
5	1	5	2	1	1
6	1	5	1	1	1

На рисунке 5 изображен график зависимость процента зависимости от сериалов от планирования своего времени. Из графика видно, что люди, которые умеют планировать свое время, имеют более низкую зависимость от сериалов. Данное утверждение соответствует реальности, т. к. люди, которые умеют планировать свое время, знают, когда надо остановиться и переключиться на другое дело.

Для исследования значимости одиночества был взят следующий пример.

На рисунке 6 изображен график зависимость процента зависимости от сериалов от одиночества. Из графика видно, что зависимость является обратной, то есть чем больше степень одиночества, тем меньше процент зависимости от сериалов. Данное утверждение соответствует реальности, т. к. одино-

кие люди чаще склонны к просмотру сериалов и отсюда получается зависимость от сериалов.

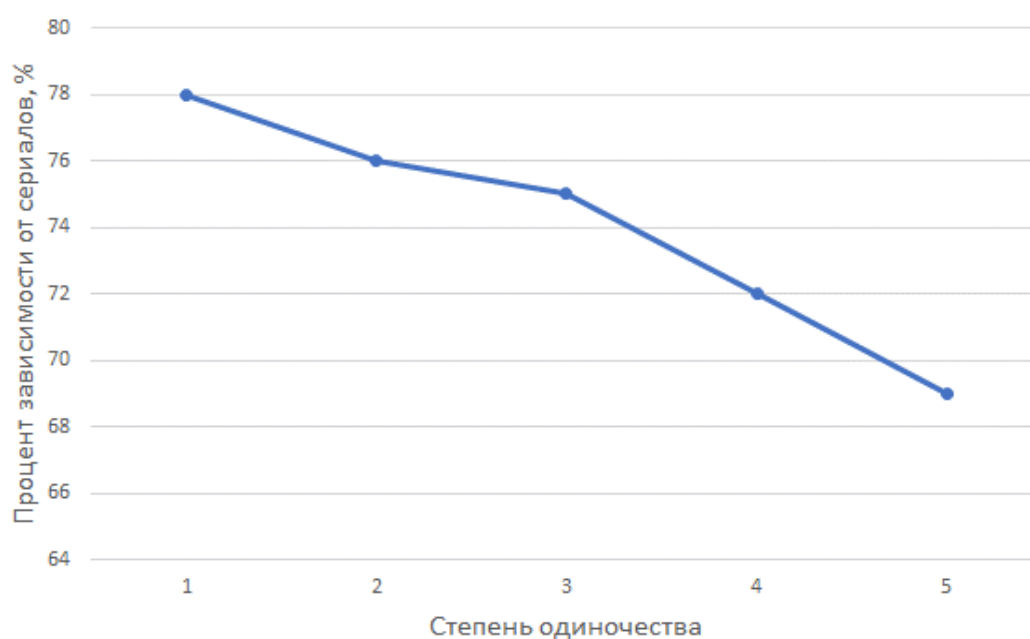


**Рисунок 5. Зависимость процента зависимости от сериалов от планирования**

**Таблица 4**

**Исследование значимости параметра одиночество**

№ студента	X1	X2	X3	X4	X5
1	1	3	2	<b>1</b>	1
2	1	3	2	<b>2</b>	1
3	1	3	2	<b>3</b>	1
4	1	3	2	<b>4</b>	1
5	1	3	2	<b>5</b>	1



**Рисунок 6. Зависимость процента зависимости от сериалов от одиночества**

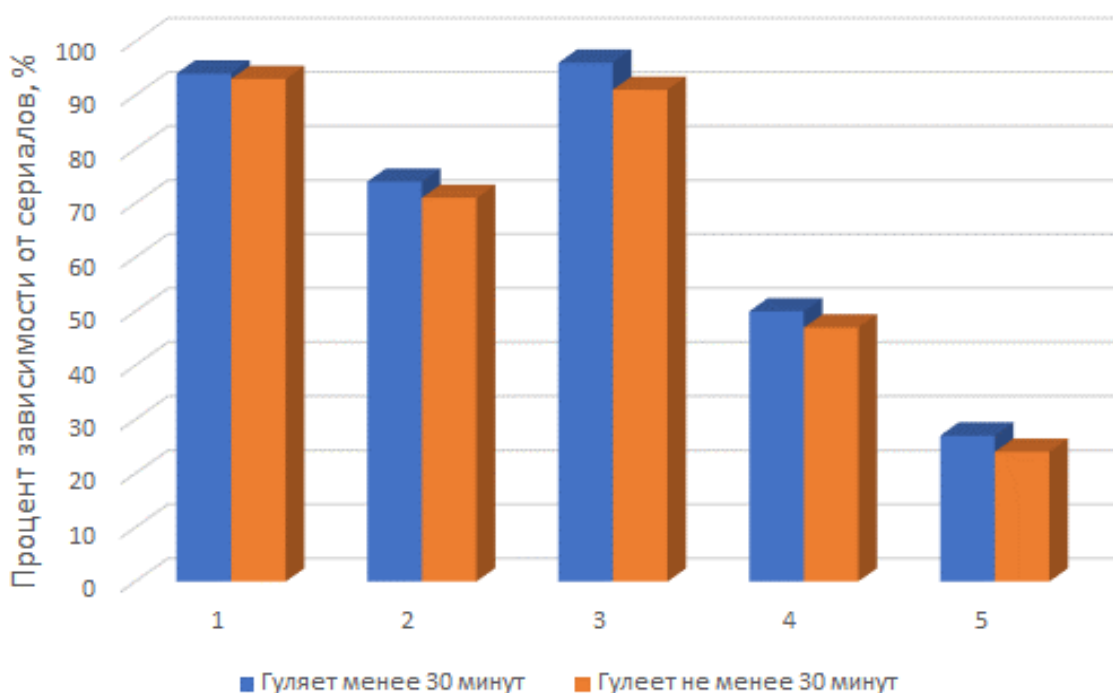
Для исследования значимости прогулок на свежем воздухе был взят следующий пример:

Таблица 5

**Исследование значимости параметра прогулка**

№ студента	X1	X2	X3	X4	X5
1	1	2	2	3	1
2	1	2	2	3	2
3	1	3	2	3	1
4	1	3	2	3	2
5	1	2	2	4	1
6	1	2	2	4	2
7	1	4	2	5	1
8	1	4	2	5	2
9	1	5	1	3	1
10	1	5	1	3	2

На рисунке 7 изображен график зависимость процента зависимости от сериалов от прогулок на свежем воздухе. Из графика видно, что люди, которые гуляют не менее 30 минут на свежем воздухе, имеют более низкую зависимость от сериалов.



**Рисунок 7. Зависимость процента зависимости от сериалов от прогулок на свежем воздухе**

Полученные результаты исследований не противоречат реальности, что подтверждает, что спроектированную нейронную сеть можно считать пригодной для определения уровня зависимости человека от сериалов.

**Заключение.** Построена система нейросетевого оценивания процента зависимости студента от сериалов. Спроектированная нейросетевая модель



обрабатывает параметры, характеризующие основные причины возникновения зависимости от сериалов. Проведены результаты оценки значимости параметров возникновения зависимости от сериалов, выполнена оценка ошибки обучения и тестирования. Полученные значения оценки ошибки тестирования хоть и являются небольшими, но сами данные являются крайне субъективными (так как каждый сам определяет для себя процент зависимости, но в тоже время определение процента зависимости основано на количестве времени, проведенном за просмотром сериалов, возможности прервать просмотр в любое время и выдержке на то, чтобы дождаться выхода новой серии). Для того, чтобы воспользоваться этой нейросетевой моделью, ее надо доработать, по максимуму исключив субъективные данные. Для этого нужны дополнительные исследования, которые помогут найти объективную оценку, а также новые входные данные.

### Библиографический список

1. Ясницкий Л.Н. Интеллектуальные системы. – М.: Лаборатория знаний, 2016.
2. Галушкин А.И. Нейронные сети: основы теории. М: Горячая линия–Телеком. 2012. 496 с.
3. Павлова О.С., Крупина М.М. Зависимость от иностранных сериалов – URL: <https://nsportal.ru/vuz/sotsiologicheskie-nauki/library/2018/03/18/issledovatel'skaya-rabota-zavisimost-ot-inostrannyh> (дата обращения 11.10.2021)
4. Закарян Г. Телесериал как инструмент воздействия на индивида – URL: <https://cyberleninka.ru/article/n/teleserial-kak-instrument-vozdeystviya-na-individa/viewer> (дата обращения 11.10.2021)
5. Акинфеев С.Н. Развлекательное телевидение: определение, классификация, жанры – URL: <https://cyberleninka.ru/article/n/razvlekatelnoe-televidenie-klassifikatsiya-opredelenie-zhanry/viewer> (дата обращения 11.10.2021)
6. MAËVA FLAYELLE, PIERRE MAURAGE, JOËL BILLIEUX Toward a qualitative understanding of binge-watching behaviors: a focus group approach – URL: <https://akjournals.com/view/journals/2006/6/4/article-p457.xml> (дата обращения 11.10.2021)
7. Черепанов Ф.М., Ясницкий Л.Н. Нейросимулятор 4.0. Свидетельство о регистрации программы для ЭВМ RUS 2014612546. Заявка № 2014610341 от 15.01.2014.
8. Ясницкий Л.Н. Интеллектуальные системы: учебник. М.: Лаборатория знаний, 2016. 221 с.

# NEURAL NETWORK SYSTEM FOR DETERMINING AND PREDICTING DEPENDENCE ON TV SERIES

*Shilyaev Vadim V.*

Perm State University, Str. Bukireva, 15, Perm, Russia, 614990, yasn@psu.ru

The article describes the development of a neural network system for determining and predicting dependence on TV series. The system allows you to determine a person's dependence on TV shows, taking into account his gender, academic performance, ability to plan his time, time spent outdoors and the degree of loneliness.

**Keywords:** artificial intelligence, neural network technologies, forecasting, forecasting dependence on TV series, serialomania.

УДК 004.855.5

## НЕЙРОСЕТЕВАЯ СИСТЕМА ПРОГНОЗИРОВАНИЯ СКЛОННОСТИ К СУИЦИДУ

*Трефилов Дмитрий Александрович*

Национальный исследовательский университет «Высшая школа экономики», 614060, Россия, г. Пермь, ул. Бульвар Гагарина, 37а, trefilowdem@mail.ru

В статье представлено описание разработки нейросетевой системы для прогнозирования склонности человека к самоубийству. Данная система позволяет на основе жалоб на состояние здоровья и убеждений о ценности жизни определить, есть ли у человека предрасположенность к суициду или нет. С помощью разработанной интеллектуальной системы проведено исследование предметной области, выявлены закономерные параметры, имеющие практическое значение. Разработанная модель может быть полезным инструментом для работы психологов и врачей.

**Ключевые слова:** искусственный интеллект, нейросетевые технологии, психологическое здоровье, выявление предрасположенности к суициду.

**Введение.** По статистике [4] в России за 2020 год в результате самоубийств погибло 16546 человека, из них 13 731 мужчин и 2 815 женщин. Самоубийство – это осознанный шаг человека лишиться себя жизни, и причины для такого поступка могут быть разными. По данным [5] основные мотивы для совершения суицида является психологическое окружение человека: семейные проблемы, разводы, супружеские измены, повторные браки, алкоголизм одного или нескольких членов семьи, конфликты между супругами, враждебность между членами семьи, болезни, потери родственников. Подобные факторы на прямую влияют на эмоциональное состояние человека и его

восприятие мира, тем самым толкая его на самоубийство, если не оказать ему своевременную помощь специалистов. Глядя на человека, нельзя явно сказать о том, нужна ли ему такая помощь или нет, это может определить только специальное тестирование и заключение врача.

Тем не менее, проблемы в семье достаточно сильно оказывает влияние на эмоциональное состояние не только взрослых, но и детей, заставляя формировать негативные ощущения и мысли, например, что ребенок никому не нужен. Ребенку или взрослому человеку со слабым эмоционально-психологическим здоровьем справиться с негативным окружением сложно. Выявление у таких людей начальных признаков к склонности к суициду крайне необходимо, поскольку это может послужить сигналом для срочного оказания им помощи специалистами.

При анализе литературных источников выяснилось, что работ на тему определения предрасположенности человека к совершению самоубийства достаточно много. К примеру, в работе [2] авторы на примере известных людей, совершившие самоубийство, определяют предрасположенность человека к такому шагу на основе личных данных: пол, возраст, семейное положение, сфера профессиональной деятельности, количество детей, финансовое положение и другие значимые параметры.

Предрасположенность людей к суициду может быть выявлена на основе их постов и записей на интернет-ресурсах. Так, [3] израильские ученые разработали искусственный интеллект, способный анализировать посты Facebook пользователей из США и определять предрасположенность человека к суициду. Система объединяет алгоритмы машинного обучения и обработки естественного языка с теоретическими и аналитическими инструментами из области психологии и психиатрии, а также использует многоуровневые нейросети.

В исследовании участвовали 1002 пользователя Facebook, которые прошли клинически обоснованный скрининг, выявляющий склонность к самоубийствам. Они добровольно предоставили ученым данные о своей активности в социальной сети за год. Как выяснилось, пользователи с суицидальными мыслями чаще используют негативные описания («плохой», «худший»), ругательства, выражения эмоционального расстройства («грустный», «больно», «плакать», «сумасшедший») и описания негативных физиологических состояний («больной», «боль», «хирургия», «больница»).

Цель данной работы заключается в создании нейросетевой модели, способной определить предрасположенность человека к самоубийству на основе его жалоб на здоровье, убеждений о ценности жизни и психоэмоционального состояния. Для этого были выбраны следующие параметры:

- X1 – проявление чувства скуки и усталости (0 – никогда, 1 – иногда, 2 – часто),
- X2 – Частота присутствия душевной боли (0 – никогда, 1 – иногда, 2 – часто),
- X3 – согласен с утверждением, что безвыходных ситуаций не бывает (0 – нет, 1 – да),
- X4 – любит себя (0 – нет, 1 – да),

- X5 – наблюдаются нарушения аппетита (0 – нет, 1 – да),
- X6 – испытывает одиночество (0 – никогда, 1 – иногда, 2 – часто),
- X7 – были попытки суицида (0 – нет, 1 – да),
- X8 – есть цели и планы в жизни (0 – нет, 1 – да),
- X9 – есть склонность опускать руки в кризисных ситуациях (0 – нет, 1 – да),
- X10 – есть прерывистый и тревожный сон (0 – нет, 1 – да),
- X11 – согласен с утверждением, что прерывание жизни избавляет человека от его проблем (0 – нет, 1 – не знаю, 2 – да),
  - X12 – посещают мысли «уснуть и не проснуться» (0 – никогда, 1 – иногда, 2 – часто),
  - X13 – периодичность заикливания на неудачах (0 – нет, 1 – иногда, 2 – да),
  - X14 – испытывает чувства вины (0 – нет, 1 – иногда, 2 – да),
  - X15 – были неудачи в жизни, когда казалось, что выхода нет (0 – нет, 1 – да)

Выходной параметр Y1 – предрасположенность к самоубийству (0 – нет предрасположенности, 1 – есть предрасположенность).

Обучающее множество было собрано на основе [1] веб-сайта определения склонности к суициду. Сбор информации, фиксирование ответов и результатов происходило вручную. Таким образом, в итоговом множестве собрано 50 примеров, из них 10 являются валидирующим множеством, 5 примеров – тестовыми, а остальные – обучающими.

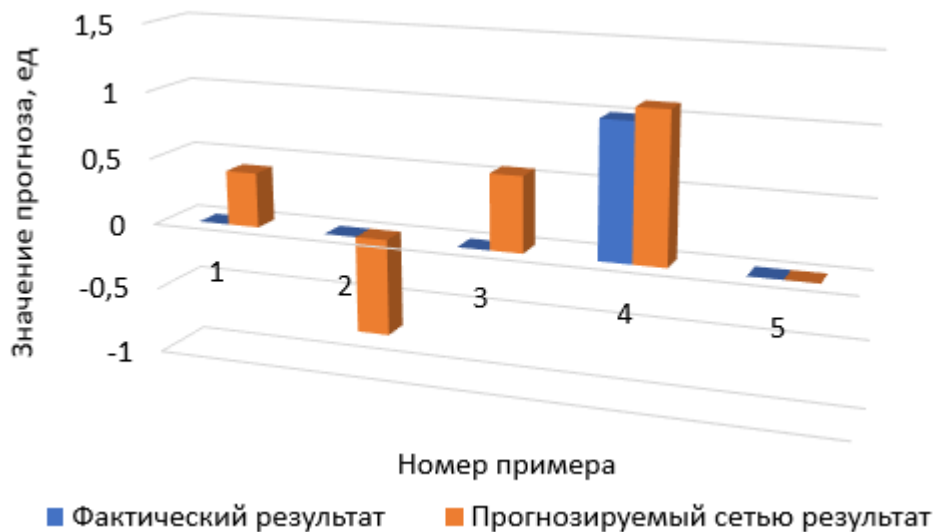
Процессы проектирования, обучения и тестирования нейронной сети выполнялись с помощью программы «Нейросимулятор 5». Полученная сеть содержит пятнадцать входных и один выходной нейрон, два скрытых слоя: на первом слое девять нейронов, на втором – пять.

Для оценки корректной работы спроектированной нейронной сети использовалось тестирующее множество, состоящее из 5 примеров. Средняя относительная ошибка тестирования для Y1 составила 35.6%. Такой результат возник из-за получения отрицательного значения, однако на всю картину прогноза это никак не влияет, поскольку получение отрицательного значения – признак 100% гарантии сети, что данный параметр должен быть равен 0. На рисунке 1 представлена гистограмма, демонстрирующая разницу между фактическими и прогнозируемыми нейросетью результатом.

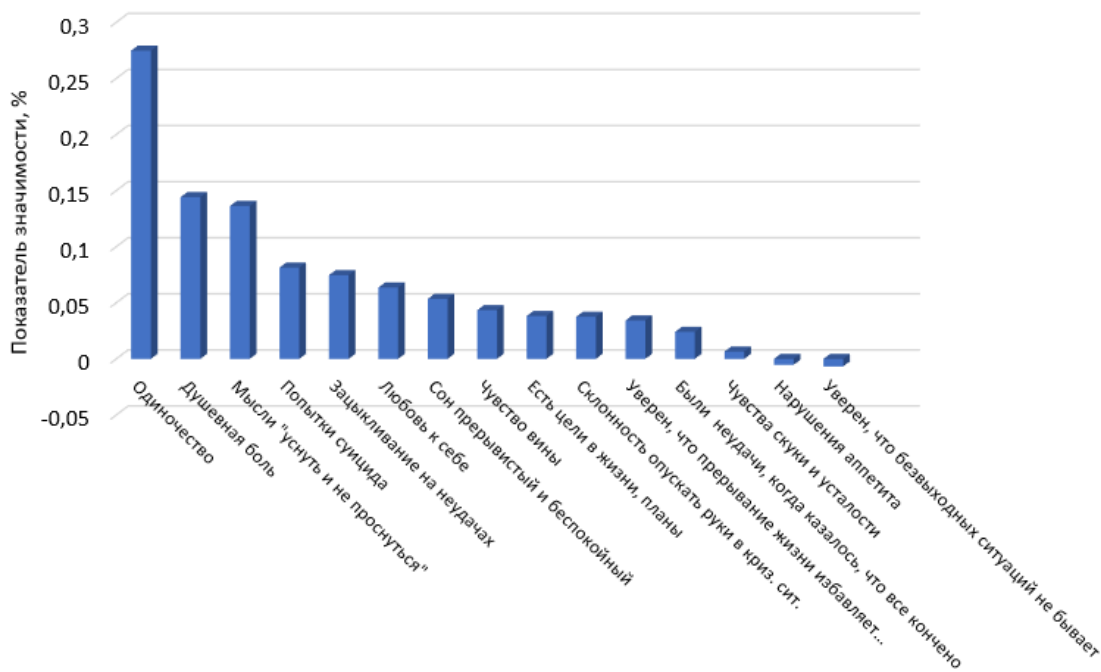
Из результата, изображенного на рисунке 1, можно сделать вывод об адекватной работе спроектированной нейронной сети. Оценка значимости входных параметров была выполнена с помощью программы «Нейросимулятор 5», результаты отображены на рисунке 2.

Как видно из рисунка 2, наиболее значимым параметром является наличие чувства одиночества. Действительно, испытывая данное чувство, человек теряет смысл существования, так как никому не нужен. Средние по значимости параметры – наличие душевной боли и мысли «уснуть и не проснуться». Обычно люди, имеющие подобные мысли, как правило, одино-

ки. Не исключено, что эти два параметра могут быть взаимосвязаны. Для проверки данной гипотезы было проведено исследование с помощью метода «замораживания» [2]. Суть метода заключается в варьировании значения одного параметра и фиксировании значений всех других параметров. Для этой цели был выбран пример из тестируемого множества, где параметр душевной боли был равен 2 (т.е. в примере отмечается частое ее проявление).



**Рисунок 1. Доказательство адекватности прогнозирования суицидальной предрасположенности человека**

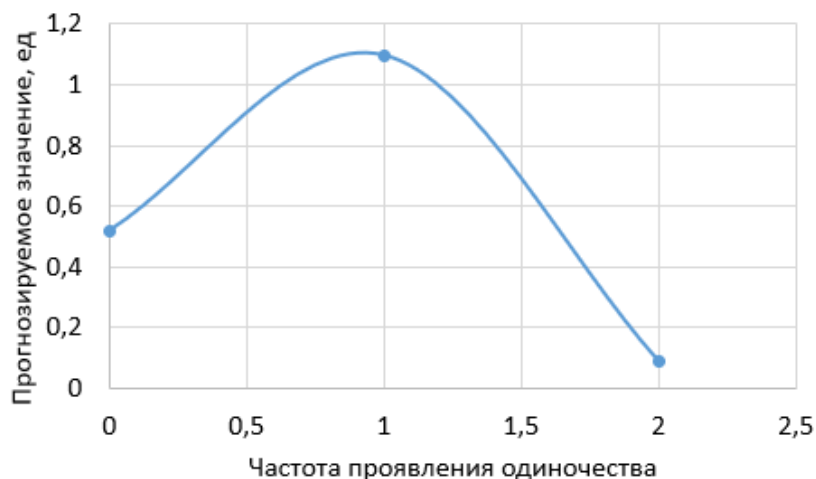


**Рисунок 2. Значимость входных параметров нейросетевой модели**

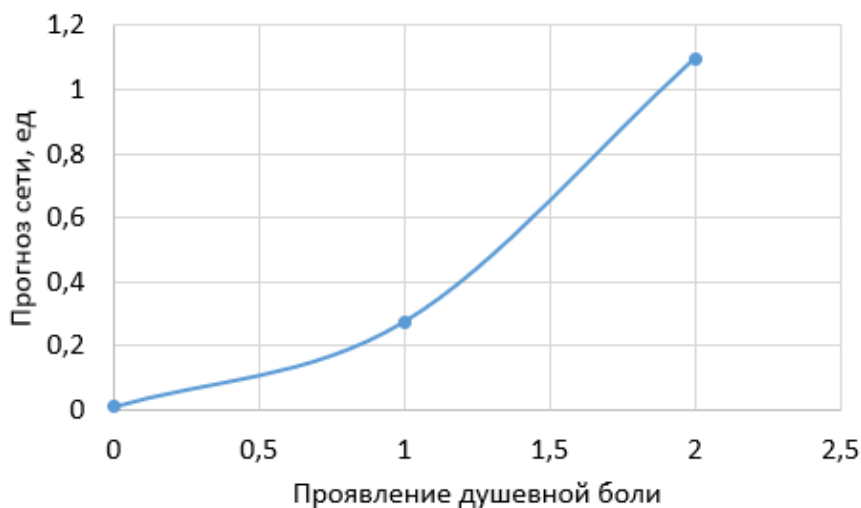
На рисунке 3 показан график зависимости изменения параметра одиночества от прогноза сети.

Как можно заметить, в случае если человек испытывает душевную боль и при этом у него не было мыслей «уснуть и не проснуться», а чувство оди-

ночества отсутствует, то у такого человека наблюдается средняя предрасположенность к суициду. Стоит подчеркнуть, что если человек иногда испытывает одиночество, то он, судя по прогнозу сети, попадает в группу риска в совершении самоубийства. Обратный эффект происходит, если человек, наоборот, очень часто испытывает одиночество, то вероятность самоубийства ниже 0,2 единиц. Попробуем провести такой же опыт, только уже с параметром проявления душевной боли, при этом значение проявления одиночества будет равен 1 (значение, равное ответу «иногда»). На рисунке 4 результат зависимости прогноза сети от проявления душевной боли.



**Рисунок 3. Зависимость прогнозируемого результата от изменения параметра одиночества**

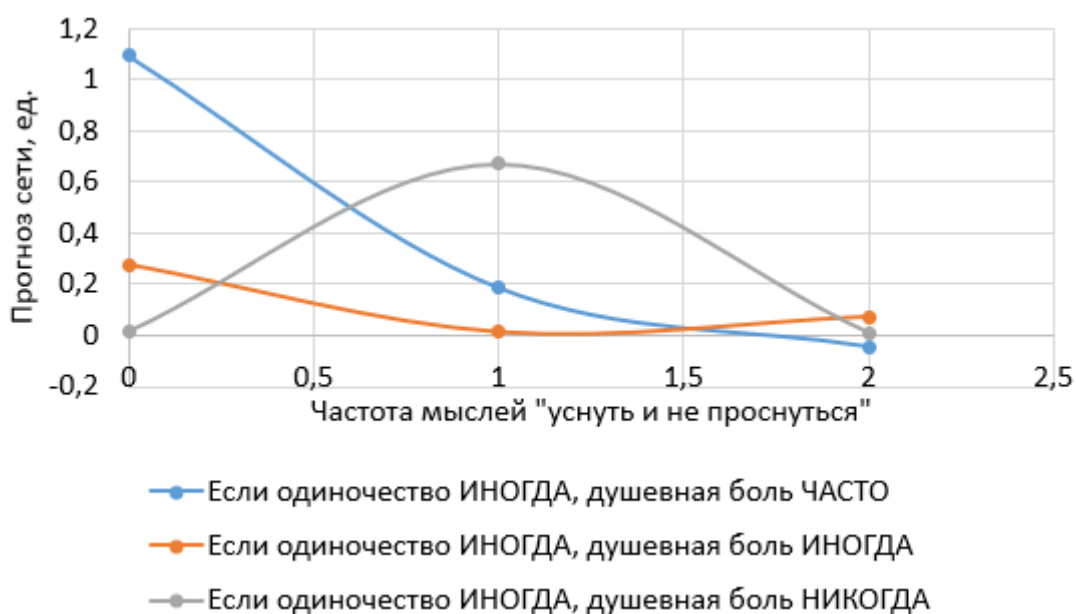


**Рисунок 4. Зависимость прогнозируемого результата от изменения параметра душевной боли**

Как можно заметить, чем чаще человек испытывает душевную боль и при этом иногда возникает ощущение одиночества, тем сильнее возрастает его склонность совершить самоубийство.

Как было сказано ранее, еще одним значимым параметром для прогноза сети является возникновение мыслей «уснуть и не проснуться». Проверим

зависимость данного показателя с учетом исследований предыдущих двух – одиночества и душевной боли. Поскольку было выявлено, что если у человека иногда возникает чувство одиночества, то его шанс совершить самоубийство намного больше, то в данном примере параметр одиночества будет равен 1 (т.е. значение «иногда»). Увеличение частоты ощущения душевной боли также влияет на прогнозируемый результат, поэтому для проверки зависимости параметра мыслей «уснуть и не проснуться» будет использовано значение душевной боли от частого проявления до его отсутствия. Таким образом на рисунке 5 получился график с результатами, в которых отражена зависимость параметра «уснуть и не проснуться» при определенных условиях предыдущих значимых параметров – одиночества и душевной боли.



**Рисунок 5. Зависимость прогнозируемого результата от частоты возникновения мыслей «уснуть и не проснуться» с учетом значимых параметров**

Из результатов видно, что в случае, если человек иногда испытывает одиночество и часто присутствует душевная боль, при этом у него никогда не было мыслей «уснуть и не проснуться», то сеть прогнозирует 100% гарантию доведения человека до самоубийства. Однако если наоборот при тех же значениях возникает мысль «уснуть и не проснуться» (иногда или часто), то шанс дойти до суицида в разы меньше. В случае, если все перечисленные параметры имеют значение «иногда», то прогнозы сети минимальны. Но если человек иногда одинок и никогда не испытывал душевной боли, то от периодических мыслей «уснуть и не проснуться» растет склонность человека к совершению самоубийства.

Полученные результаты исследований не противоречат реальности, что подтверждает, что спроектированную нейронную сеть можно считать пригодной для использования в работе психологов и врачей.

**Заключение.** Построена нейросетевая система прогнозирования склонности к суициду. Спроектированная нейросетевая модель учитывает 15

параметров, включающие в себя личные убеждения человека о ценности жизни, его недомоганиях и внутренних состояниях. Методом сценарного прогнозирования построены графики зависимостей значимых параметров. Применение такого набора параметров в модели позволяет предсказывать склонность человека к совершению самоубийства.

### Библиографический список

1. Сайт Testometrika. Опросник суицидального риска. [Электронный ресурс] // URL: <https://testometrika.com/personality-and-temper/the-questionnaire-of-suicidal-risk-for-adolescents/>
2. Ясницкий Л.Н. Введение в искусственный интеллект. М.: Издательский центр «Академия», 2005. – 176 с.
3. Сайт RB.ru. Израильские ученые создали алгоритм, который определяет склонность к суициду на ранних стадиях. [Электронный ресурс] // URL: <https://rb.ru/story/suicide-risk/>
4. Сайт Demoscop. Самоубийства и их профилактика в Российской Федерации, 2021 год: основные факты. [Электронный ресурс] // URL: <http://www.demoscope.ru/weekly/2021/0911/suicide.php>
5. Сайт Федерального государственного бюджетного учреждения «Центр защиты прав и интересов детей». Причины, мотивы, симптомы, виды суицидального поведения. [Электронный ресурс] // URL: <https://fcprc.ru/value-of-life/prichiny-motivy-simptomyy-vidy-suicidalnogo-povedeniya/>

## NEURAL NETWORK SYSTEM FOR PREDICTING SUICIDAL TENDENCIES

*Trefilov Dmitry A.*

National Research University “Higher School of Economics”  
Str. Gagarin Boulevard, 37a, Perm, Russia, 614060, [trefilowdem@mail.ru](mailto:trefilowdem@mail.ru)

The article describes the development of a neural network system for predicting a person's propensity to commit suicide. This system makes it possible, based on complaints about the state of health and beliefs about the value of life, to determine whether a person has a predisposition to suicide or not. With the help of the developed intellectual system, a study of the subject area was carried out, regular parameters of practical importance were identified. The developed model can be a useful tool for psychologists and doctors.

**Keywords:** artificial intelligence, neural network technologies, psychological health, identification of predisposition to suicide.



## НЕЙРОСЕТЕВОЕ ПРОГНОЗИРОВАНИЕ НАЛИЧИЯ НЕРВНО-ПСИХИЧЕСКОЙ НЕУСТОЙЧИВОСТИ

*Зейтунян Аким Арсенович*

Пермский государственный национальный исследовательский университет,  
ПМИ. 614990, Россия, г. Пермь, ул. Букирева, 15, akim.zeytunyan@mail.ru

Рассмотрены способы разработки математической нейросетевой модели, предназначенной для определения предрасположенности человека к нервно-психической неустойчивости. Для проектирования модифицированных моделей и их исследования написана программа на языке Python с использованием библиотеки Keras, предназначенной для работы с искусственными нейронными сетями. Также разработана аналогичная модель средствами нейросимулятора [2]. Рассмотрены различные технологии и алгоритмы проектирования моделей, проведено сравнение полученных результатов.

**Ключевые слова:** искусственный интеллект, нейронная сеть, закономерности, прогнозирование, нервно-психическая неустойчивость, Python, Keras.

**Введение.** Нервно-психическая неустойчивость, характеризующаяся эмоциональной изменчивостью, тревогой, низким самоуважением, вегетативными расстройствами [3, с 253], несомненно, является серьезной проблемой современного общества. Низкий уровень психологического воспитания, образования окружающих людей вкуче с высокой чувствительностью нервной системы отдельного индивида создает благоприятные условия для развития у него данного расстройства. Так, при проведении автором опроса в рамках работы [1], 62% респондентов отметили наличие у себя симптомов НПН. Лечение таких людей не в нашей компетенции, однако в наших силах попытаться решить задачу, которая в свою очередь позволит своевременно проводить соответствующие мероприятия по подготовке и решению данной проблемы – задачу прогнозирования НПН, что и является целью данной работы.

Нейросетевая модель для прогнозирования НПН в предыдущей работе автора [1] была спроектирована, оптимизирована, обучена и исследована с помощью нейросимулятора [2], а ее среднеквадратичная процентная ошибка тестирования (обобщения) составила 15%. Данными для обучения и тестирования выступали результаты опроса (с содержимым можно ознакомиться ниже), общее число респондентов составило 150 человек. Методом сценарного прогнозирования [5, с. 89] были построены графики зависимости вероятности НПН от некоторых входных параметров. В данной статье мы попытаемся улучшить имеющуюся модель, а также провести более качественное исследование предметной области.

**Нейронные сети в психодиагностике.** В настоящее время нейронные сети успешно используются в научных исследованиях. С момента их появле-

ния они доказали свою эффективность в решении широкого спектра задач, связанных с распознаванием образов, классификацией, принятием решений, адаптивным управлением, кластеризацией, прогнозированием, аппроксимацией, сжатием данных и ассоциативной памятью [6].

Благодаря внедрению информационных технологий в психологическую науку появилась возможность решения задач практической психологии с помощью нейронных сетей. Важное место в современной психологии отводится психодиагностике – отрасли психологии, целью которой стоит измерение индивидуально-психологических особенностей личности, их оценка. В современной науке нейросети применяются во многих психодиагностических исследованиях. Так, например, Пермским отделением Научного Совета РАН по методологии искусственного интеллекта получены положительные результаты в создании интеллектуальных систем, предназначенных для диагностики сердечно-сосудистых заболеваний [7], выявления склонности к насилию [8], выявления у человека скрытых талантов, способностей [9, 10].

**Методика.** В качестве входных параметров модели в работе [1] была выбрана общая информация о человеке (пол, возраст, образование) и предполагаемые 11 причин НПП, а для построения датасета проводился соответствующий опрос, в котором также предлагалось отметить наличие у себя симптомов НПП, что определяло выходной параметр модели. Общий список входов и выходов был следующий:

$x_1$  – Пол:

*0 – мужской; 1 – женский.*

$x_2$  – Возраст:

$x_3$  – Образование:

*0 – неполное среднее; 1 – среднее/среднее специальное; 2 – неполное высшее; 3 – высшее.*

$x_4$  – Тип темперамента:

*0 – сангвиник; 1 – флегматик; 2 – холерик; 3 – меланхолик.*

$x_5$  – Тип мышления:

*0 – технический; 1 – гуманитарный.*

$x_6$  – Подверженность в детстве «гиперопеке» со стороны родителей/родственников:

*0 – был подвержен в большом количестве; 1 – был подвержен в умеренном количестве; 2 – не был подвержен.*

$x_7$  – Наличие психических отклонений, таких как частые депрессивные состояния, неврозы, нервные тики, маниакальные расстройства:

*0 – да; 1 – нет.*

$x_8$  – Рос в полной семье:

*0 – да; 1 – нет, рос без отца; 2 – нет, рос без матери; 3 – нет, рос без родителей.*

$x_9$  – Наличие хронических заболеваний:

*0 – да; 1 – нет.*

$x_{10}$  – Правильное питание:

*0 – да; 1 – пытаюсь правильно питаться, но у меня не всегда это выходит; 2 – редко правильно питаюсь; 3 – нет.*

$x_{11}$  – Подвергался психологическим травмам, в т. ч. насильственного физического и/или психологического характера:

*0 – да; 1 – нет.*

$x_{12}$  – Наличие проблем, которые не с кем обсудить:

*0 – да; 1 – нет.*

$x_{13}$  – Трудно знакомиться с новыми людьми:

*0 – да; 1 – нет.*

$x_{14}$  – Знакомых в интернете больше, чем в реальной жизни:

*0 – да; 1 – нет.*

$u_1$  – Наличие симптомов НПН:

*0 – да; 1 – нет.*

Так как большинство параметров проектируемой модели являются категориальными, то есть имеют нечисловые значения, в прошлой работе мы сопоставляли этим значениям некоторые числа. Однако, серьезным минусом данного способа кодирования является то, что алгоритмы машинного обучения, как правило, считают большее число на входе более важным, чем меньшее. В нашем же случае все входные данные не имеют ранжирования и являются независимыми друг от друга, поэтому более или менее важных категорий нет. В связи с этим в данной работе мы применим способ кодирования данных, называемый в литературе «One Hot Encoding». Согласно нему, мы конвертируем каждое категориальное значение в новый категориальный столбец и присваиваем этим столбцам двоичное значение 1 или 0. Так, например, параметр « $x_1$  – Пол» в рассматриваемой модели будет конвертирован в двоичный вектор длины 2, значение «мужской» будет представлено вектором [1, 0], значение «женский» – вектором [0, 1]. Закодировав таким способом все категориальные параметры, мы получим 37 входных значений модели вместо 14. На выходном слое остается, как и прежде, 1 нейрон, отвечающий за спрогнозированную моделью степень уверенности в наличии у пациента НПН, где 0 – полная уверенность в отсутствии НПН, 1 – полная уверенность в наличии. Получаемое на выходе значение также можно интерпретировать, как вероятность у пациента данного расстройства.

Множество полученных результатов опроса (150 ответов) было разделено на два подмножества: обучающее (90%) и тестирующее (10%), а в процессе обучения на валидирующее множество выделялось 20% от обучающего.

Найденная методом проб и ошибок модель в нейросимуляторе [2] содержит 37 входных нейронов, один скрытый слой с 12-ю нейронами и один выходной нейрон. Функцией активации для всех слоев является гиперболический тангенс. В качестве алгоритма обучения был выбран алгоритм упругого распространения.

При использовании Keras структура найденной «оптимальной» нейросети содержит 37 входных нейронов, один скрытый слой с 14-ю нейронами с активационными функциями *relu* и один выходной нейрон с сигмоидной функцией активации. Для обучения модели был выбран модифициро-

ванный метод градиентного спуска Adam. Идея данного метода состоит в изменении величины скорости обучения learning rate в методе градиентного спуска в зависимости от значения производной функции ошибки модели, чтобы не допускать слишком «больших скачков» по этой функции [4]. Для оценки качества модели использовалась среднеквадратичная ошибка

$$RMSE = \frac{1}{N} \sqrt{\sum_{i=1}^N (y_i - \hat{y}_i)^2},$$

а ее перевод в процентную величину проводился по формуле

$$MSPE = \frac{RMSE}{|\max(y_i) - \min(y_i)|} * 100$$

Также при обучении модели была применена так называемая дропаут-регуляризация, согласно которой на каждом новом примере обучения каждый нейрон «выбрасывается» из сети с определенной вероятностью  $p$ . После завершения обучения выходной сигнал каждого нейрона умножается на вероятность  $1 - p$ , с которой он оставался в сети. В нашем случае величина  $p = 0.2$ .

**Исследование результатов.** При использовании Keras путем множества попыток обучения получилось прийти к случаю, в котором ошибка  $MSPE$  модели на тестирующем множестве составила 18.6%. В нейросимуляторе аналогичная ошибка составила 19.2%. То есть качество полученных моделей довольно близко. Однако в работе [1] та же ошибка составила 15%, что предположительно может означать, что использование кодирования входных категориальных переменных в выбранной предметной области отрицательно сказывается на качество модели.

На рис. 1, 2 представлены сравнения реальных и спрогнозированных значений на тестирующих множествах.

Как мы видим, хоть модель и запомнила основные закономерности предметной области, ошибка прогноза довольно велика и в отдельных случаях совсем не соответствует действительности. Можно сделать вывод, что данной модели все еще необходима доработка, а одной из основных причин невысокой точности прогноза, по мнению автора, является небольшой размер датасета.

На рис. 3 можно видеть оценки значимостей входных параметров, найденные средствами нейросимулятора. Для получения оценки значимости конкретного параметра находилось среднее значение элементов соответствующего ему вектора, полученного при использовании «One Hot Encoding».

Заметим, что при использовании кодирования категориальных переменных самым значимым показателем является «Наличие проблем, которые не с кем обсудить», тогда как без кодирования таковым являлся показатель

«Большее количество знакомых в интернете, чем в реальной жизни». Однако, общая картина остается прежней – в первые шесть наиболее значимых параметров в обоих случаях попали одни и те же показатели, а наименее значимым оказался пол. Наиболее значимые параметры определенно дают понять, что нервно-психическая неустойчивость возникает в основном у одиноких, не уверенных в себе людей.

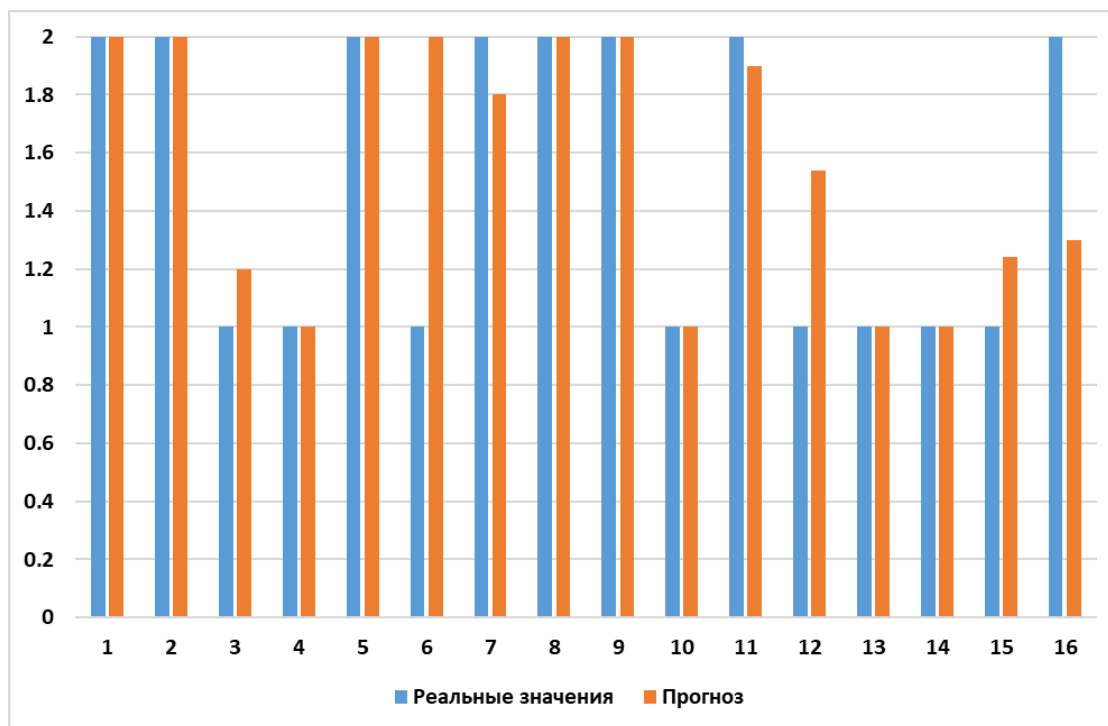


Рисунок 1. Keras. Сравнение реальных и прогнозных значений

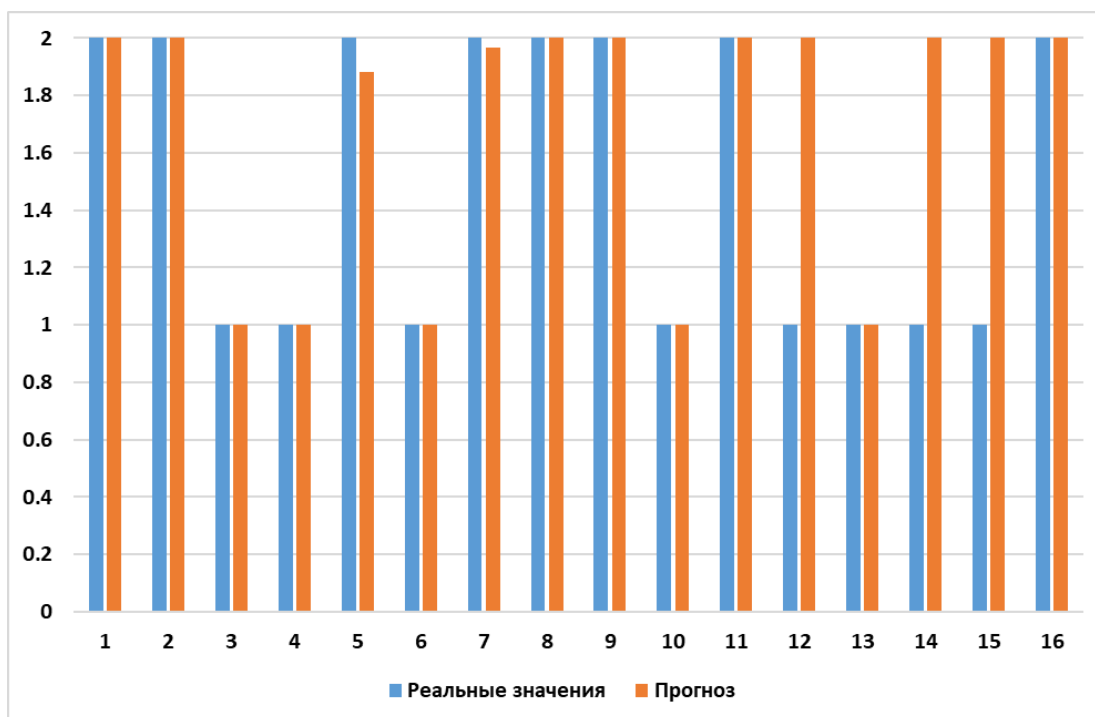
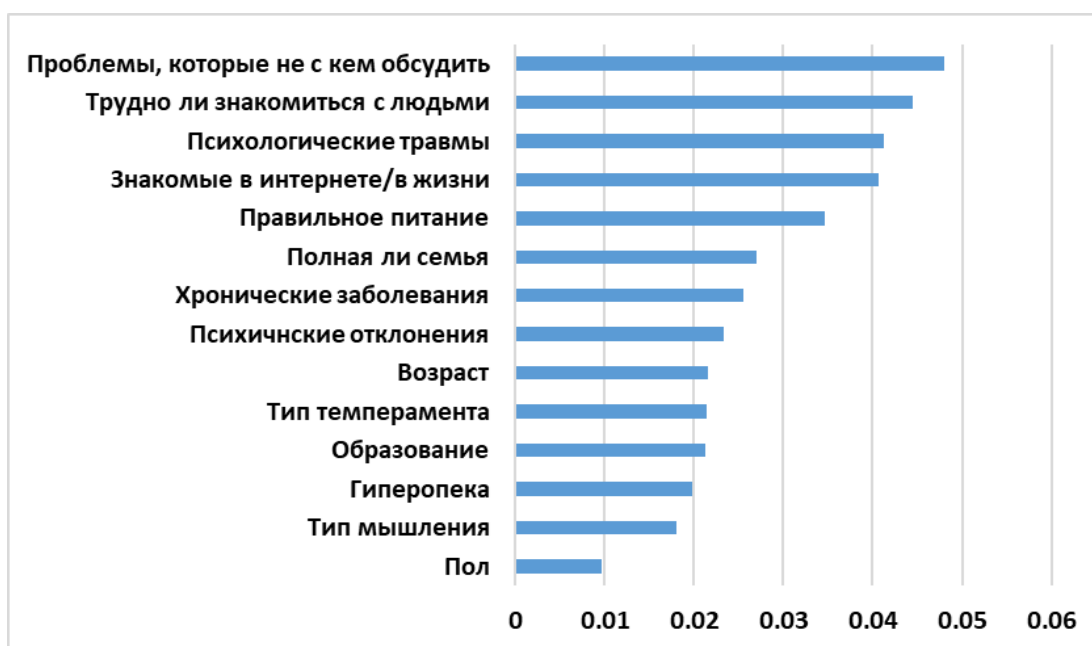


Рисунок 2. Нейросимулятор. Сравнение реальных и прогнозных значений

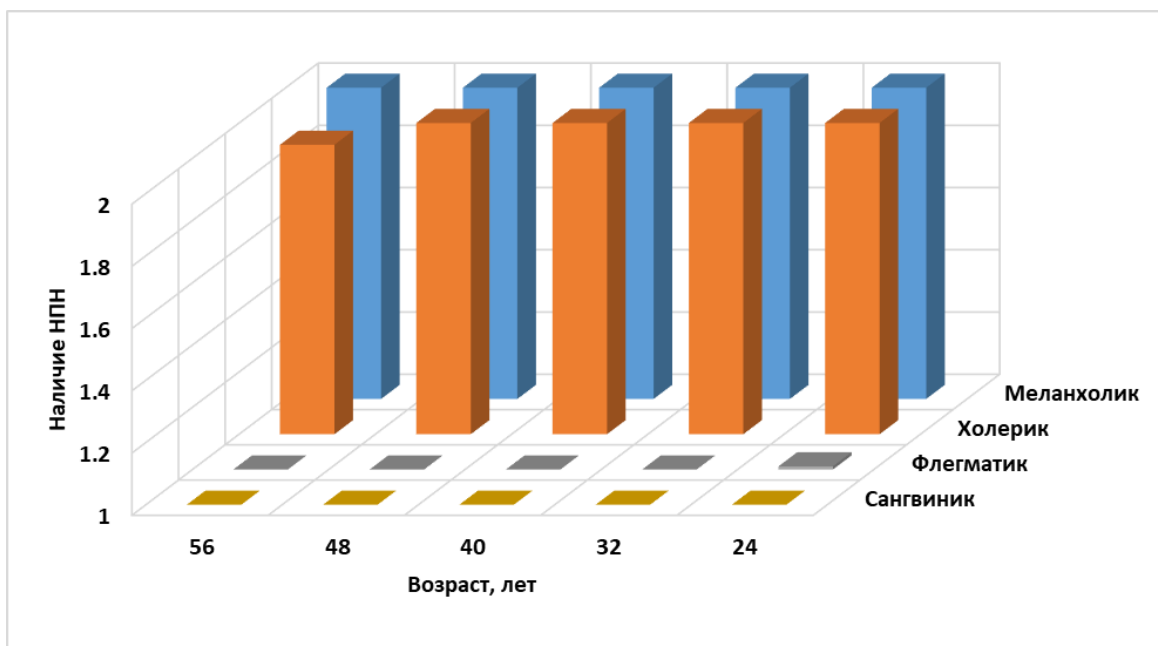


**Рисунок 3. Нейросимулятор. Сравнение реальных и прогнозных значений**

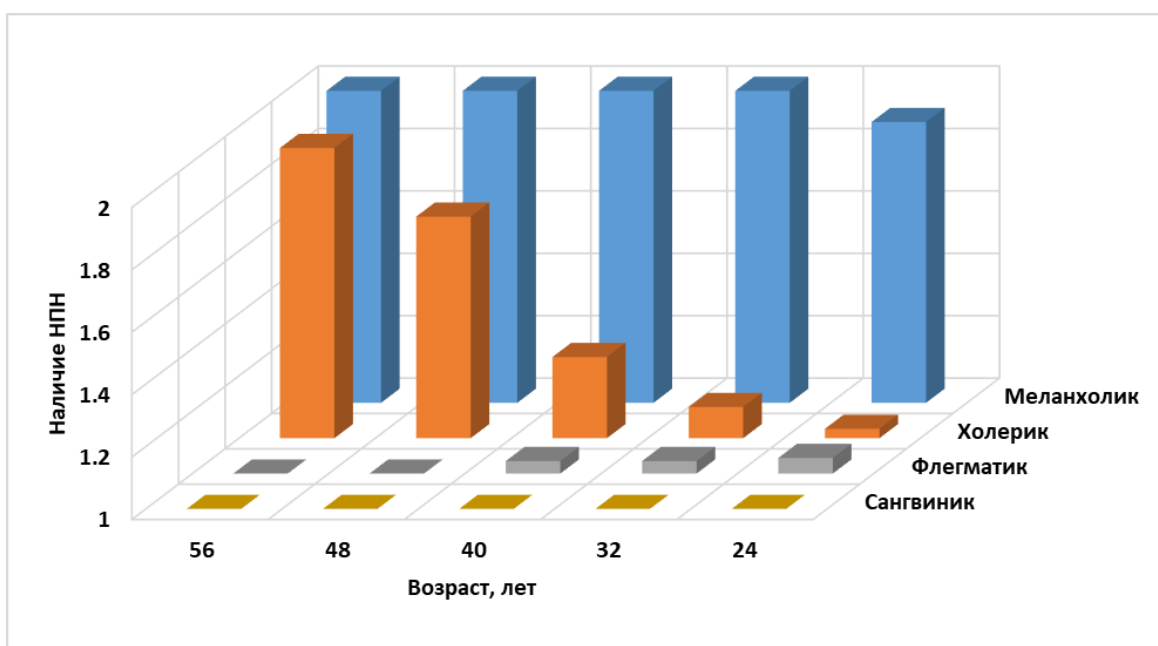
Далее, с помощью изменения какого-либо показателя и «заморозки» всех остальных, проведем исследование полученных зависимостей между входными параметрами и результатом. Такое исследование уже проводилось в работе [1], однако в нем выбор значений входных параметров проводился случайно, что, вообще говоря, с точки зрения рекомендаций по избавлению от НПН, не оптимально, так эти рекомендации будут иметь смысл только для людей с аналогичными параметрами. Сейчас же, чтобы выявить зависимости наиболее полезные для большинства, выберем в качестве входных параметров наиболее популярные ответы в опроснике. Получим следующие значения:

- Образование: высшее;
- Тип темперамента: холерик;
- Тип мышления: гуманитарный;
- Подверженность гиперопеке: в умеренных количествах;
- Наличие психических отклонений: отсутствуют;
- Рос в полной семье: да;
- Наличие хронических заболеваний: отсутствуют;
- Питание: умеренно правильное;
- Психологические травмы: отсутствуют;
- Проблемы, которые не с кем обсудить: присутствуют;
- Трудно знакомиться с людьми: нет;
- Больше количество знакомых в интернете, чем в жизни: да;

Первым делом рассмотрим зависимость наличия НПН от возраста и типа темперамента. Полученные результаты можно видеть на рис. 4, 5. Единица по оси «Наличие НПН» означает абсолютную уверенность нейросетевой модели в отсутствии у объекта нервно-психической неустойчивости, а двойка, соответственно, абсолютную уверенность в её присутствии.



**Рисунок 4. Зависимость наличия НПН у мужчины от возраста и типа темперамента**

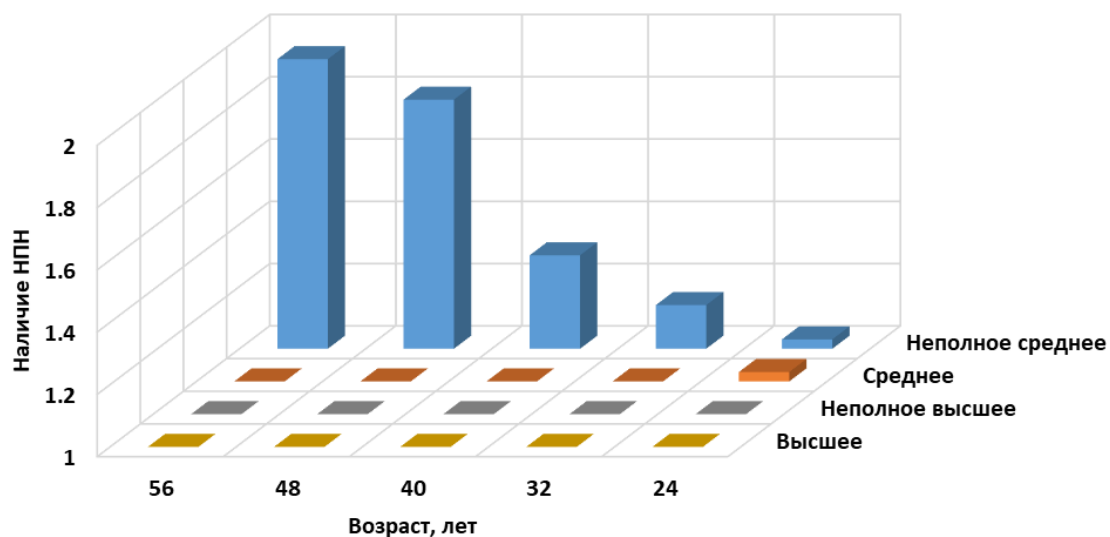


**Рисунок 5. Зависимость наличия НПН у женщины от возраста и типа темперамента**

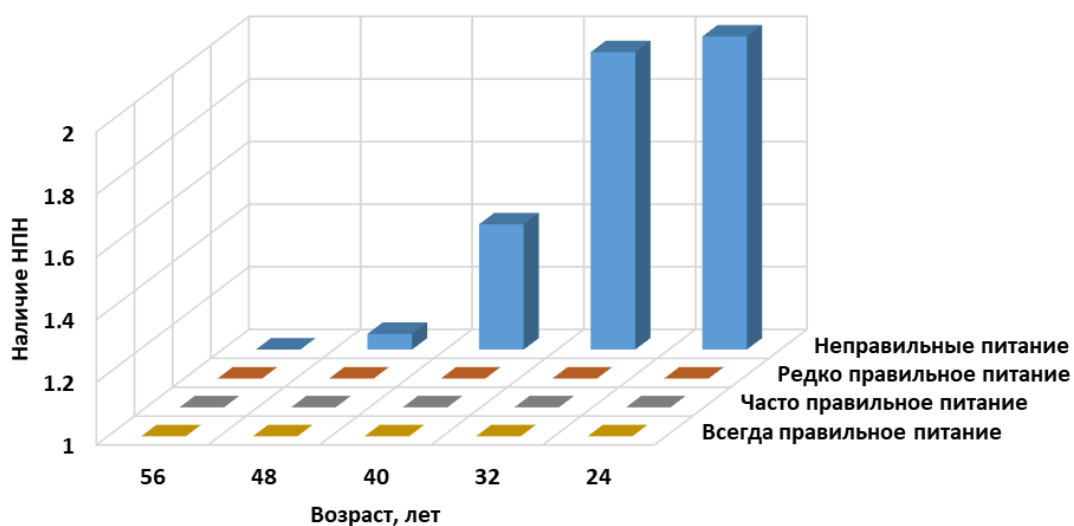
Из приведенных гистограмм можно сделать вывод, что среди людей с параметрами среднестатистического респондента проведенного опроса, наиболее подверженными НПН будут меланхолики и холерики, при чем зависимость от возраста просматривается только у холериков женского пола.

На рис. 6, 7 представлены результаты аналогичного исследования с показателями образования и питания.

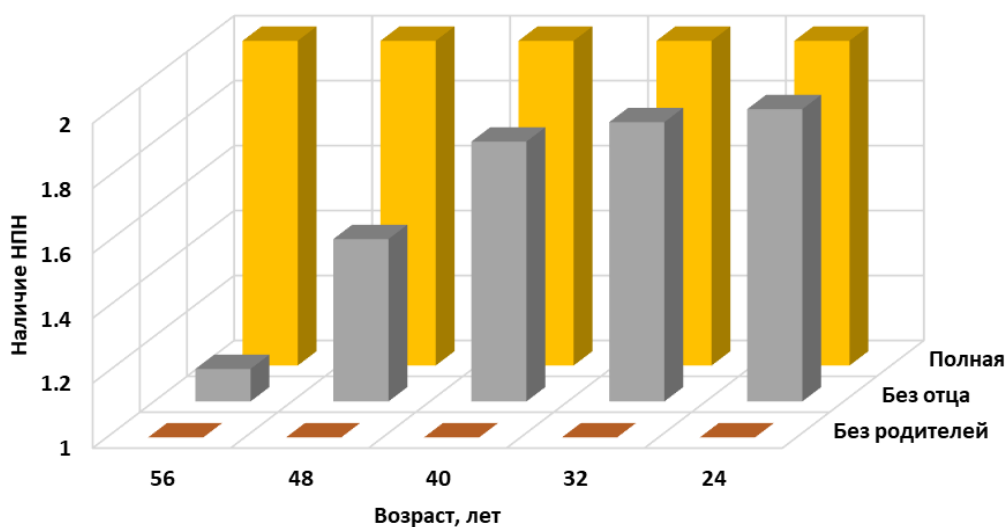
Видно, что модель прогнозирует НПН только у среднестатистической женщины без среднего образования, при чем по мере увеличения возраста вероятность НПН возрастает.



**Рисунок 6. Зависимость наличия НПН у женщины от возраста и образования**



**Рисунок 7. Зависимость наличия НПН у женщины от возраста и питания**



**Рисунок 8. Зависимость наличия НПН у мужчины от состава семьи, в которой он вырос**



С точки зрения питания вероятность НПН существенна лишь при полностью неправильном питании, и чем меньше возраст, тем эта вероятность выше. При построении таких же гистограмм для лиц мужского пола зависимость не видна.

Рассмотрим также зависимость НПН у мужчины от состава семьи, в которой он вырос (рис. 8).

Как ни странно, модель дает прогноз, что лицо мужского пола, выросшее в более полной семье, больше подвержено НПН, а мужчина, росший без родителей, НПН не подвержен совсем.

**Заключение.** Рассмотрены способы разработки нейросетевой математической модели, предназначенной для определения у человека наличия нервно-психической неустойчивости с помощью нейросимулятора [2], а также инструментами библиотеки Keras ([https://github.com/kknstnt/МП-research/blob/master/МИИ\\_исследование/МИИ\\_исследование.py](https://github.com/kknstnt/МП-research/blob/master/МИИ_исследование/МИИ_исследование.py)). Проведены сравнения полученных моделей, анализ значимости параметров. Методом сценарного прогнозирования [5, с. 89] путем изменения одного из входов нейросети и сохранения неизменными всех остальных построены графики зависимости наличия НПН от некоторых входных параметров.

### Библиографический список

1. Зейтунян, А. А. Нейросетевое прогнозирование наличия нервно-психической неустойчивости / А. А. Зейтунян // Интеллектуальные системы в науке и технике. Искусственный интеллект в решении актуальных социальных и экономических проблем XXI века: Сборник статей по материалам Международной конференции и Шестой всероссийской научно-практической конференции, Пермь, 12–18 октября 2020 года / Под редакцией Л.Н. Ясницкого. – Пермь: Пермский государственный национальный исследовательский университет, 2020. – С. 409-419.
2. Черепанов Ф.М., Ясницкий Л.Н. Нейросимулятор 4.0. Свидетельство о государственной регистрации программы для ЭВМ № 2014618208. Заявка Роспатент № 2014614649. Зарегистрировано в Реестре программ для ЭВМ 12 августа 2014г.
3. Александровский Ю.М. Пограничные психические расстройства. – СПб: Питер, 2006.
4. Аббакумов В.Л. (2018, сентябрь 24). Лекция 2. Нейронные сети. Теория и первый пример (Анализ данных на Python в примерах и задачах. Ч2) [Видео файл]. Взято из <https://www.youtube.com/watch?v=RKKhzFBmEBg&list=PLlb7e2G7aSpT1ntsozWmWJ4kGUsUs141Y&index=2>
5. Ясницкий Л.Н. Интеллектуальные системы / Лаборатория знаний, 2016. —221 с.
6. Хайкин С. Нейронные сети: Полный курс. – Изд. 2-е. – М.: Вильямс, 2006 – 1101 с.
7. Ясницкий Л.Н., Думлер А.А., Полещук А.Н., Богданов К.В., Черепанов Ф.М. Нейросетевая система экспресс-диагностики сердечно-сосудистых заболеваний // Пермский медицинский журнал. – 2011. – №4. – С. 77-86.

8. Ясницкий Л.Н. Использование методов искусственного интеллекта в изучении личности серийных убийц / Л.Н. Ясницкий, С.В. Ваулева, Д.Н. Сафонова, Ф.М. Черепанов // Криминологический журнал Байкальского государственного университета экономики и права. — 2015. — Т. 9, № 3. — С. 423–430. — DOI: 10.17150/1996-7756.2015.9(3).423-430.

9. Ясницкий Л.Н., Михалева Ю.А., Черепанов Ф.М. Возможности методов искусственного интеллекта для выявления и использования новых знаний на примере задачи управления персоналом // International Journal of Unconventional Science. Журнал Формирующихся Направлений Науки. 2014. Вып. 6; URL: <http://www.unconv-science.org/n6/yasnitsky/>

10. Байдин Д.Ю. Определение способности к научной деятельности с помощью нейронных сетей // Искусственный интеллект: философия, методология, инновации. Сборник трудов VII Всероссийской конференции студентов, аспирантов и молодых ученых. Часть 2. г. Москва, МГТУ МИРЭА, 13-15 ноября 2013 г. – М.: «Радио и Связь», 2013. – С. 10-15.

## NEURAL NETWORK FORECASTING THE PRESENCE OF NEUROPSYCHIATRIC INSTABILITY

*Zeytunyan Akim A.*

Perm State University

614990, Russia, Perm, Bukireva st., 15, [akim.zeytunyan@mail.ru](mailto:akim.zeytunyan@mail.ru)

This article discusses ways to improve a mathematical neural network model designed to determine a person's predisposition to neuropsychiatric instability, developed in the article [1].

**Keywords:** artificial intelligence, neural network, regularities, forecasting, neuro-mental instability, Python, Keras.

УДК 004.032.26

## НЕЙРОСЕТЕВАЯ СИСТЕМА ОЦЕНКИ ЗАВИСИМОСТИ ЦВЕТОВ ФОТОГРАФИИ ПРОФИЛЯ В СОЦИАЛЬНОЙ СЕТИ И КОЛИЧЕСТВА ОТМЕТОК «НРАВИТСЯ»

*Шарипова Алиса Олеговна*

Национальный исследовательский университет «Высшая школа экономики»

614070, Россия, г. Пермь, ул. Студенческая, 38, [inforperm@hse.ru](mailto:inforperm@hse.ru)

В статье рассмотрена степень текущей разработанности темы продвижения профилей в социальной сети с помощью алгоритмов машинного обучения. Также описано получение и подготовка данных для обучения нейронной сети, проведено её обучение, проверено качество обучения сети.

После этого выявляется закономерность, имеющее практическое значение, а также представлен прогноз нейронной сети количества отметок «нравится» для нескольких новых фотографий.

**Ключевые слова:** искусственный интеллект, нейронная сеть, социальные сети, продвижение, анализ фотографий.

**Введение.** В современном мире трудно представить существование малого бизнеса без присутствия в сети Интернет, в социальных сетях. В частности, как показало исследование Крыловой Д. С., малый и средний бизнес всё чаще выбирает в качестве канала коммуникации с целевой аудиторией, а также в качестве способа продвижения социальные сети. И в том числе этим обуславливается рост компаний [1].

Однако одного присутствия в социальных сетях недостаточно для роста бизнеса. Представители малого бизнеса, индивидуальные предприниматели вынуждены или нанимать специалиста по продвижению в социальных сетях, или изучать этот вопрос сами.

В последнем случае предпринимателям будет полезно иметь основу для своего продвижения. В частности, иметь результаты анализа собственного профиля для того, чтобы сделать выводы для последующих публикаций. Данное исследование сконцентрировано на анализе визуальной составляющей профилей социальной сети «ВКонтакте» – фотографий. В работе будет оценена зависимость между цветами фотографий и количеством у них отметок «нравится».

Предполагается обучение модели на наборе фотографий, выявление наиболее удачного для фотографий цвета, а также оценка количества отметок «нравится» у фотографий, не вошедших в обучающее и валидирующее множества.

**Подготовка набора данных.** Как было сказано выше, данное исследование заключаться в анализе фотографий профилей представителей малого бизнеса ВКонтакте. Для получения этой информации могут быть использованы разные методы: использование сторонних сервисов, скачивание вручную, использование технологий веб-скрапинга и подключение к социальной сети с помощью API (application programming interface) «ВКонтакте». В данной работе был использован последний метод, так как он легальный и бесплатный.

Скачивание фотографий выполнялось на языке python в среде «Jupyter Notebook». Для подключения к API использовалась библиотека языка python – vk. На рисунке 1 приведена функция для получения сведений о фотографиях по id пользователя.

```
def GetAllPhotos (id):
    collection = []
    length = vkapi.photos.getAll(owner_id=id,count=1)['count']
    for part in range(0, math.ceil(length/200)):
        response = vkapi.photos.getAll(owner_id=id,extended=1,count=200,offset=part*200)
        temp = response['items']
        collection.extend(temp)
        time.sleep(0.5)
    return collection
```

**Рисунок 1. Функция для получения фотографий**

На рисунке 2 приведена функция для получения ссылок на фотографии.

```
def GetLinks (collection):
    urls = []
    for photo in collection:
        url = photo["sizes"][len(photo["sizes"])-1]["url"]
        urls.append(url)
    return urls
```

**Рисунок 2. Функция для получения ссылок на фотографии**

Далее были использованы возможности Azure Cognitive Services Компьютерное зрение. Для этого также было использовано подключение к API сервиса с помощью индивидуального ключа. Сервис позволяет анализировать фотографии с точки зрения того, что на них изображено, основных цветов, категорий и наличия запрещённого контента. В данном исследовании в первую очередь были использованы возможности определения цветов на фотографиях. Это было сделано с помощью следующей функции, которая представлена на рисунке 3.

```
def GetCategoriesAndColors (urls):
    results = []
    features = ["categories", 'color']
    for url in urls:
        result = computervision_client.analyze_image(url , features)
        results.append(result)
    return results
```

**Рисунок 3. Функция для анализа категорий и цветов фотографий**

После этого необходимо было сформировать таблицу из полученных результатов, для этого была использована библиотека pandas языка python и написана следующая функция, представленная на рисунке 4.

```
def CreateDataFrame (results, phot, size):
    df = pd.DataFrame(index=range(size), columns=range(6))
    for i in range(0, size):
        df.iloc[i,0] = results[i].color.accent_color
        df.iloc[i,1] = results[i].color.dominant_color_background
        df.iloc[i,2] = results[i].color.dominant_color_foreground
        if results[i].color.dominant_colors!=[]:
            df.iloc[i,3] = results[i].color.dominant_colors
            df.iloc[i,4] = results[i].color.dominant_colors[0]
        df.iloc[i,5] = phot[i]['likes']['count']
    df = df.rename(columns={0: "accent_color", 1: "dominant_color_background", 2: "dominant_color_foreground", 3: "dominant_colors", 4: "dominant_color", 5: "likes"})
    df = df.dropna()
    return df
```

**Рисунок 4. Функция для формирования таблицы**

Далее каждый цвет фотографии был вынесен в отдельный столбец и закодирован методом One-Hot Encoding. Полученная таблица была выгружена в csv файл. Ниже приведён фрагмент такой таблицы, который представлен на рисунке 5.

Black	Blue	Brown	Green	Grey	Orange	Pink	Purple	Red	Teal	White	Yellow	likes
X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12	D
0	0	0	1	0	0	0	0	0	0	0	0	69
1	0	0	0	0	1	0	0	0	0	0	1	3
0	0	0	0	0	0	0	0	0	0	0	1	2
0	0	1	0	0	0	0	0	0	0	0	1	2
1	0	0	1	0	0	0	0	0	0	0	1	4
0	0	0	0	1	0	0	0	0	0	0	1	7
0	0	1	0	1	0	0	0	0	0	0	0	3
1	0	0	0	0	0	0	0	0	0	0	0	2
1	0	0	0	0	1	0	0	0	0	0	1	45

**Рисунок 5. Фрагмент набора данных**

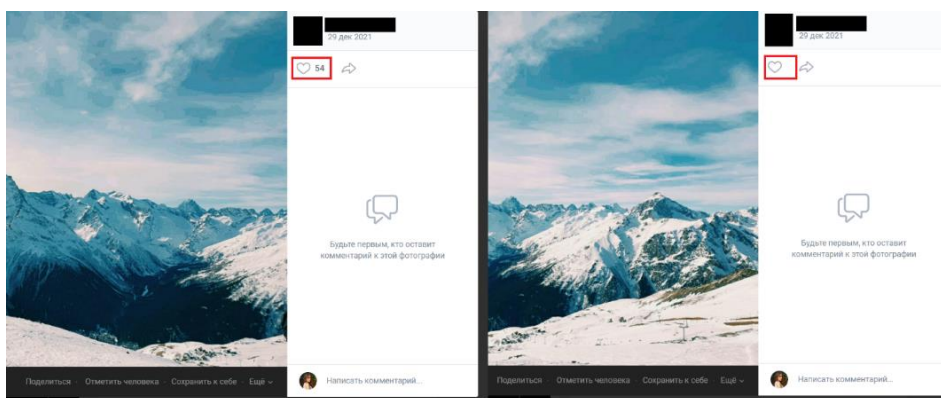
Полученный набор данных содержал около 1100 строк и был разделён на обучающую, валидационную и тестовую выборки.

**Построение нейронной сети.** Для построения модели были получены следующие параметры: X1 – чёрный цвет; X2 – синий цвет; X3 – коричневый цвет; X4 – зелёный цвет; X5 – серый цвет; X6 – оранжевый цвет; X7 – розовый цвет; X8 – фиолетовый цвет; X9 – красный цвет; X10 – сине-зелёный цвет; X11 – белый цвет; X12 – жёлтый цвет. Выходной параметр – количество отметок «нравится».

Проектирование, обучение, тестирование нейронной сети выполнялись с помощью программы «Нейросимулятор 5» [2] по методике [3]. В результате оптимизации спроектированная нейронная сеть представляла собой персептрон, который имеет двенадцать входных нейронов, один выходной и один скрытый слой с 38 нейронами на нём.

Для оценки корректной работы спроектированной нейронной сети использовалось тестирующее множество, состоящее из 58 примеров. Средняя относительная ошибка тестирования составила 35%, что не является приемлемым результатом даже для данных из социальной сети. Следовательно, требуется либо доработка модели, либо чистка исходного набора данных поиск выбросов в нём.

При детальном рассмотрении исходного набора данных оказалось, что в нём присутствуют «противоречащие» примеры, связанные с особенностью социальной сети. В частности, очень похожие фотографии из одной серии, имеющие одинаковые цвета отмечены разным количеством «лайков». Это можно увидеть на рисунке 6.



**Рисунок 6. Почти одинаковые фотографии с разным количеством «лайков»**

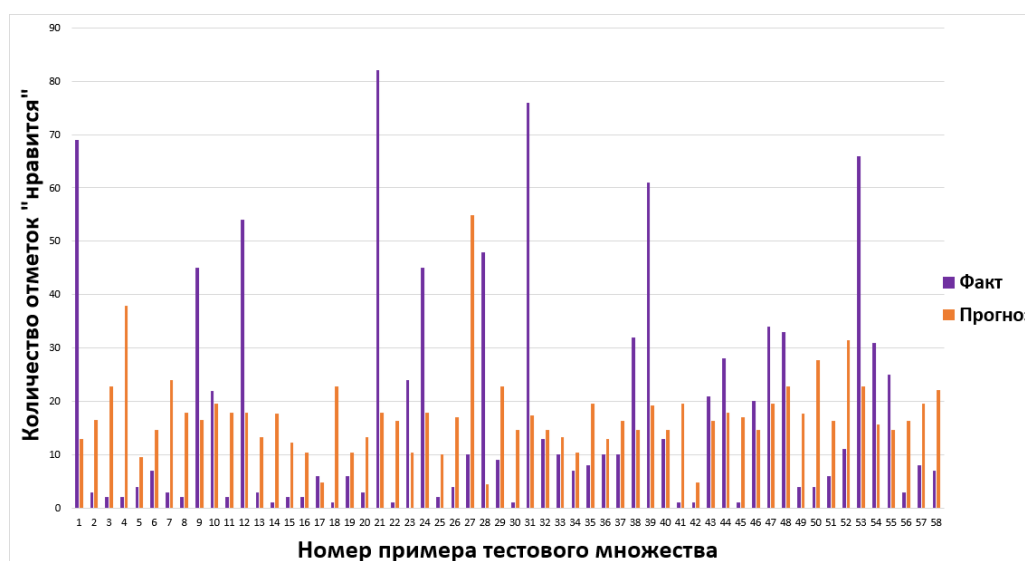
После выявления таких «дубликатов» из всех множеств были удалены примеры с нулевым количеством отметок «нравится». Также было решено немного упростить сеть уменьшив количество нейронов на скрытом слое до 13 и установить линейную активационную функцию выходного слоя.

На рисунке 6 представлена гистограмма, демонстрирующая разницу между фактическим и прогнозируемым нейросетью количеством отметок «нравится». Средняя относительная ошибка тестирования составила 27,8%. Расчёт среднеквадратичной ошибки в нейросимуляторе производится с помощью следующей формулы:

$$E = \frac{\sqrt{\frac{\sum_{n=1}^N (d_n - y_n)^2}{N}}}{|\max(d_n) - \min(d_n)|} 100\% ,$$

где  $N$  – количество элементов выборки,  $d_n$  – фактическое значение моделируемой переменной,  $y_n$  – значение переменной, вычисленное с помощью нейронной сети.

Данная средняя квадратичная относительная ошибка тестирования нейронной сети является допустимой для данных, собранных вручную из социальной сети.



**Рисунок 7. Результат тестирования нейронной сети**

Присмотревшись к полученному графику, можно заметить, что присутствуют отклонения в предсказаниях нейронной сети. Это подтверждается и коэффициентом детерминации, который близок к 0. В будущем планируется уменьшение таких отклонений с более глубокой работой с исходными данными для обучения нейронной сети. Например, планируется строить отдельные модели для отдельных профилей. Так как у каждого профиля своя аудитория и предпочтения, соответственно есть основания для проектирования и обучения сети на фотографиях конкретного пользователя.

**Результаты.** Оценка значимости входных параметров была выполнена с помощью программы «Нейросимулятор 5». Была оценена значимость на обучающем и на тестовом множествах, далее был вычислен средний результат и по нему построена диаграмма, которая представлена на рисунке 8.

Как видно из рисунка 8, наиболее значимым цветом на фотографиях является жёлтый. Теперь необходимо понять положительно ли влияет наличие жёлтого цвета на фотографии или отрицательно.

Для этого был применён метод сценарного прогнозирования [4-8]. Необходимо было рассмотреть ранее неопубликованные фотографии, которые не входили в обучающее, тестовое и валидирующее множества. Фотографии и их цвета представлены на рисунке 9.

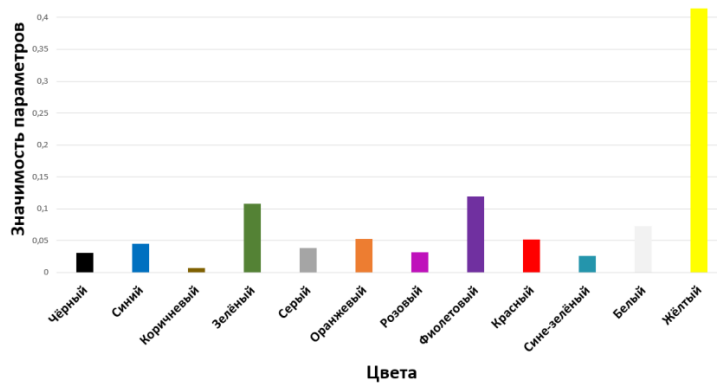


Рисунок 8. Значимость входных параметров нейросетевой модели

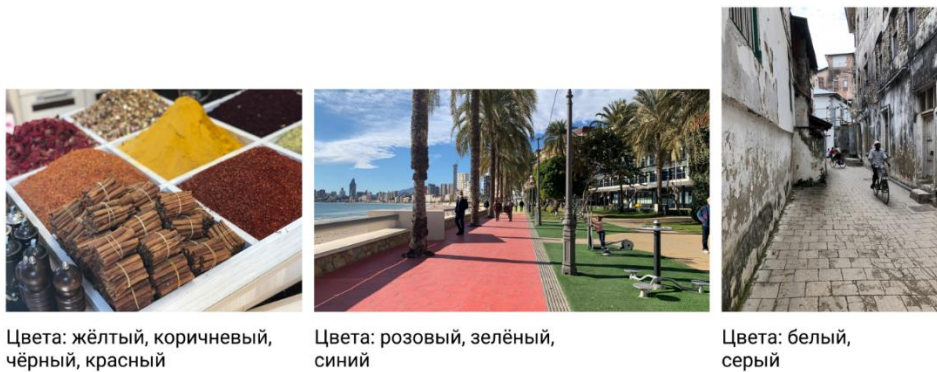


Рисунок 9. Фотографии и их цвета для прогноза

В Нейросимулятор были внесены следующие записи для прогноза с ожидаемым результатом, которые представлены в таблице 1.

Таблица 1

Цвета фотографий для прогноза и ожидаемый результат

Фото	Цвета	Ожидаемый результат
1	<b>Жёлтый</b> , коричневый, чёрный, красный	Высокое количество «лайков»
1	Коричневый, чёрный, красный	Количество «лайков» ниже, чем в первом случае
2	Розовый, зелёный, синий	Низкое количество «лайков»
2	Розовый, зелёный, синий, <b>жёлтый</b>	Количество «лайков» выше, чем в первом случае
3	Белый, серый	Низкое количество «лайков»
3	Белый, серый, <b>жёлтый</b>	Количество «лайков» выше, чем в первом случае

Записи были занесены в Нейросимулятор и был построен следующий прогноз отметок «нравится», который представлен на рисунке 10.

Как оказалось, ожидаемые результаты полностью совпали с прогнозом Нейросимулятора. Жёлтый цвет действительно определяет успех первой фотографии и его присутствие может повысить успех других фотографий. Получается, что яркие фотографии с присутствием жёлтого цвета самые выигрышные для публикации на странице в социальной сети. Это может быть



объяснено тем, что психологически жёлтый цвет у людей ассоциируется с солнцем, радостью. Он нравится творческим людям и поднимает настроение.

	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12	Y1
1	1	0	1	0	0	0	0	0	1	0	0	1	80.7807
2	1	0	1	0	0	0	0	0	1	0	0	0	20.1283
3	0	1	0	1	0	0	1	0	0	0	0	0	9.7598
4	0	1	0	1	0	0	1	0	0	0	0	1	52.3056
5	0	0	0	0	1	0	0	0	0	0	1	0	18.5041
6	0	0	0	0	1	0	0	0	0	0	1	1	50.4837

**Рисунок 10. Прогнозы нейросимулятора**

**Заключение.** В рамках данной работы была построена система нейросетевого прогнозирования количества отметок «нравится» для цветов фотографий профиля социальной сети. Спроектированная нейросетевая модель учитывает 12 цветов: чёрный, синий, коричневый, зелёный, серый, оранжевый, розовый, фиолетовый, красный, сине-зелёный, белый и жёлтый цвета. Оказалось, что наиболее всего на высокое количество «лайков» влияет присутствие жёлтого цвета на фотографиях.

Похожая нейросеть была также построена на языке python. Полученная нейросеть почти не отличалась от полученной в модели в нейросимуляторе.

#### **Библиографический список**

1. Крылова Дарья Сергеевна Влияние социальных сетей на рост компаний малого и среднего бизнеса // Стратегии бизнеса. 2017
2. Черепанов Ф.М., Ясницкий Л.Н. Нейросимулятор 4.0. Свидетельство о регистрации программы для ЭВМ RUS 2014612546. Заявка № 2014610341 от 15.01.2014.
3. Ясницкий Л.Н. Интеллектуальные системы: учебник. М.: Лаборатория знаний, 2016. 221 с.
4. Ясницкий Л.Н., Павлов И.В., Черепанов Ф.М. Прогнозирование результатов олимпийских игр 2014 года в неофициальном командном зачете методами искусственного интеллекта // Современные проблемы науки и образования. 2013. № 6. С. 985.
5. Ясницкий Л.Н., Vnukova O.V. Прогноз результатов олимпиады-2014 в мужском одиночном фигурном катании методами искусственного интеллекта // Современные проблемы науки и образования. 2014. № 1. С. 189.
6. Ясницкий Л.Н., Кировоса А.В., Ратегова А.В., Черепанов Ф.М. Прогноз результатов чемпионата мира-2015 по легкой атлетике методами искусственного интеллекта // Современные проблемы науки и образования. 2014. № 4. С. 624.
7. Ясницкий Л.Н. Введение в искусственный интеллект. М.: Издательский центр «Академия», 2005. – 176 с.



# NEURAL NETWORK SYSTEM FOR ASSESSING THE DEPENDENCE OF THE COLORS OF A PROFILE PHOTO IN A SOCIAL NETWORK AND THE NUMBER OF LIKES

*Sharipova Alisa O.*

National Research University “Higher School of Economics”  
Studencheskaya Ulitsa, 38, Perm, Russia, 614070, infoperm@hse.ru

The article considers the degree of current development of promoting profiles in a social network using machine learning algorithms. It also describes the acquisition and preparation of data for training a neural network and how its training was carried out. Then, the quality of network training was checked. After that, a pattern of practical importance is revealed, and a neural network forecast of the number of “likes” for several new photos is also presented.

**Keywords:** artificial intelligence, neural network, social networks, promotion, photo analysis.

УДК 004.032.26

## НЕЙРОСЕТЕВАЯ СИСТЕМА ПРОГНОЗИРОВАНИЯ ИНДЕКСА СЧАСТЬЯ СТРАНЫ

*Мельников Павел Александрович*

Пермский государственный национальный исследовательский университет, ПМИ. 614990, Россия, г. Пермь, ул. Букирева, 15, pav3luchenik@yandex.ru

В статье представлено описание разработки нейросетевой системы для прогнозирования индекса счастья в стране. Система позволяет с большой точностью предсказать индекс выбранной страны на основании данных о определенных показателях. С помощью разработанной интеллектуальной системы проведено исследование предметной области, выявлены закономерности, имеющие практической значение.

**Ключевые слова:** искусственный интеллект, нейросетевые технологии, прогнозирование, счастье, индекс.

**Введение.** В последнее время в мировом сообществе набирают популярность нейросетевые технологии, многие всемирные организации все чаще стали использовать их в расчетах и прогнозировании тех или иных событий и процессов, а также для выявления важных факторов, влияющих на конечный результат. При анализе литературных источников, выяснилось, что работ на тему прогнозирования рейтингов, индексов и других характеристик было проведено мало. Так ООН предложили использовать уровень ВВП на душу населения, социальную поддержку граждан, продолжительность жизни, уро-

вень щедрости и достатка, отношение к коррупции и кумовству, свобода слова и жизни граждан.

Основная цель настоящей работы заключается в сборе множества публичных данных о странах, а также создание и обучение нейросетевой модели этих данных. Конечным результатом должна быть нейросетевая система, способная с высокой точностью прогнозировать индекс счастья выбранной страны.

Для создания нейросетевой системы были выбраны следующие параметры: X1 – Остаточная антиутопия (дистопия), X2 – ВВП на душу населения, X3 – Социальная поддержка, X4 – Ожидаемая продолжительность здоровой жизни, X5 – Свобода граждан к самостоятельным решениям, X6 – Уровень щедрости, X7 – Отношение к коррупции. Выходной параметр – индекс счастья страны в интервале от 0 до 10 включительно.

Множество примеров было собрано опираясь на интернет ресурсы [5, 6] от каждой страны мира были взяты нужные параметры. Примеры проверялись на адекватность данных. Также проверялось существование страны в двадцать первом веке. Также были исключены страны, у которых не было полных данных, необходимых для прогнозирования индекса счастья. Таким образом, объем итогового множества включает в себя 146 примеров.



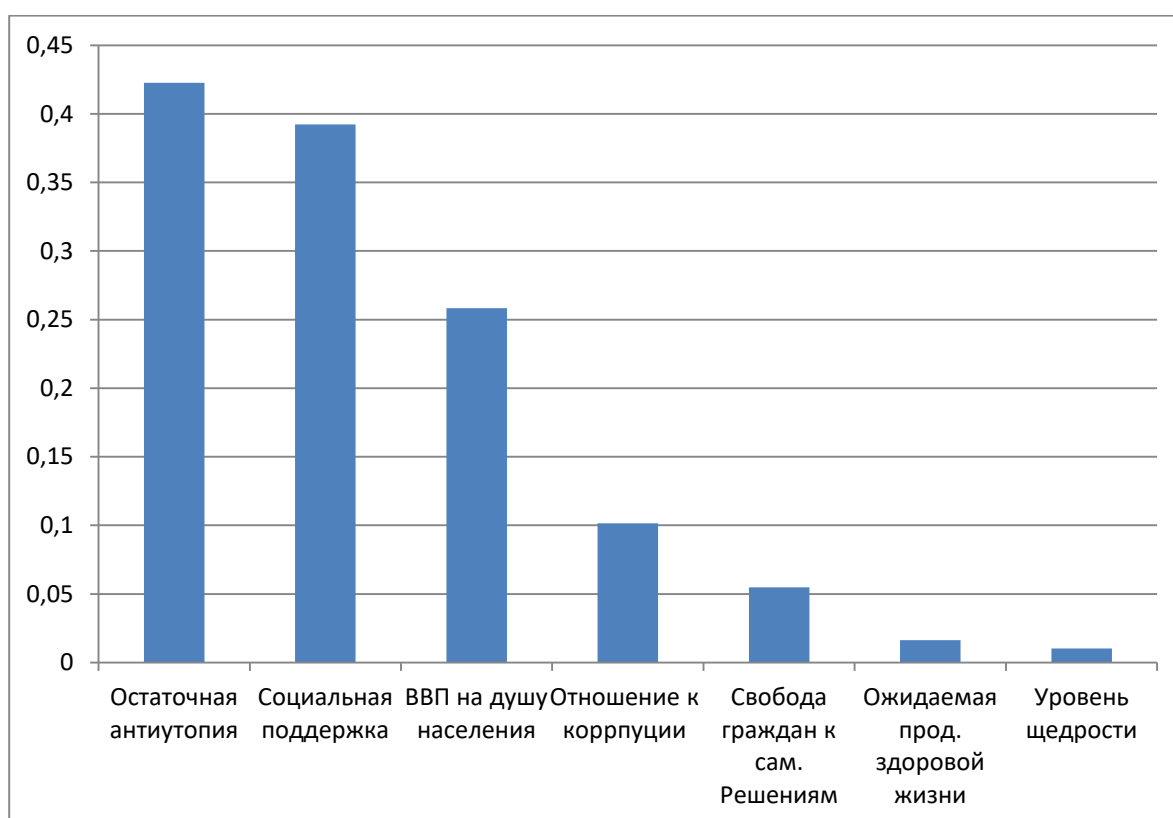
**Рисунок 1. Результат тестирования нейронной сети**

Проектирование, обучение, тестирование нейронной сети выполнялись с помощью программы «Нейросимулятор 5» [7] по методике [8]. В результате оптимизации спроектированная нейронная сеть представляла собой персептрон, который имеет семь входных нейронов, один выходной и один скрытый слой с двумя нейронами.

Для оценки корректной работы спроектированной нейронной сети использовалось тестирующее множество, состоящее из 24 примеров. Средняя относительная ошибка тестирования составила 1.02%, что можно считать приемлемым результатом. На рисунке 1 представлена гистограмма, демонстрирующая разницу между фактическим и прогнозируемым значением индекса счастья.

Из результатов, изображенных на рисунке 1, можно сделать вывод об адекватной работе спроектированной нейронной сети.

Оценка значимости входных параметров была выполнена с помощью программы «Нейросимулятор 5», результат отображен на рисунке 2.



**Рисунок 2. Значимость входных параметров нейросетевой модели**

Как видно из рисунка 2, наиболее значимыми являются значение остаточной антиутопии, социальная поддержка, ВВП на душу населения и т.д. Как и ожидалось, наиболее влиятельными параметрами являются Остаточная антиутопия и ВВП на душу населения. Индекс счастья, напрямую связан с уровнем и качеством жизни в стране, если у страны высокий уровень Валового внутреннего продукта, это говорит о росте экономики, о повышении зарплат и укреплении национальной валюты. Также остаточная антиутопия влияет на эмоциональное и душевное состояние жителей страны, из-за низкого уровня этого параметра, люди чаще впадают в депрессии, чаще совершают попытки суицида и становятся несчастными.

Далее было проведено исследование полученных зависимостей между входными параметрами и полученным индексом. Исследование производилось с помощью метода «замораживания» [1, 8], суть которого заключается в

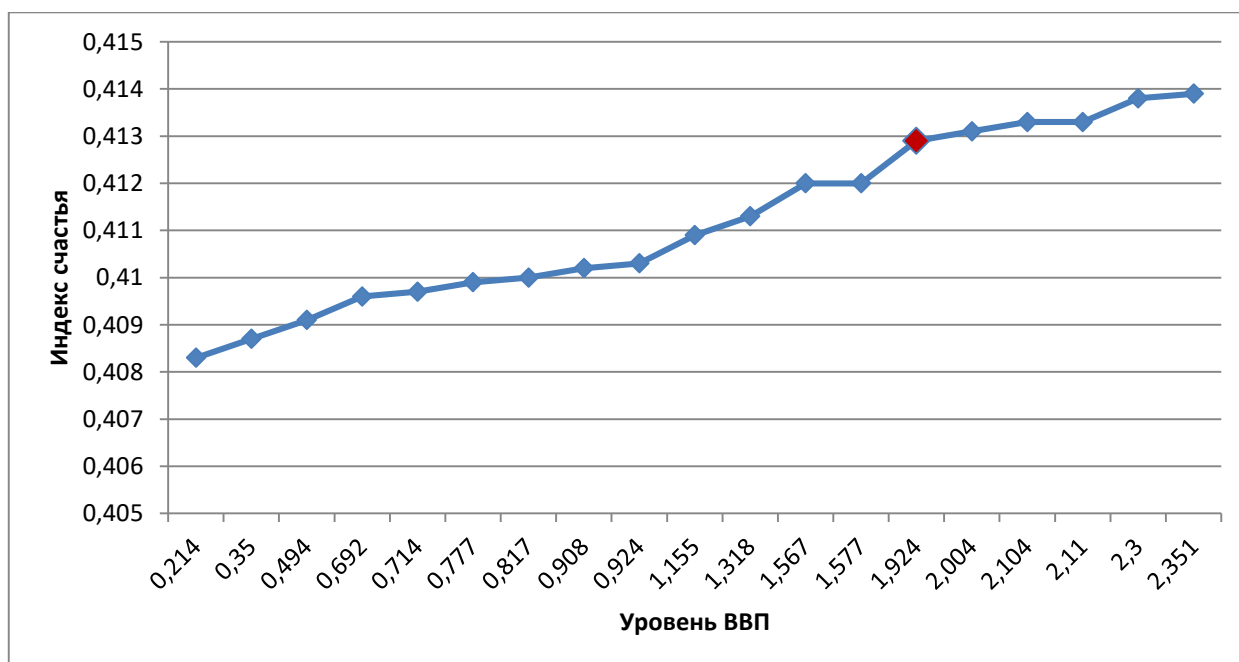
варьировании значения одного параметра и фиксировании значений всех других параметров. Данный метод позволяет выявить влияние исследуемого параметра на выходной. Для этой цели была отобрана страна (Германия), информация о параметрах которого отражена в таблице 1.

**Таблица 1**

**Характеристики страны, выбранной для исследования**

Входные параметры	Расшифровка	Характеристики
X1	Остаточная антиутопия	2,142
X2	ВВП на душу населения	1,924
X3	Социальная поддержка	1,088
X4	Ожидаемая продолжительность здоровой жизни	0,776
X5	Свобода граждан к самостоятельным решениям	0,585
X6	Уровень щедрости	0,163
X7	Отношение к коррупции	0,358

На рисунке 3 показан график зависимости прогнозируемого индекса. В том случае, когда уровень ВВП страны высокий, вероятность на получение наибольшего индекса счастья повышается и нейросеть прогнозирует повышение уровня счастья в стране.



**Рисунок 3. Зависимость индекса счастья от величины ВВП страны**

**Заключение.** Построена система нейросетевого прогнозирования и расчета индекса счастья страны. Спроектированная нейросетевая модель учитывает 7 параметров: Остаточная антиутопия (дистопия), уровень ВВП на душу населения, социальная поддержка, ожидаемая продолжительность здоровой жизни, свобода граждан к самостоятельным решениям, уровень щедрости, отношение к коррупции. Методом сценарного прогнозирования по-

строены графики зависимостей индекса счастья страны от изменения входных параметров. Применение такого набора параметров в модели позволяет с высокой точностью прогнозировать индекс счастья.

### Библиографический список

1. <https://worldhappiness.report/>
2. [kaggle.com/datasets/hemil26/world-happiness-report-2022](https://kaggle.com/datasets/hemil26/world-happiness-report-2022)
3. <https://worldpopulationreview.com/country-rankings/happiest-countries-in-the-world>
4. [stud-me.org/197430/sotsiologiya/sotsialnoe\\_prognozirovanie\\_proektirovanie\\_modelirovanie](https://stud-me.org/197430/sotsiologiya/sotsialnoe_prognozirovanie_proektirovanie_modelirovanie)
5. [kaggle.com/datasets/hemil26/world-happiness-report-2022](https://kaggle.com/datasets/hemil26/world-happiness-report-2022)
6. [un.org/ru/](https://un.org/ru/)
7. Черепанов Ф.М., Ясницкий Л.Н. Нейросимулятор 4.0. Свидетельство о регистрации программы для ЭВМ RUS 2014612546. Заявка № 2014610341 от 15.01.2014.
8. Ясницкий Л.Н. Введение в искусственный интеллект. М.: Издательский центр «Академия», 2005. – 176 с.

## NEURAL NETWORK SYSTEM FOR FORECASTING THE COUNTRY'S HAPPINESS INDEX

*Melnikov Pavel A.*

Perm State University

Str. Bukireva, 15, Perm, Russia, 614990, pav3luchenik@yandex.ru

The article describes the development of a neural network system for predicting the happiness index of a country. The system allows you to predict the index of the selected country with great accuracy on the basis of data on certain indicators. With the help of the developed intelligent system a study of the subject area was carried out, the regularities of practical importance were identified.

**Keywords:** artificial intelligence, neural network, prediction, happy, index.

## НЕЙРОСЕТЕВАЯ СИСТЕМА ПРОГНОЗИРОВАНИЯ УПОТРЕБЛЕНИЯ АЛКОГОЛЯ СТУДЕНТАМИ

*Густокашина Виктория Михайловна*

Пермский государственный национальный исследовательский университет  
ПМИ. 614990, Россия, г. Пермь, ул. Букирева, 15,  
vgustokashina@gmail.com

В статье представлено описание разработки нейросетевой системы для прогнозирования употребления алкоголя студентами. Система позволяет предсказать количество употребляемого студентами алкоголя в неделю. С помощью разработанной интеллектуальной системы проведено исследование предметной области, выявлены закономерности, имеющие практического значения.

**Ключевые слова:** искусственный интеллект, нейросетевые технологии, прогнозирование, алкоголь, спиртные напитки, студенты.

**Введение.** Согласно Докладу Общественной Палаты РФ о злоупотреблении алкоголем в Российской Федерации, потери государства, связанные с алкоголизмом, составляют около 1,7 триллиона рублей в год. В эту сумму входит содержание медицинских учреждений, направленных на лечение алкогольной зависимости, выплата пособий, расходы на обеспечение лиц, совершивших уголовные преступления в состоянии алкогольного опьянения, покрытие издержек ущерба чрезвычайных ситуаций (пожары, ДТП), причиной которых было употребление алкоголя [1]. Одной из наиболее восприимчивых к преобразованиям социальных групп является студенчество. В начале 2000-х годов прослеживался высокий уровень потребления алкоголя среди молодежи [2]. Рассмотрим какие факторы их жизни в большей степени влияют на это. Выявление причин алкоголизма в будущем поможет улучшить как физическое, так и моральное состояние жителей страны.

Работ на тему потребления алкоголя среди студентов много, но проблема остается актуальной. Исследователи здоровья студентов В.А. Медик и А.М. Осипов утверждают, что алкоголь укоренился в вузовской системе на протяжении десятилетий в виде традиции. Погруженность части студентов в алкогольную традицию ведет к искаженному восприятию ими общественного мнения по проблемам алкоголя [3]. Исследователи здоровья студентов отмечают, что именно употребление алкоголя является самой распространенной вредной привычкой среди студентов. Алкоголь с разной частотой употребляют 40,1% учащихся вузов.

Основная цель настоящей работы заключается в сборе множества данных о студентах и их условиях жизни, а также создание и обучение нейросетевой модели на этих данных. Конечным результатом должна быть нейросе-

тевая система, способная прогнозировать количество выпиваемого алкоголя в больше чем 70% случаев.

Для создания нейросетевой системы были выбраны следующие параметры:

X1 – пол (1 – женский, 2 – мужской)

X2 – возраст,

X3 – кол-во детей в семье (1 – 3 или меньше, 2 – больше 3)

X4 – образование матери (0 – нет, 1 – начальное образование, 2 – 5-9-й классы, 3 – среднее образование, 4 – высшее образование),

X5 – образование отца (0 – нет, 1 – начальное образование, 2 – 5-9-й классы, 3 – среднее образование, 4 – высшее образование),

X6 – работа матери (1 – "учитель", 2 – "здравоохранение", 3 – "службы", 4 – "на дому", 5 – "другие"),

X7 – работа отца (1 – "учитель", 2 – "здравоохранение", 3 – "службы", 4 – "на дому", 5 – "другие"),

X8 – путь от дома до учебного заведения (1 – <15 мин., 2 – 15-30 мин., 3 – 30 мин.-час, от 4 – >1 часа),

X9 – учебное время (1 – <2 часов, 2 – 2-5 часов, 3 – 5-10 часов, 4 – >10 часов),

X10 – кол-во пересдач (4 – больше трех),

X11 – посещение детского сада (0 – нет, 1 – да),

X12 – состоит в романтических отношениях (0 -нет, 1 – да),

X13 – качество семейных отношений (от 1 – очень плохо до 5 – отлично),

X14 – прогулки с друзьями (от 1 – редко до 5 – часто),

X15 – состояние здоровья (от 1 – очень плохое до 5 – очень хорошее)

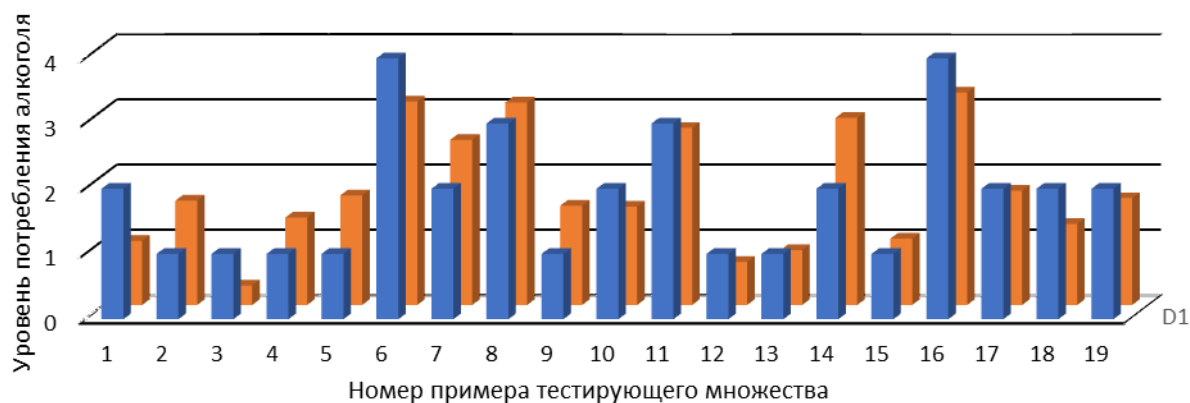
X16 – кол-во пропусков.

Выходной параметр – потребление алкоголя в выходные дни (числовое значение: от 1 – очень низкое до 5 – очень высокое).

Обучающее множество было взято с сайта kaggle [4]. Перед переходом к проектированию и обучению нейросетевой модели, была выполнена очистка исходного множества от противоречивых примеров, выбросов, дубликатов. Таким образом, объем итогового множества включает в себя 693 примера. Данное множество было разделено на обучающее, валидационное и тестирующее в соотношении 70%, 20% и 10% соответственно.

Проектирование, обучение, тестирование нейронной сети выполнялись с помощью программы «Нейросимулятор 5» [5] по методике [6]. В результате оптимизации спроектированная нейронная сеть представляла собой персептрон, который имеет шестнадцать входных нейронов, один выходной, без скрытых слоев.

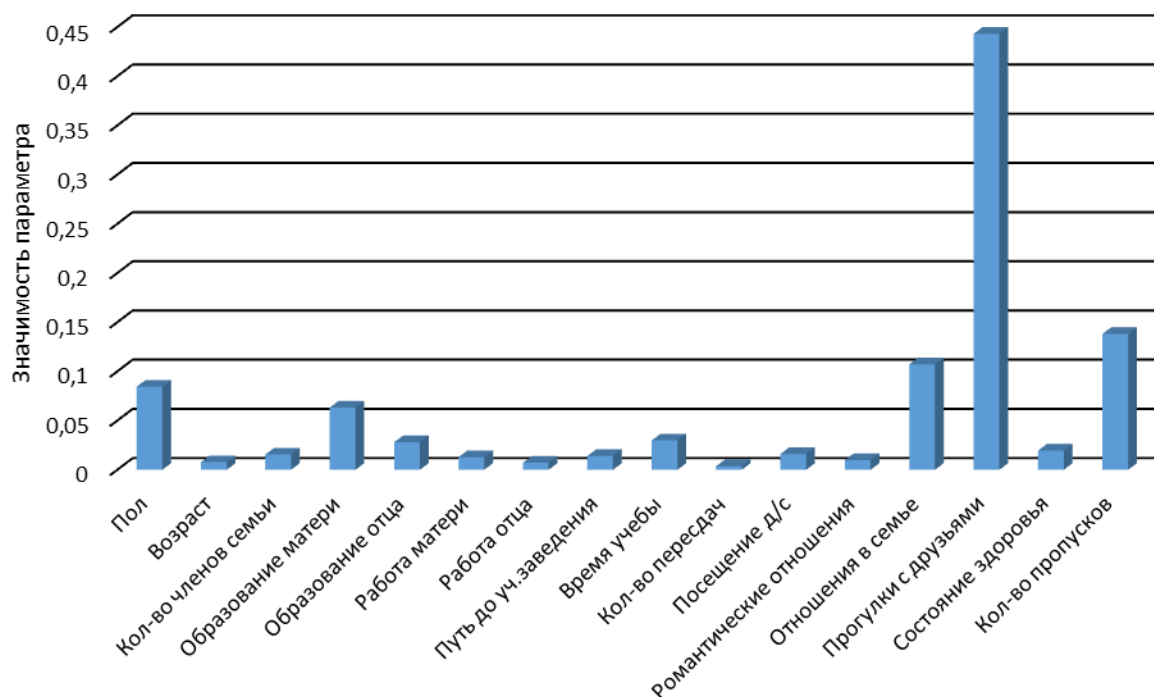
Для оценки корректной работы спроектированной нейронной сети использовалось тестирующее множество, состоящее из 80 примеров. Средняя относительная ошибка тестирования составила 14.7%, что можно считать приемлемым результатом. На рисунке 1 представлена гистограмма, демонстрирующая разницу между фактическим и прогнозируемым количеством потребляемого алкоголя.



**Рисунок 1. Результат тестирования нейронной сети**

Из результатов, изображенных на рисунке 1, можно сделать вывод об адекватной работе спроектированной нейронной сети.

Оценка значимости входных параметров была выполнена с помощью программы «Нейросимулятор 5», результат отображен на рисунке 2.



**Рисунок 2. Значимость входных параметров нейросетевой модели**

Как видно из рисунка 2, наиболее значимыми являются прогулки с друзьями, кол-во пропусков, качество отношений в семье и т.д. На поведение человека оказывает сильное влияние его окружение. Проблемы в семье, ссоры с друзьями могут сказаться на душевном состоянии студента, из-за чего могут возникнуть проблемы с алкоголем.

Далее было проведено исследование полученных зависимостей между входными параметрами и количеством потребляемого алкоголя. Исследование производилось с помощью метода «замораживания» [6], суть которого заключается в варьировании значения одного параметра и фиксировании

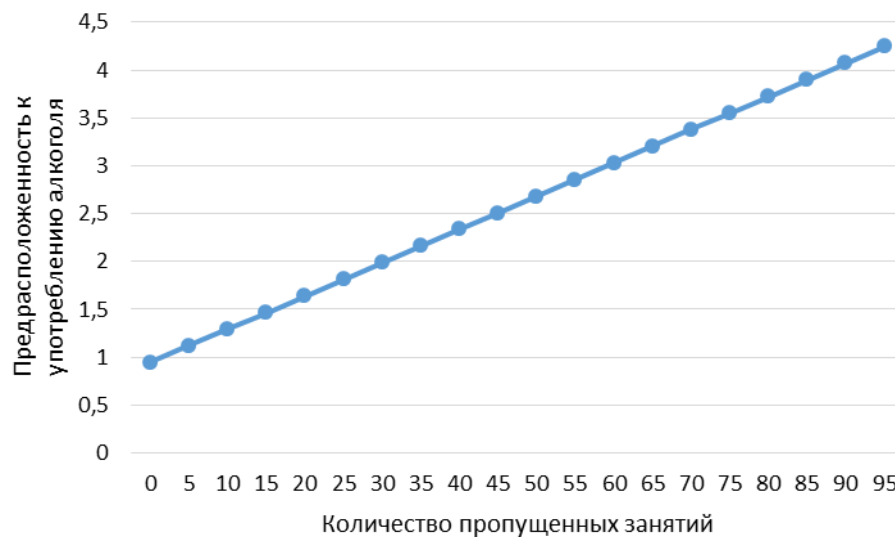


значений всех других параметров. Для этой цели была отобрана 17 – летняя студентка с хорошим состоянием здоровья, не пропускающая занятия.

Входные параметры выбранного примера:

X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12	X13	X14	X15	X16
1	17	2	4	3	1	5	2	3	0	1	1	4	2	4	0

На рисунке 3 показан график зависимости предрасположенности к употреблению алкоголя от количества пропущенных занятий. Чем больше занятий пропущено, тем чаще студент употребляет алкоголь.



**Рисунок 3. Зависимость предрасположенности к употреблению алкоголя от количества пропущенных занятий**

На рисунке 4 продемонстрирована зависимость предрасположенности к употреблению алкоголя от частоты прогулок с друзьями. Можно заметить, что количество прогулок негативно влияет на потребление алкоголя.



**Рисунок 4. Зависимость предрасположенности к употреблению алкоголя от частоты прогулок с друзьями**

Полученные результаты исследований не противоречат реальности, что подтверждает, что спроектированную нейронную сеть можно считать пригодной для прогнозирования потребления алкоголя студентами.

**Заключение.** Построена система нейросетевого прогнозирования употребления алкоголя студентами. Спроектированная нейросетевая модель учитывает 16 параметров: пол, возраст, количество членов семьи, образование и должность обоих родителей, путь от дома до учебного заведения, время учебы, количество пересдач, посещение детского сада, романтические отношения, качество отношений в семье, частоту прогулок с друзьями, текущее состояние здоровья и количество пропусков занятий. Методом сценарного прогнозирования построены графики зависимостей количества потребляемого алкоголя от изменения входных параметров. Применение такого набора параметров в модели позволяет с высокой точностью прогнозировать уровень употребления спиртных напитков. Данный набор параметров может быть изменен для прогнозирования количества употребляемого алкоголя для других возрастных групп.

### **Библиографический список**

1. Комиссия Общественной палаты Российской Федерации по социальной и демографической политике, Общественный совет Центрального федерального округа, Доклад, Москва, 2009
2. Шрага М. Х. Арктика и Север 2011 № 4(ноябрь) / М. Х. Шрага, К. Колчина, Н. Сморгалова // Алкоголизация студентов г. Архангельска в начале 21 века. — Архангельск 2011. с. 122–134.
3. Медик В.А., Осипов А.М. Университетское студенчество: образ жизни и здоровье. М.: Логос, 2003; 200 с.
4. Kaggle.com – сообщество специалистов по обработке данных, <https://www.kaggle.com/datasets/uciml/student-alcohol-consumption>.
5. Черепанов Ф.М., Ясницкий Л.Н. Нейросимулятор 4.0. Свидетельство о регистрации программы для ЭВМ RUS 2014612546. Заявка № 2014610341 от 15.01.2014.
6. Ясницкий Л.Н. Введение в искусственный интеллект. М.: Издательский центр «Академия», 2005. – 176 с.

## **NEURAL NETWORK SYSTEM FOR PREDICTING ALCOHOL CONSUMPTION BY STUDENTS**

*Gustokashina Viktoria M.*

Perm State University

Str. Bukireva, 15, Perm, Russia, 614990, [vgustokashina@gmail.com](mailto:vgustokashina@gmail.com)

The article describes the development of a neural network system for predicting alcohol consumption by students. The system allows you to predict the amount of alcohol consumed by students per week. With the help of the devel-

oped intellectual system, a study of the subject area was carried out, patterns of practical importance were identified.

**Keywords:** artificial intelligence, neural network technologies, forecasting, alcohol, alcoholic beverages, students.

УДК 004.032.26

## НЕЙРОСЕТЕВАЯ СИСТЕМА ПРОГНОЗИРОВАНИЯ УРОВНЯ УСПЕХА В ИТ-СФЕРЕ

*Кузьминых Ульяна Ивановна*

Пермский государственный национальный исследовательский университет,  
КМБ. 614990, Россия, г. Пермь, ул. Букирева, 15, ulyana280701111@yandex.ru

В статье представлено описание разработки нейросетевой системы для прогнозирования уровня успеха в ИТ-сфере. Система позволяет с большой точностью предсказать уровень успеха в ИТ-сфере на основании определённых данных о человеке.

**Ключевые слова:** искусственный интеллект, нейросетевые технологии, прогнозирование, ИТ, профессия, программирование.

**Введение.** В данной работе рассматриваются люди, которые никогда не изучали ИТ сферу, никогда в ней не работали, но хотели бы с ней познакомиться и добиться успеха. Часто люди не знают, получится ли у них реализоваться в ИТ сфере, а времени на попытки не всегда достаточно. Это исследование может помочь школьникам, которые выбирают направление своей дальнейшей учёбы, профессию или людям, которые решили сменить свою сферу деятельности и попробовать себя в чём-то новом. Как показал опыт Пермской научной школы искусственного интеллекта, например, [1-2], правильно натренированная нейросетевая система может добиться успешности большей части прогнозов. Знание уровня своего возможного успеха может помочь школьникам и людям, ищущим работу в ИТ сфере сэкономить время на попытки стать успешными в ИТ сфере.

При анализе электронных источников выяснилось, что существует достаточно много материала про критерии успешного работника ИТ сферы. Но информации настолько много, что в ней легко запутаться. Также не было найдено информации про связь ИТ сферы и типа темперамента.

Основная цель данной работы заключается в сборе множества данных о людях, которые хотят попробовать себя в ИТ сфере, а также создание и обучение нейросетевой модели на этих данных. Конечным результатом должна быть нейросетевая система, способная прогнозировать уровень успеха человека в ИТ сфере в больше чем 70% случаев.

Для создания нейросетевой системы были выбраны следующие параметры:

- X1 – пол человека (1 – женский, 2 – мужской),
- X2 – возраст человека (от 16 до 55 лет),
- X3 – тип темперамента (1 – холерик, 2 – сангвиник, 3 – флегматик, 4 – меланхолик),
- X4 – наличие высшего образования (1 – есть, 2 – нет).

Выходной параметр – уровень успеха в IT сфере от 0 до 10.

20% обучающего множества было собрано вручную, 80% – смоделировано. Уровень успеха определялся с помощью информации, собранной в источниках [5-7]. После изучения этой информации для каждого параметра обозначился коэффициент важности: пол – 1, возраст – 4, тип темперамента – 3, наличие высшего образования – 2. Далее каждому человеку за каждый параметр назначались баллы:

- Мужской пол – 1 балл, женский – 0 баллов,
- Возраст до 35 лет – 1 балл, после 35 лет – 0 баллов,
- Тип темперамента: сангвиник, флегматик, меланхолик – 1 балл, холерик – 0 баллов,
- Наличие высшего образования – 1 балл, его отсутствие – 0 баллов.

Уровень успеха рассчитывался по формуле, построенной на основе метода векторов предпочтений [8]:

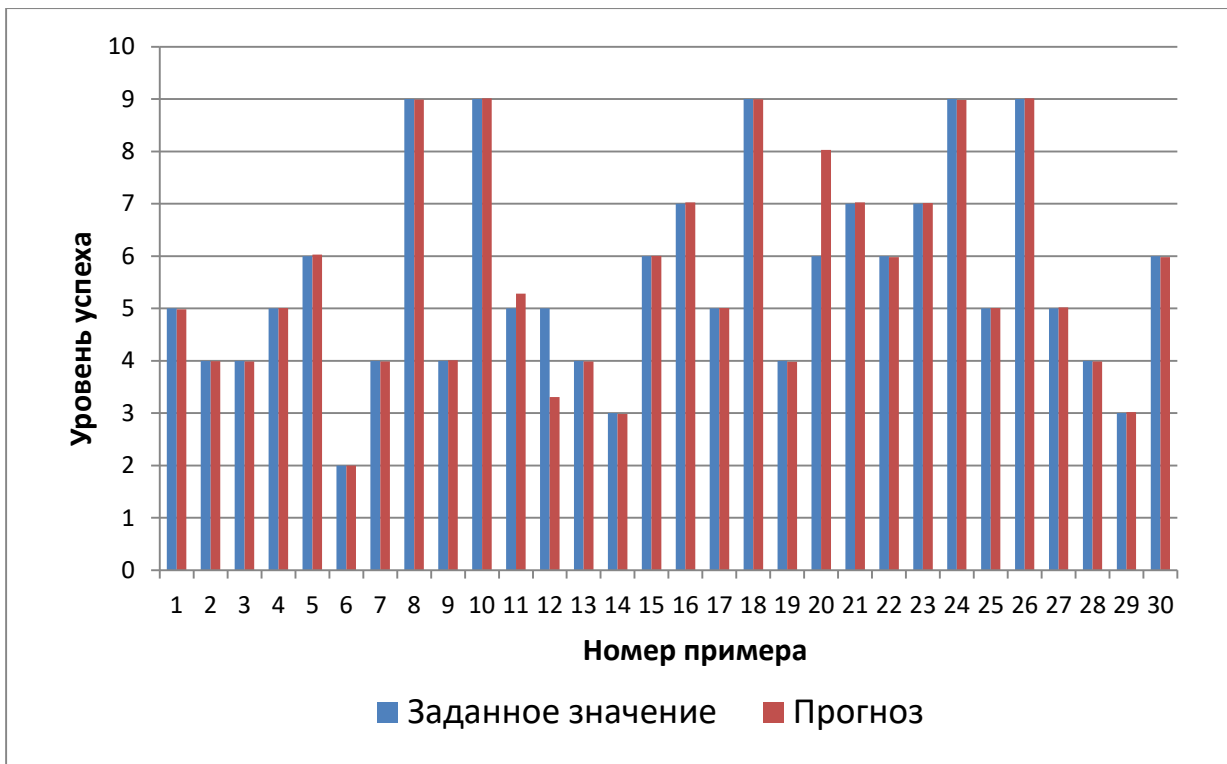
$$\begin{aligned} & \text{кол-во баллов за параметр}_1 * \text{коэффициент важности параметра}_1 + \\ & \text{кол-во баллов за параметр}_2 * \text{коэффициент важности параметра}_2 + \\ & \text{кол-во баллов за параметр}_3 * \text{коэффициент важности параметра}_3 + \\ & \text{кол-во баллов за параметр}_4 * \text{коэффициент важности параметра}_4. \end{aligned}$$

Перед переходом к проектированию и обучению нейросетевой модели, была выполнена очистка исходного множества от противоречивых примеров, выбросов, дубликатов. Например, некорректными примерами считались те, где мужчина холерик в возрасте около 55 лет без высшего образования имеет уровень успеха больше 5. Таким образом, объем итогового множества включает в себя 180 примеров. Данное множество было разделено на обучающее и тестирующее в соотношении 80% и 20% соответственно.

Проектирование, обучение, тестирование нейронной сети выполнялись с помощью программы «Нейросимулятор 5» [3] по методике [4]. В результате оптимизации спроектированная нейронная сеть представляла собой персептрон, который имеет четыре входных нейронов, один выходной и один скрытый слой с тремя нейронами.

Для оценки корректной работы спроектированной нейронной сети использовалось тестирующее множество, состоящее из 30 примеров. Средняя относительная ошибка тестирования составила 7%, что можно считать хорошим результатом.

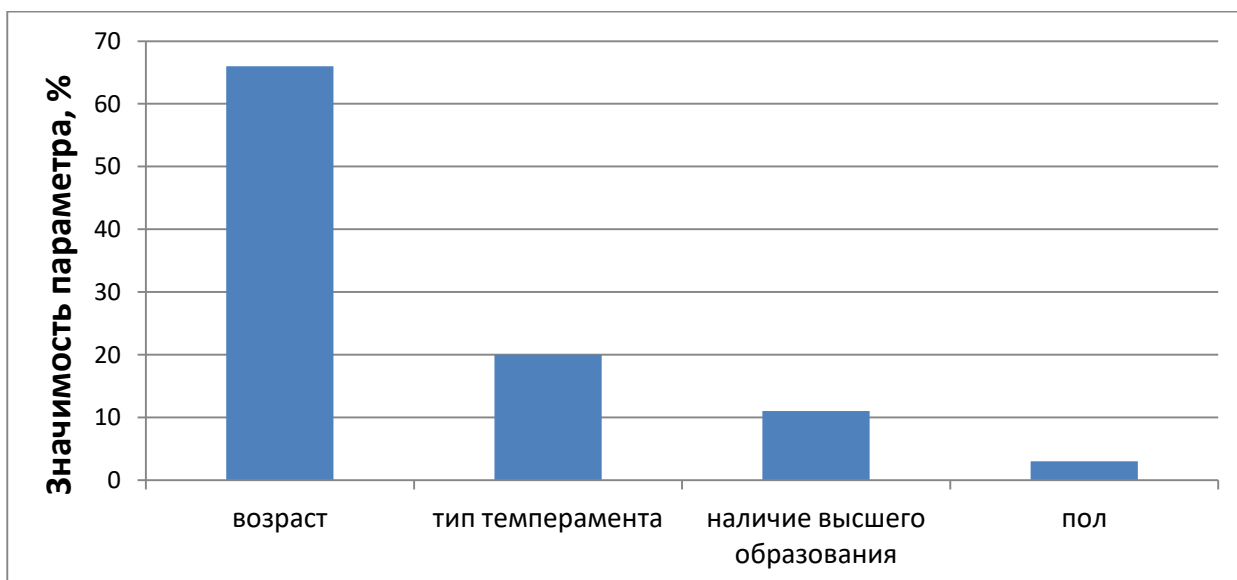
На рисунке 1 представлена гистограмма, демонстрирующая разницу между фактическим и прогнозируемым нейросетью уровнем успеха человека.



**Рисунок 1. Результат тестирования нейронной сети**

Из результатов, изображенных на рисунке 1, можно сделать вывод об адекватной работе спроектированной нейронной сети.

Оценка значимости входных параметров была выполнена с помощью программы «Нейросимулятор 5», результат отображен на рисунке 2.



**Рисунок 2. Значимость входных параметров нейросетевой модели**

Как видно из рисунка 2, наиболее значимым параметром является возраст, т.к. IT сфера начала развиваться относительно недавно и в основном успеха в ней добиваются те, кто начал изучать её в более раннем возрасте. Человеку, никогда не изучавшему эту сферу и решившему добиться в ней

успеха после 35 лет, будет сложнее, т.к. в основном на работу берут более молодых людей.

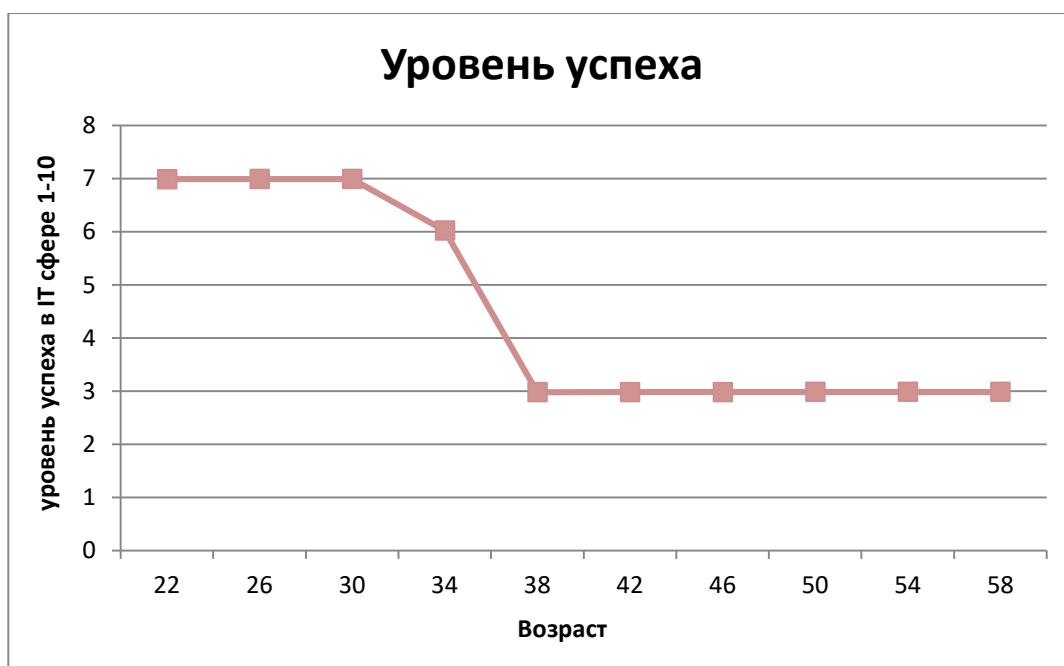
Тип темперамента влияет намного меньше. Почти все типы темперамента могут добиться успеха в этой сфере. Сложнее всего будет холерику, т.к. он не приспособлен к монотонной работе и долгой концентрации, однако такие способности можно развить.

Наличие высшего образования в IT сфере не является сильно значимым параметром, т.к. сейчас существует множество книг и курсов, по которым можно обучиться самостоятельно. Часто на работу берут студентов, которые ещё не закончили своё обучение.

Пол является наименее значимым параметром. По статистике в IT сфере работает больше мужчин, чем женщин, но этот почти не понижает уровень возможного успеха женщины.

Далее было проведено исследование полученных зависимостей между входными параметрами и победителем матча. Исследование производилось с помощью метода «замораживания», суть которого заключается в варьировании значения одного параметра и фиксировании значений всех других параметров. Для этой цели был отобран случайный пример.

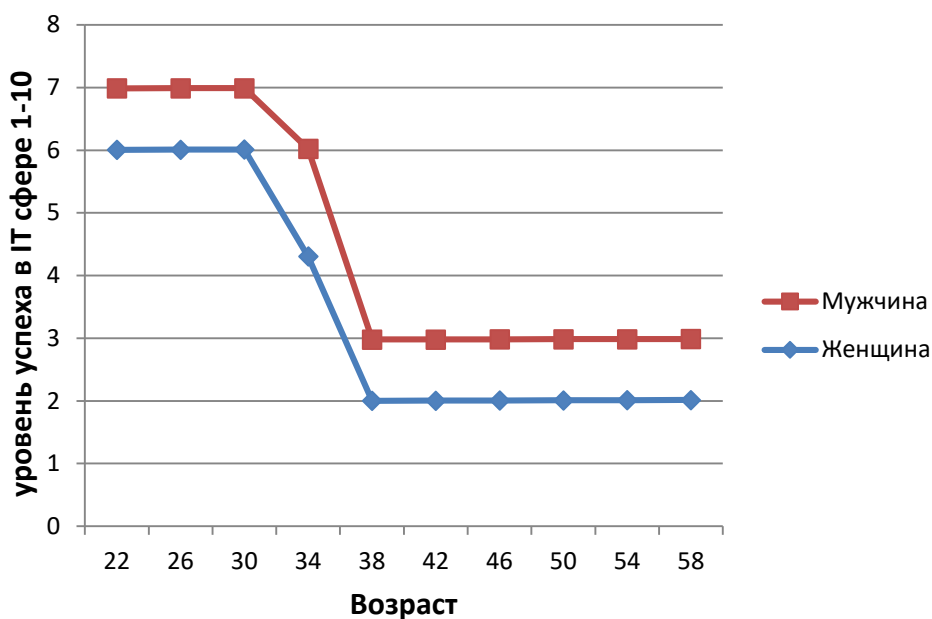
На рисунке 3 показан график зависимости уровня успеха от возраста. Можем заметить, что, чем старше человек, тем ниже становится его уровень успеха. Особенно заметно он падает после 35 лет.



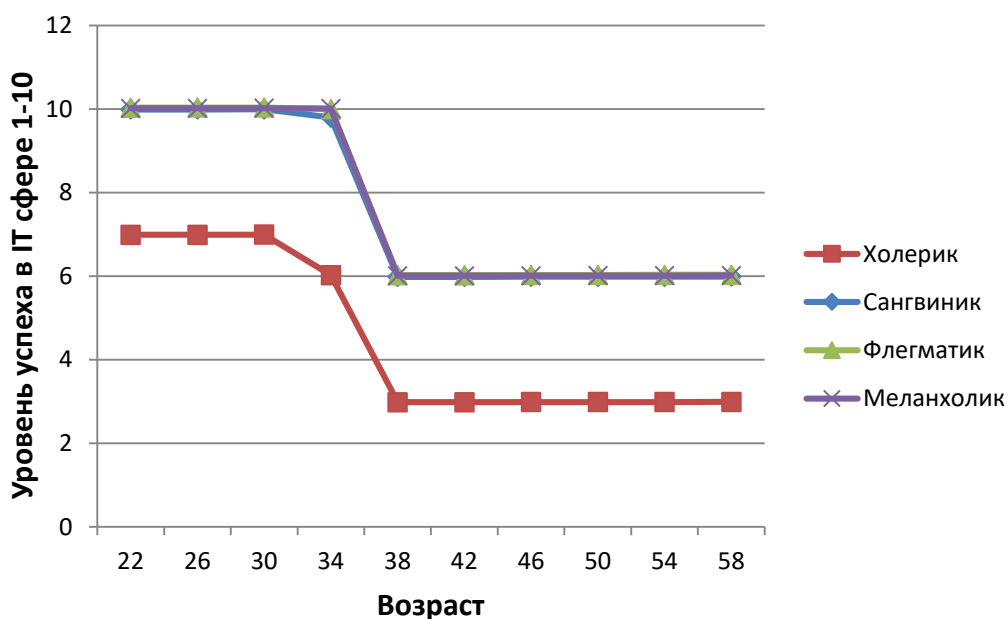
**Рисунок 3. Зависимость уровня успеха от возраста**

На рисунке 4 продемонстрирована зависимость уровня успеха от пола человека. Можно заметить, что у женщин уровень успеха немного меньше.

На рисунке 5 изображен график зависимости уровня успеха от типа темперамента. Как видно из графика, все типы темперамента примерно равны, кроме холерика.



**Рисунок 4. Зависимость уровня успеха от пола человека**



**Рисунок 5. Зависимость уровня успеха от типа темперамента**

На рисунке 6 продемонстрирована зависимость успеха от наличия высшего образования. Заметно, что отсутствие высшего образования достаточно сильно влияет на уровень успеха в IT сфере.

Полученные результаты исследований не противоречат реальности, что подтверждает, что спроектированную нейронную сеть можно считать пригодной для прогнозирования уровня успеха в IT сфере.

**Заключение.** Построена система нейросетевого прогнозирования уровня успеха человека в IT сфере. Спроектированная нейросетевая модель учитывает 4 параметра: пол, возраст, тип темперамента и наличие высшего образования. Методом сценарного прогнозирования построены графики зависимостей уровня успеха от изменения входных параметров.

Применение такого набора параметров в модели позволяет с высокой точностью прогнозировать уровень успеха человека в IT сфере, учитывая то, что он ранее не изучал её.

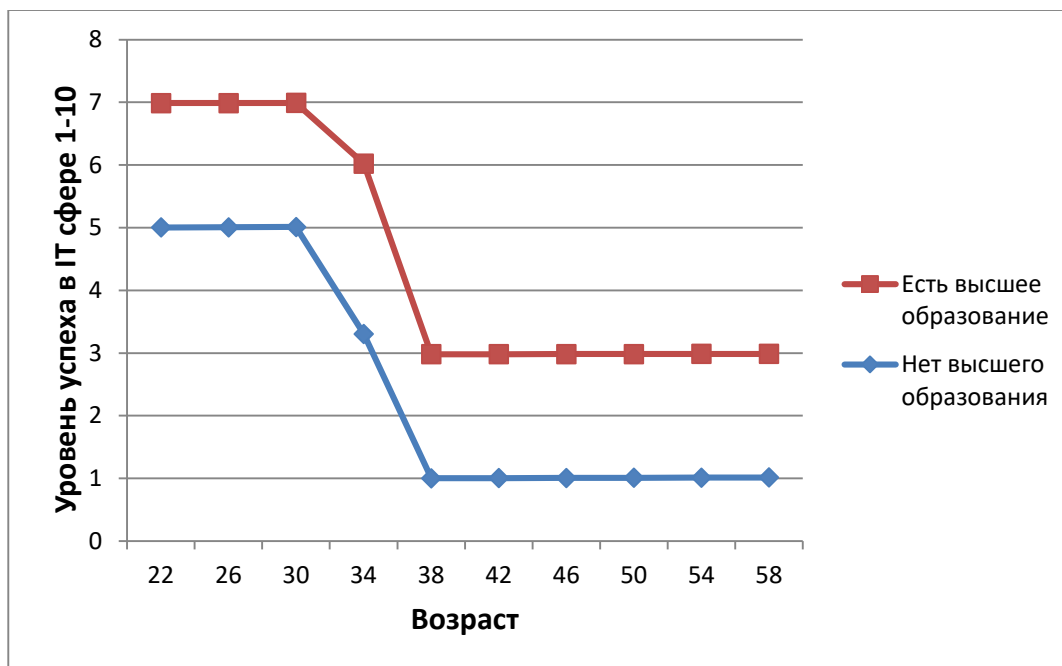


Рисунок 6. Зависимость успеха от наличия высшего образования

### Библиографический список

1. Ясницкий Л.Н., Грацилёв В.И., Куляшова Ю.С., Черепанов Ф.М. Возможности моделирования предрасположенности к наркозависимости методами искусственного интеллекта // Вестник Пермского университета. Философия. Психология. Социология. 2015. № 1 (21). С. 61-73.
2. Ясницкий Л.Н., Сичинава З.И. Нейросетевые алгоритмы анализа поведения респондентов // Нейрокомпьютеры: разработка, применение. 2011. № 10. С. 59-64.
3. Черепанов Ф.М., Ясницкий Л.Н. Нейросимулятор 4.0. Свидетельство о регистрации программы для ЭВМ RUS 2014612546. Заявка № 2014610341 от 15.01.2014.
4. Ясницкий Л.Н. Введение в искусственный интеллект. М.: Издательский центр «Академия», 2005. – 176 с.
5. proforientator.ru – статья Евгении Лепешовой на тему связи типа темперамента и профессии. [Электронный ресурс] Режим доступа: <https://proforientator.ru/publications/articles/temperament-i-budushchaya-professiya.html>
6. habr.com – Результат опроса разработчиков на Stack Overflow 2020. [Электронный ресурс] Режим доступа: <https://habr.com/ru/company/skillfactory/blog/504194/>
7. vc.ru – Результаты исследования Stack Overflow среди разработчиков [Электронный ресурс] Режим доступа: <https://vc.ru/hr/46800-rezultaty-issledovaniya-stack-overflow-sredi-razrabotchikov-chast-1>



8. bstudy.net – Метод векторов предпочтений [Электронный ресурс]
9. Режим доступа: [https://bstudy.net/643614/ekonomika/metod\\_mnozhestvennyh\\_sravneniy](https://bstudy.net/643614/ekonomika/metod_mnozhestvennyh_sravneniy)

## NEURAL NETWORK SYSTEM FOR PREDICTING THE LEVEL OF SUCCESS IN THE IT SPHERE

*Kuzminykh Ulyana Ivanovna*

Perm State University,  
614990, Russia, Perm, Bukireva str., 15, ulyana280701111@yandex.ru

The article describes the development of a neural network system for predicting the level of success in the IT field. The system allows you to accurately predict the level of success in the IT field based on certain data about a person.

**Keywords:** artificial intelligence, neural network technologies, forecasting, IT, profession, programming.

УДК 004.032.26

## НЕЙРОСЕТЕВАЯ СИСТЕМА ПРОГНОЗИРОВАНИЯ УХОДА СОТРУДНИКА ИЗ КОМПАНИИ В ТЕЧЕНИЕ ДВУХ ЛЕТ

*Цыплякова Елена Александровна*

Национальный исследовательский университет «Высшая школа экономики»  
614060, Россия, г. Пермь, ул. Бульвар Гагарина, 37, eatsyplyakova@edu.hse.ru

В статье представлено описание разработки нейросетевой системы прогноза ухода сотрудника из компании. Система позволяет прогнозировать покинет ли будущий сотрудник компанию в течение двух лет. Было проведено ряд исследований, результаты которых приводятся в данной статье.

**Ключевые слова:** компания, сотрудники, уход, прогноз, нейронная сеть, уровень образования, год вступления в компанию, уровень зарплаты, возраст сотрудника, опыт работы в текущей сфере.

**Введение.** Неожиданный уход сотрудника из компании – это всегда проблема для бизнеса. Не только потому, что потеря ключевого сотрудника вредит результатам в бизнесе, но и потому, что руководители не любят сюрпризы [1]. Текучесть кадров дорого обходится компаниям. Исследования Центра Американского прогресса говорят, что компании обычно тратят около одной пятой части годовой зарплаты сотрудника, чтобы найти ему замену

[2]. Именно поэтому отделу кадров компании важно предсказывать, покинут ли некоторые сотрудники компанию в ближайшие 2 года.

При анализе литературных источников выяснилось, что работ на тему прогнозирования результатов ухода сотрудников из компании было проведено достаточно. Однако, было найдено лишь одно исследование в данной области, при котором использовались методы машинного обучения [3]. Каждая компания разрабатывает свой процесс идентификации ухода сотрудника. Обычно такие методы идентификации включают в себя индивидуальные «stay interviews», которые помогают выявить основные фрустраторы, из-за которых сотрудники задумывались об увольнении [4]. Для высокой точности прогнозирования такие интервью проводятся в виде встречи один на один с сотрудником, что занимает очень много времени. Следующий способ прогнозирования ухода сотрудника – это поиск индикаторов в интернете, то есть просмотр профиля каждого сотрудника в таких сервисах как LinkedIn и hh.ru. Однако такой способ требует постоянного мониторинга и не дает точных результатов [4].

Целью данного исследования было построение прогностической модели, предсказывающей перспективы будущего и настоящего сотрудника. Для этого необходимо было собрать множество публичных данных о сотрудниках и факторов, влияющих на их уход из компании, а также создать и обучить нейросетевую модель на этих данных. Конечным результатом должна быть нейросетевая система, способная прогнозировать уход сотрудника из компании в течение двух лет.

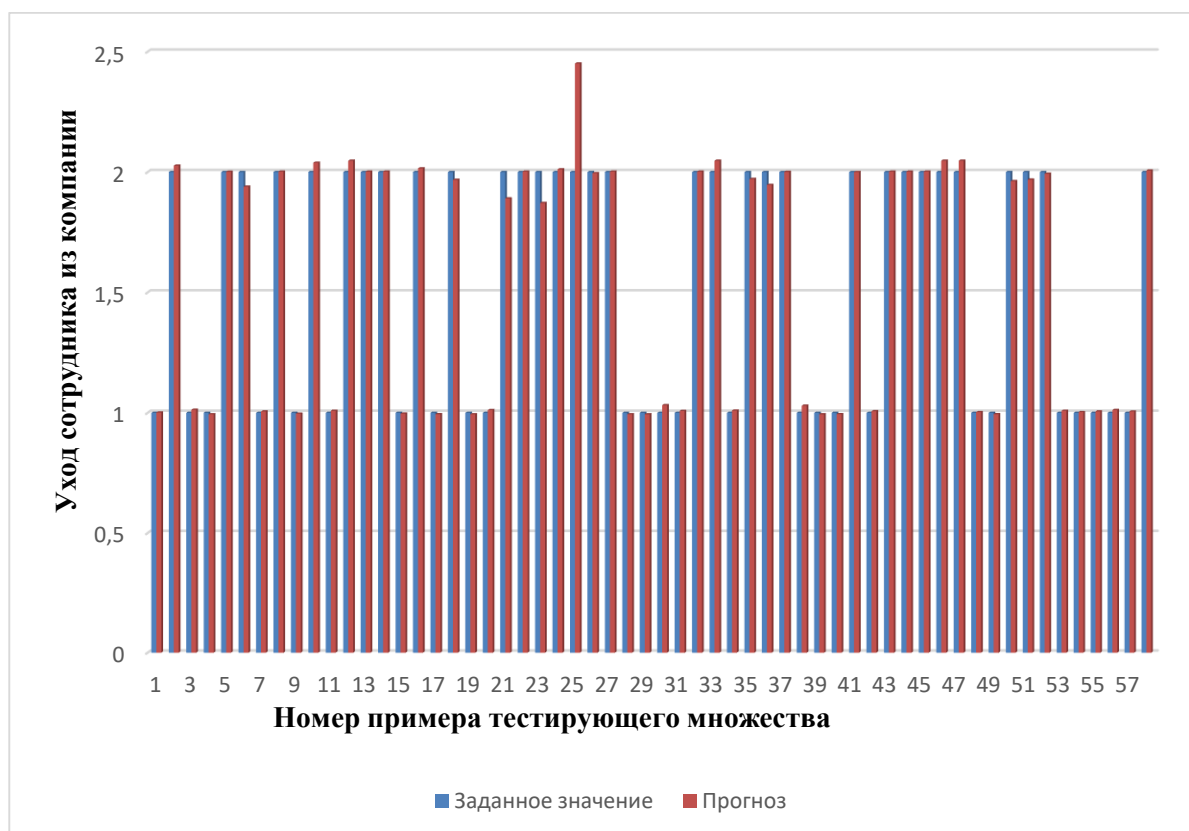
В результате анализа были выделены следующие параметры: X1 – уровень образования (бакалавриат = 1, магистратура = 2, PhD = 3), X2 – год вступления в компанию (год), X3 – уровень заработной платы (высший уровень = 1, средний уровень = 2, самый низкий уровень = 3), X4 – возраст (лет), X5 – пол (мужской = 0, женский = 1), X6 – параметр, отвечающий за участие сотрудника в проектах в течение 1 месяца или более (нет = 0, да = 1), X7 – опыт работы в текущей сфере. Выходной параметр: D1 – Прогноз ухода сотрудника из компании в ближайшие 2 года (не покинет = 0, покинет = 1).

Итоговое множество для нейросети содержит разнообразные данные сотрудников компаний. Перед переходом к проектированию и обучению нейросетевой модели, была выполнена очистка исходного множества от противоречивых примеров, выбросов, дубликатов. Например, некорректными примерами считались те, где сотрудник при высокой заработной плате и только вступивший на должность сразу увольнялся из компании. Таким образом, объем итогового множества включает в себя 885 примеров. Данное множество было разделено на обучающее и тестирующее. Данные были собраны с сайта kaggle [5].

Проектирование, обучение, тестирование нейронной сети выполнялись с помощью программы «Нейросимулятор 5» [6] по методике [7]. Обнаружение выбросов статистической информации в исходном множестве данных выполнялось по методике [9]. В результате оптимизации спроектированная

нейронная сеть представляла собой персептрон, который имеет семь входных нейронов, один выходной и один скрытый слой с тремя нейронами.

Для оценки корректной работы спроектированной нейронной сети использовалось тестирующее множество, состоящее из 57 примеров. Средняя относительная ошибка тестирования составила 9,85%, что можно считать приемлемым результатом. На рисунке 1 представлена гистограмма, демонстрирующая разницу между фактическим и прогнозируемым нейросетью результатом ухода сотрудника из компании.



**Рисунок 1. Результат тестирования нейросетевой системы**

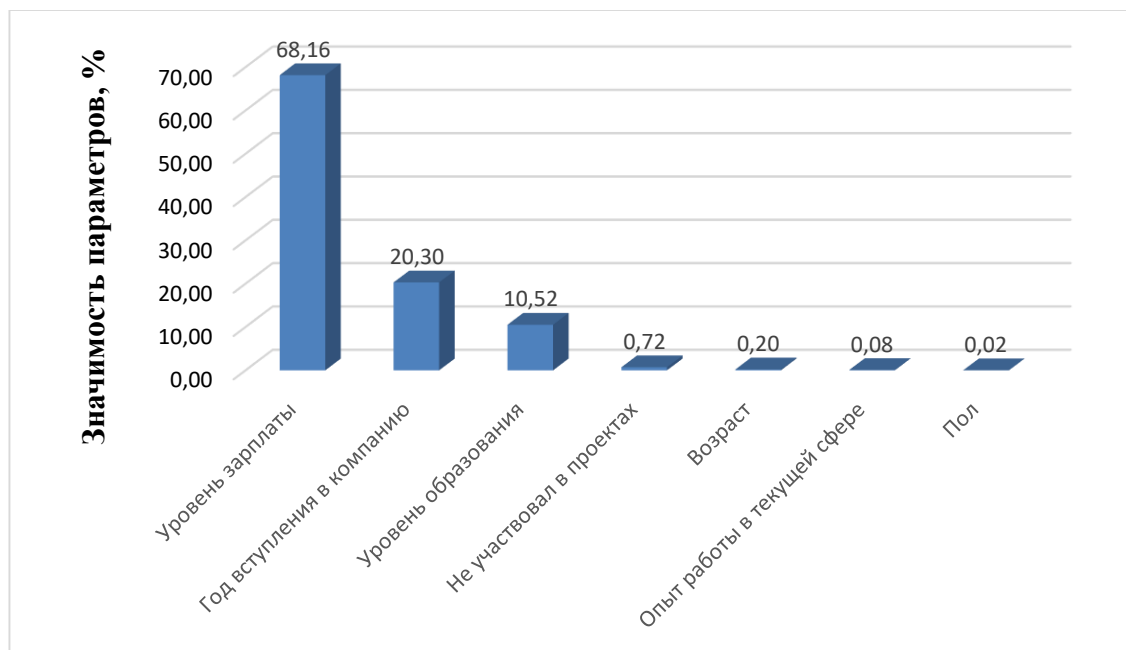
Из результатов, изображенных на рисунке 1, можно сделать вывод об адекватной работе спроектированной нейронной сети.

Оценка значимости входных параметров была выполнена с помощью программы «Нейросимулятор 5». На рисунке 2 представлена гистограмма, показывающая распределение значимости параметров.

Как видно из рисунка 2, самыми значимыми входными параметрами оказались по степени весомости соответственно: уровень зарплаты (уровень значимости – 68%), год вступления в компанию (уровень значимости – 20%), уровень образования (уровень значимости – 10%), неучастие в определенный момент времени в проекте (уровень значимости – 1%). Полностью неэффективные и незначимые параметры: возраст, пол, опыт работы в текущей сфере.

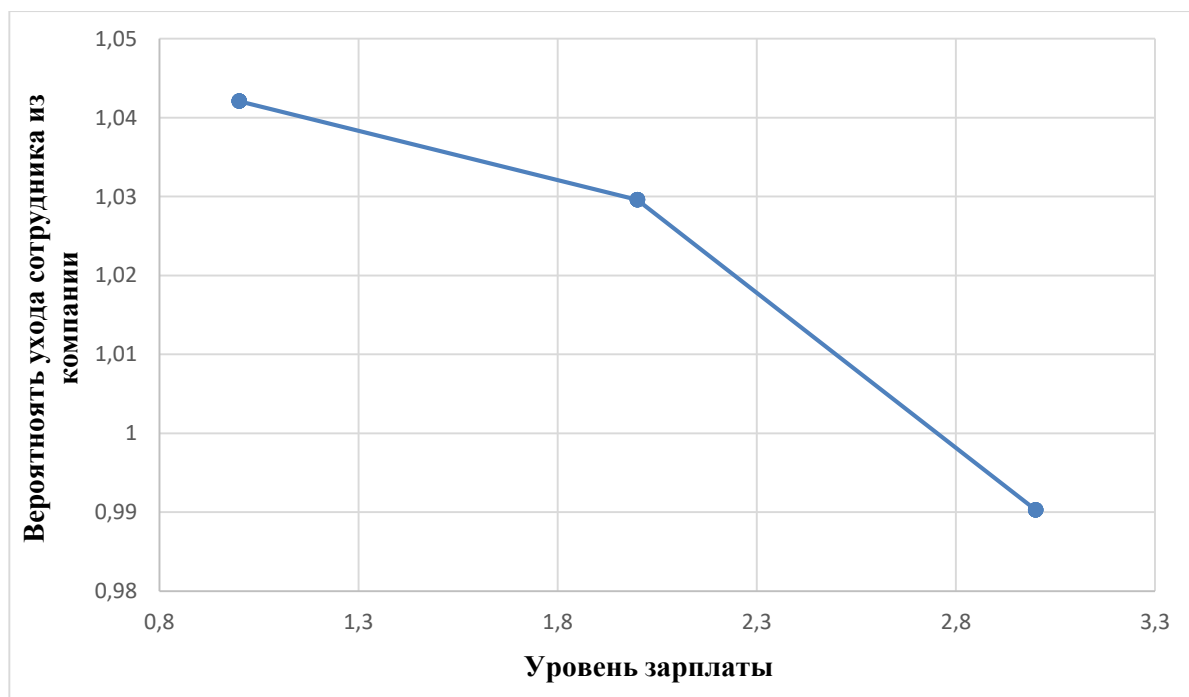
Далее было проведено исследование полученных зависимостей между входными параметрами и вероятностью ухода сотрудника из компании. Исследование производилось с помощью метода «замораживания» [7], суть ко-

торого заключается в варьировании значения одного параметра и фиксации значений всех других параметров. Для этой цели был отобран «нейтральный пример».



**Рисунок 2. Значимость входных параметров нейросетевой модели**

На рисунке 3 показан график зависимости прогнозируемого ухода сотрудника из компании от уровня зарплаты сотрудника. В том случае, когда у сотрудника растет заработная плата его вероятность ухода из компании падает. В исследованиях зачастую именно зарплата является основным стимулом удержания сотрудника в компании [8].



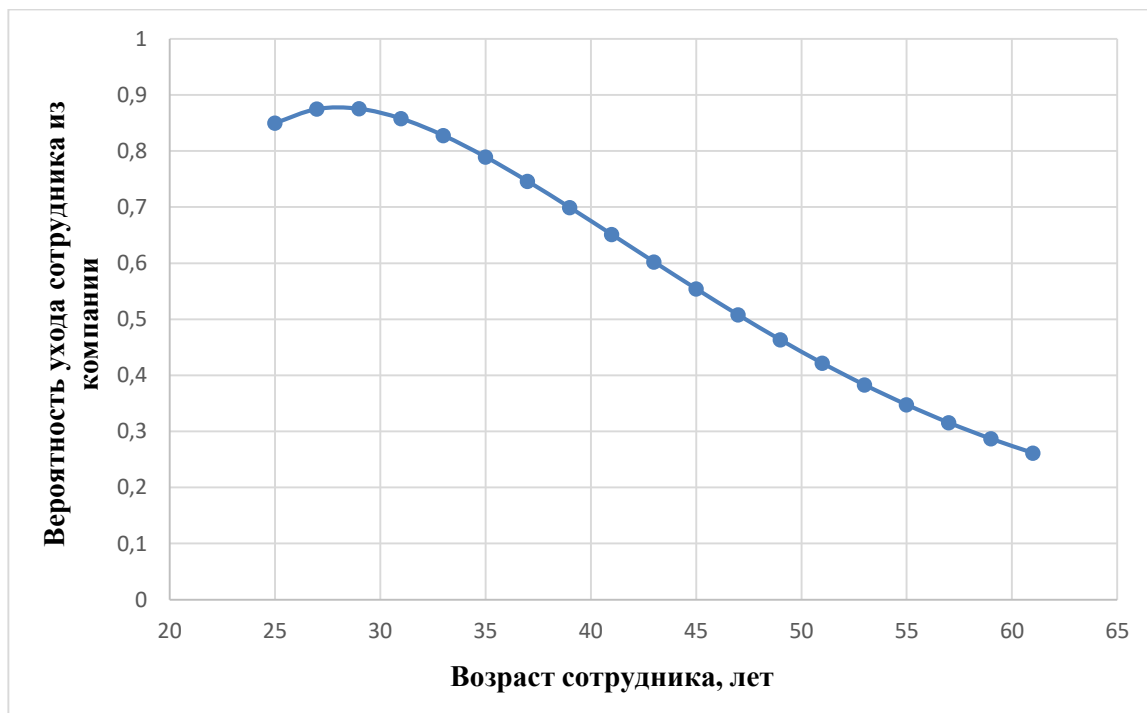
**Рисунок 3. Зависимость ухода сотрудника из компании от уровня зарплаты**

На рисунке 4 продемонстрирована зависимость прогнозируемого ухода сотрудника из компании от уровня образования. Можно заметить, что от уровня образования вероятность ухода сотрудника из компании растет.



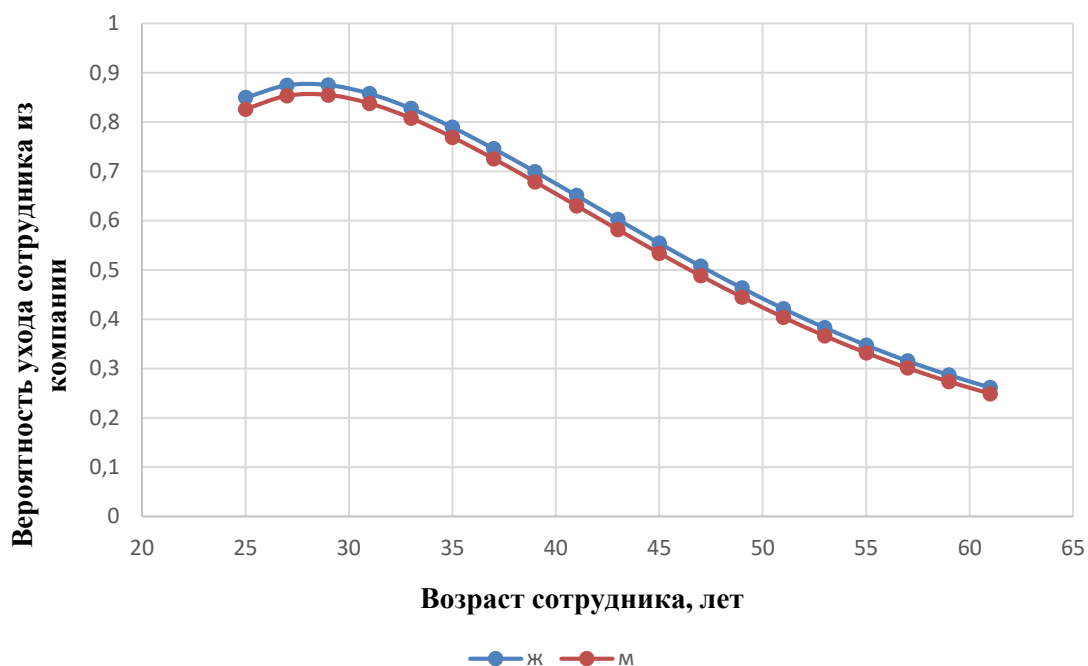
**Рисунок 4. Зависимость ухода сотрудника из компании от уровня образования**

На рисунке 5 изображен график зависимости вероятности ухода сотрудника из компании от возраста сотрудника. Как видно из графика, с возрастом сотрудник все меньше подвержен уходу из компании.



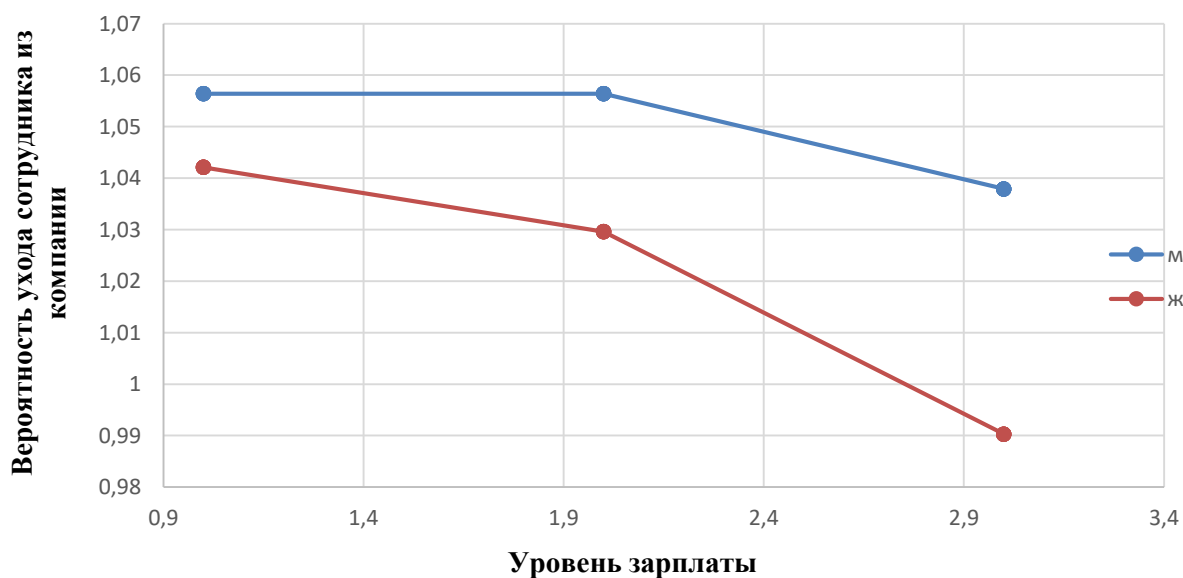
**Рисунок 5. Зависимость ухода сотрудника из компании от возраста сотрудника**

На рисунке 6 продемонстрирована зависимость ухода сотрудника из компании в зависимости от пола и возраста. Как видно из графика, мужчины более постоянны. Однако, наибольшую постоянность демонстрирует рисунок 7.



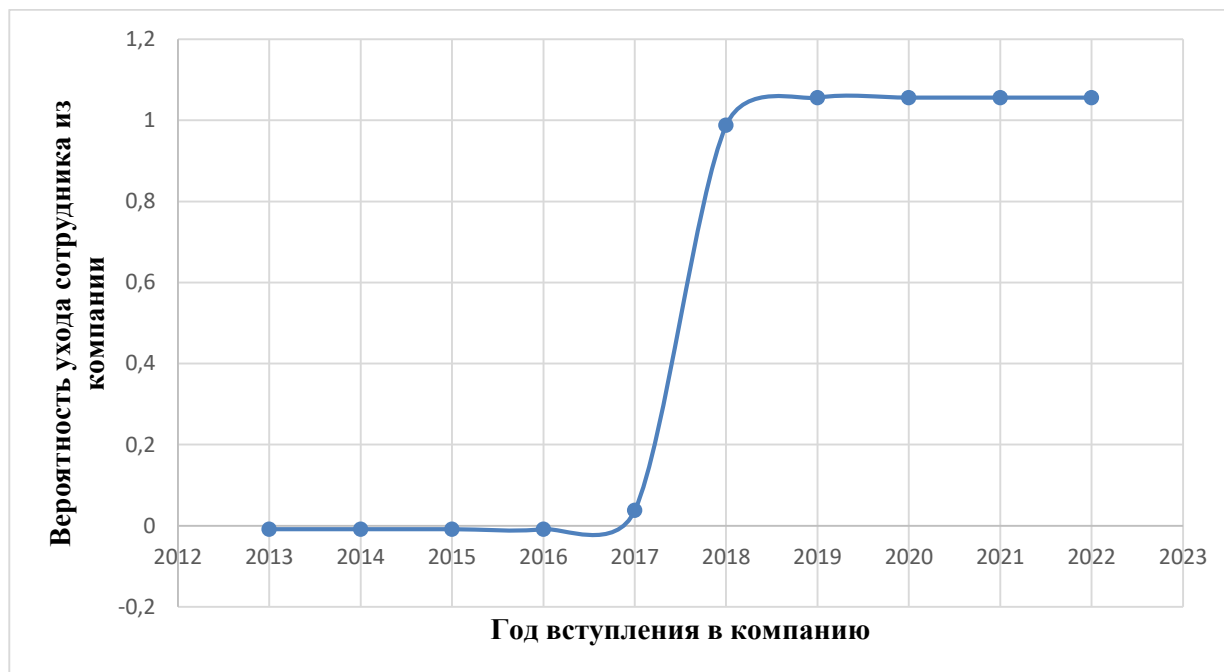
**Рисунок 6. Зависимость ухода сотрудника из компании от пола и возраста**

На рисунке 7 продемонстрирована зависимость ухода сотрудника из компании от пола и уровня зарплаты. Как видно из графика мужчины более постоянны, в то время как женщины больше подвержены переменам. То есть можно сделать вывод, что мужчины более постоянны в выборе при уходе с работы, в то время как женщины имеют тенденцию к сомнениям.



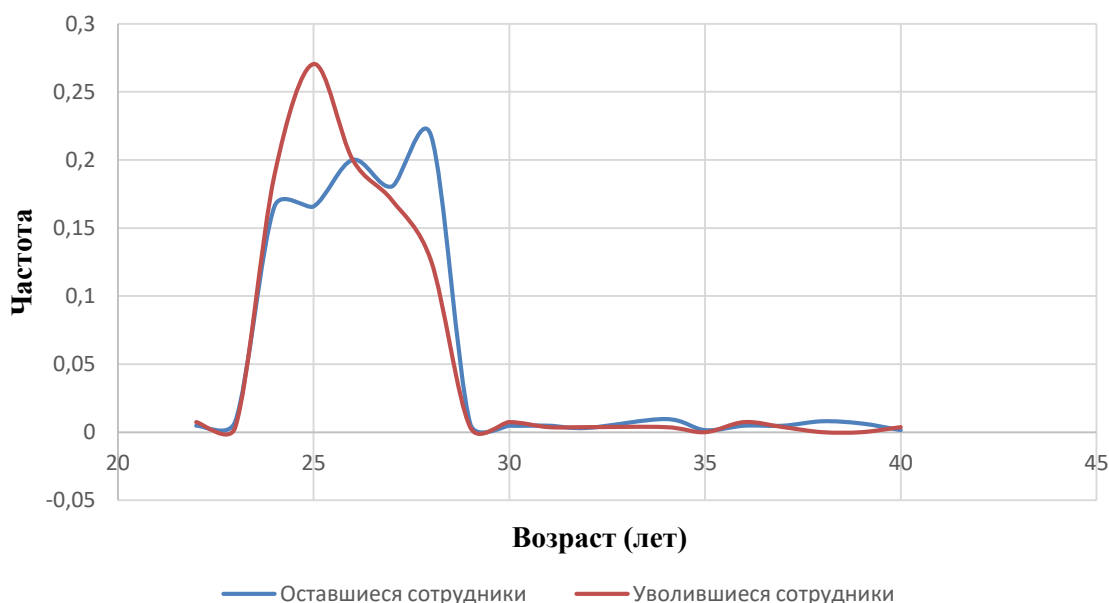
**Рисунок 7. Зависимость ухода сотрудника из компании от пола и уровня зарплаты**

На рисунке 8 продемонстрирована зависимость ухода сотрудника из компании от года вступления в компанию. По диаграмме можно сделать вывод, что раньше люди были менее подвержены уходу из компании. Начиная с 2017 года тенденция к частому уходу из компании среди сотрудников начинает расти.



**Рисунок 2. Зависимость ухода сотрудника из компании от года вступления в компанию**

Распределение по возрасту для работающих и уволившихся сотрудников не отличается. Средний возраст сотрудников, решивших уйти из компании, составляет 26,6 лет, а все еще работающих сотрудников – 26 лет.



**Рисунок 9. Средний возраст сотрудников, оставшихся в компании и сотрудников, уволившихся из компании**

Полученные результаты исследований не противоречат реальности, что подтверждает, что спроектированную нейронную сеть можно считать пригодной для прогнозирования ухода сотрудников из компании в течение двух лет.

Далее построение нейросетевой модели происходило в интерактивном облачном сервисе блокноте Google Colab на языке программирования Python с помощью библиотеки Keras. Были получены аналогичные результаты исследования что и в нейросимуляторе. Однако, прогнозирование на нейросимуляторе имеет преимущества в виде простоты изучения, а также нейросимулятор обладает интуитивно понятным интерфейсом и для построения модели прогноза не требуется знание языка программирования Python.

**Заключение.** Построена система нейросетевого прогнозирования ухода сотрудника из компании в течение двух лет. Спроектированная нейросетевая модель учитывает 7 параметров: уровень образования, год вступления в компанию, уровень заработной платы, возраст сотрудника, пол, параметр, отвечающий за неучастие сотрудника в проектах в течение 1 месяца или более, опыт работы в текущей сфере. Методом сценарного прогнозирования построены графики зависимостей прогнозируемой вероятности ухода сотрудника из компании от изменения входных параметров. Применение такого набора параметров в модели позволяет с высокой точностью прогнозировать уход сотрудника из компании.

### Библиографический список

1. Текучесть кадров: причины и следствие, пути выхода из ситуации [Электронный ресурс]. Режим доступа: <https://hr-portal.ru/article/tekuchest-kadrov-prichiny-i-sledstvie-puti-vyhoda-iz-situacii>.
2. There Are Significant Business Costs to Replacing Employees [Электронный ресурс]. Режим доступа: <https://www.americanprogress.org/wp-content/uploads/2012/11/CostofTurnover.pdf>
3. Построение модели оттока сотрудников для разработки стратегии удержания [Электронный ресурс]. Режим доступа: <https://habr.com/ru/post/582304/>
4. 10 способов, позволяющих определить, что сотрудник собирается уволиться [Электронный ресурс]. Режим доступа: [https://neohr.ru/korporativnaya-kultura/article\\_post/10-sposobov-pozvolyayuschih-opredelit-cto-sotrudnik-sobirae](https://neohr.ru/korporativnaya-kultura/article_post/10-sposobov-pozvolyayuschih-opredelit-cto-sotrudnik-sobirae)
5. Employee future prediction [Электронный ресурс]. Режим доступа: <https://www.kaggle.com/tejashvi14/employee-future-prediction>
6. Черепанов Ф. М., Ясницкий Л.Н. Нейросимулятор 4.0. Свидетельство о регистрации программы для ЭВМ RUS 2014612546. Заявка № 2014610341 от 15.01.2014.
7. Ясницкий Л.Н. Интеллектуальные системы: учебник. М.: Лаборатория знаний, 2016. 221 с.
8. Как удержать сотрудников: пошаговая инструкция [Электронный ресурс]. Режим доступа: <https://dasreda.ru/learn/blog/article/1090>



9. Yasnitsky L.N. Algorithm for searching and analyzing abnormal observations of statistical information based on the Arnold – Kolmogorov – Hecht-Nielsen theorem // International Journal of Advanced Trends in Computer Science and Engineering. 2020. 9(2). Pp. 1814-1819. DOI: 10.30534/ijatcse/2020/139922020 <http://www.warse.org/IJATCSE/static/pdf/file/ijatcse139922020.pdf>

## **NEURAL NETWORK SYSTEM FOR FORECASTING THE EMPLOYEE'S LEAVING THE COMPANY WITHIN TWO YEARS**

*Tsypliakova Elena A.*

National Research University “Higher School of Economics”  
St. Gagarin Boulevard, 37, Perm, Russia, 614060, [eatsyplyakova@edu.hse.ru](mailto:eatsyplyakova@edu.hse.ru)

The article presents a description of the development of a neural network system for predicting the departure of an employee from the company. The system allows predicting whether a future employee will leave the company within two years. A few studies have been carried out, the results of which are presented in this article.

**Keywords:** company, employees, care, forecast, neural network, level of education, year of joining the company, salary level, employee's age, work experience in the current field.

**ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ  
В ЭКОНОМИКЕ, ПРОИЗВОДСТВЕ  
И БИЗНЕСЕ**

## НЕЙРОСЕТЕВАЯ СИСТЕМА ПРОГНОЗИРОВАНИЯ СТОИМОСТИ АВИАБИЛЕТА

*Черемных Александр Александрович*

Национальный исследовательский университет  
«Высшая школа экономики», 614070, Россия, г. Пермь,  
ул. Студенческая, 38, cherema049@yandex.ru

В статье представлено описание разработки нейросетевой системы для прогнозирования цены билетов на авиаперелеты. Система позволяет с большой точностью оценить стоимость авиабилета по таким показателям как количество дней между приобретением и датой вылета, общее время в пути, время и дата вылета, категории авиаперевозчика и другими. С помощью разработанной интеллектуальной системы проведено исследование предметной области, выявлены закономерности, имеющие практической значение.

**Ключевые слова:** искусственный интеллект, нейросетевые технологии, прогнозирование, авиапутешествия, стоимость авиабилета.

**Введение.** Современный мир стремительно развивается. Ежедневно в нашу жизнь внедряются тысячи новых технологий, существенно меняя действующие устои. Благодаря этому отношения между странами и континентами последние десятилетия становятся все ближе, стираются возможные границы и барьеры для сотрудничества. Одним из ключевых факторов тенденции сближения является бурное развитие авиационной индустрии, услуги которой стали основой большинства бизнес-путешествий по всему миру.

Согласно различным исследованиям, проводимым специалистами ВЦИОМ и исследовательскими центрами Росавиации, фиксируется увеличение процента населения страны, пользующегося услугами отечественных и зарубежных авиаперевозчиков. В 2019 году он составил более 35%, в то время как в 2014-м году всего 19% [1]. Также, в 2019 году показатель «количество перевезенных пассажиров в год» достиг рекордной отметки 129 миллионов человек [2].

Приведённые тенденции указывают на увеличение спроса россиян на услуги авиаперевозок и прямое воздействие отрасли на население.

Именно цена на билет определяет эффективность транспортной отрасли в целом по отношению к другим его участникам.

На сегодняшний день изучение факторов, влияющих на процесс ценообразования билетов на авиаперевозки актуально и востребовано. Политика, принимаемая авиакомпаниями для его реализации, должна позволять достигать максимальной эффективности, обеспечивать прибыльность даже в условиях стихийных и непредвиденных ситуаций, происходящих на рынке пассажироперевозок и экономики в целом.

Качественно новым и более интересным подходом к решению данной задачи является прогнозирование самих цен на базе имеющихся данных. Благодаря аналитическому подходу, который поможет корректно смоделировать нейросетевую систему, использованию искусственного интеллекта, возможно добиться успехов в большей части прогнозов.

Результаты работы в перспективе могут быть полезны авиаперевозчикам в процессе формирования стоимости билетов для достижения максимальной эффективности и прибыльности бизнеса. Используются в качестве справочной информации для авиапассажиров, авиапутешественников при поиске и покупке авиабилетов с наилучшими условиями и по самому выгодному предложению.

При анализе научных работ выяснилось, что исследование на тему прогнозирования стоимости авиабилетов с помощью методов машинного обучения, не представляется новым.

В одной из зарубежных работ по прогнозированию стоимости авиабилетов в Греции, К. Циридис и его коллегами использована модель, которая основывается на данных международного рейса по маршруту «Салоники – Штутгарт» национального греческого авиаперевозчика Aegean Airlines в течение нескольких лет [3]. Основными данными выступили следующие описывающие переменные: время вылета, время прилета, количество бесплатного регистрируемого багажа по тарифу, количество дней перед датой вылета и покупной билета, число остановок в пути, день недели и две бинарных (dummy) переменных: выходные (1 – да; 0 – нет) и ночной перелет (1 – да; 0 – нет).

Основываясь на вышеперечисленных параметрах, авторы создали большое количество моделей классификации для прогнозирования стоимости авиабилета, выполнили полноценное сравнение полученных моделей с помощью ML<sup>1</sup>. Целью исследования греческих авторов, является не прогнозирование цены тарифа, а построение (обучение) моделей для прогноза с помощью входных данных, общее количество которых 1814 единиц. Полученные модели использованы ими для составления полезных прогнозов на основе новых данных, полученных из того же распределения, которое используется для подготовки моделей.

По результатам исследования К. Циридису и другим удалось достичь 88% точности прогноза с помощью моделей трех видов: Bagging Regression Tree, Regression Tree и Random Forest Regression Tree. Это исследование позволяет улучшить результаты подобных работ путем объединения большего количества статистических данных по авиаперелетам.

Цель настоящей работы заключается в самостоятельном сборе множества данных по авиабилетам, создание и обучение нейросетевой модели на основе этих данных. В качестве конечного результата ожидается получение нейросетевой системы, способной прогнозировать цену тарифа с наибольшей вероятностью.

Для создания нейросетевой системы выбран следующий набор показателей: X1 – тип авиакомпании, X2 – вылет в вечер пятницы, выходные и утро

---

<sup>1</sup> ML – Machine Learning (перевод. Машинное обучение).

понедельника, X3 – время вылета, X4 – вылет в час-«пик», X5 – время в пути (мин.), X6 – количество остановок-пересадок (ед.), X7 – количество дней между приобретением авиабилета и вылетом (дн.). Выходной параметр D1 – стоимость перелета (нац. валюта).

Часть параметров закодирована: тип авиакомпании (1 – национальный перевозчик; 2 – бюджетный авиаперевозчик, который осуществляет внутренние рейсы; 3 – бюджетный авиаперевозчик, который осуществляет внутренние и международные рейсы; 4 – код-шеринг), вылет в вечер пятницы, выходные и утро понедельника (1 – да; 0 – нет). Отметим, что в качестве «выходных» дней определены суббота и воскресенье, вылеты в «вечер пятницы» – это время вылета в промежутке [17:00-22:00], в «утро понедельника» в течение [04:00-12:00] соответственно. Вылет в часы «пик» (1 – да; 0 – нет).

Основываясь на анализе литературы, часами «пик» принято считать время с 7:00 до 10:00 утра и с 15:00 до 19:00 вечера [4], количество пересадок (0 – без пересадок; 1 – одна пересадка; 2 – две пересадки; 3 – три пересадки; 4 – четыре пересадки).

Данные взяты с интернет-ресурса **Kaggle** [5], а также дополнены вручную. Собранный набор данных содержит информацию о вариантах бронирования авиабилетов с сайта **EaseMyTrip** на внутренних и международных перелетах из/в Индию [6].

Исходная выборка содержала более 500-700 тыс. строк. Для удобства обучения и тестирования случайным образом объем итогового набора данных сокращен до 168 примеров. Данные проверены на наличие пропущенных значений и выбросов с помощью реализованных функций в программе «Нейросимулятор 5» [7] и статистического пакета R [8]. По результатам успешной проверки данных выборка разделена на обучающую и валидирующую, в соотношении 80% и 20% соответственно. Обучающее множество разделено на обучающее и тестирующее, в соотношении 90% и 10% соответственно.

Первый этап работы по проектированию, обучению, тестированию нейронной сети выполнен с помощью программы «Нейросимулятор 5» по методике [13]. В результате оптимизации спроектированная нейронная сеть представляла собой персептрон, имеющий семь входных нейронов, один выходной и один скрытый слой с одним нейроном.

Функция активации на входном и скрытом слоях – тангенс. На выходном слое применяется линейная функция активации. На рисунке 1 отображено графическое представление созданной сети.

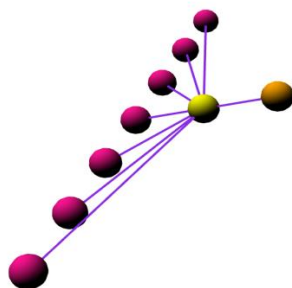
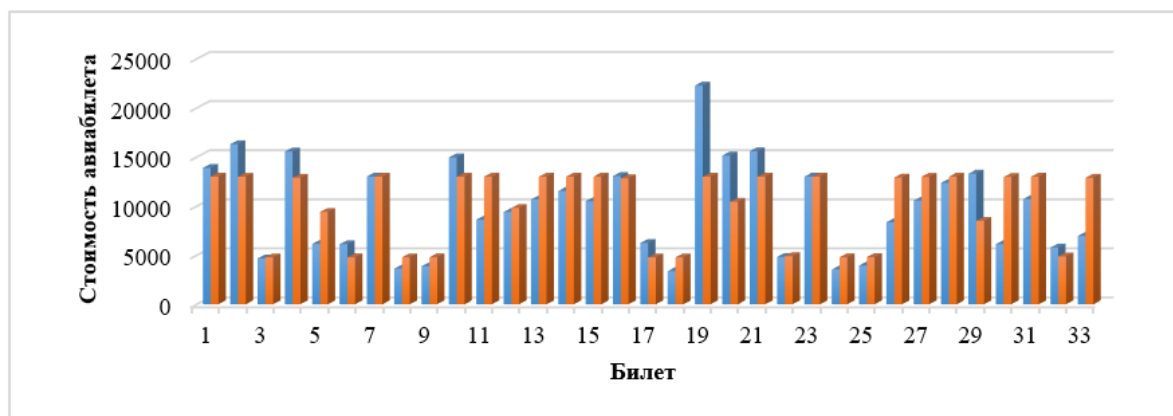


Рисунок 1. Графическое представление сети

Для оценки адекватности и корректной работы спроектированной нейронной сети использовалось тестирующее множество, состоящее из 33 примеров. В итоге, средняя относительная ошибка тестирования составила 15.12%, что вполне можно считать приемлемым результатом моделирования.

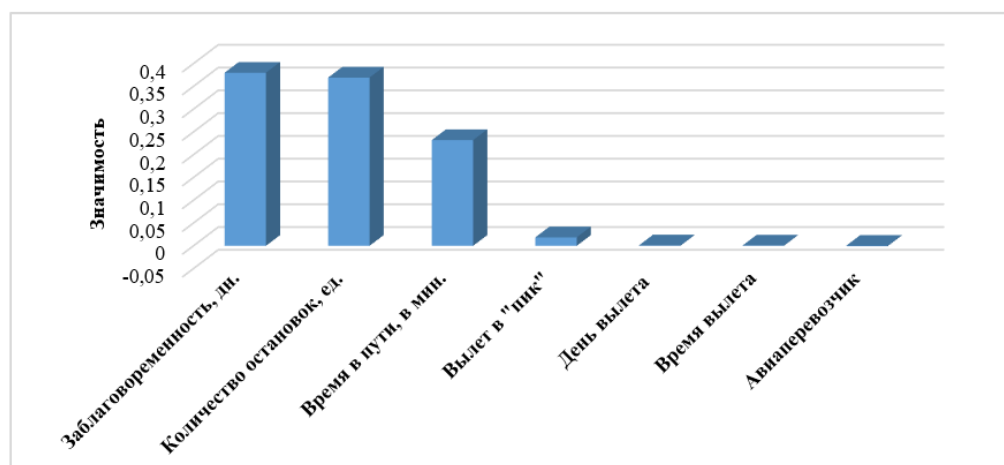
На рисунке 2 представлена гистограмма, демонстрирующая разницу между фактическим и прогнозируемым показателем стоимости авиабилета.



**Рисунок 2. Результат тестирования нейронной сети**

Из результатов, изображенных на рисунке 2, сделан вывод об адекватной работе спроектированной нейронной сети.

Оценка значимости входных параметров выполнена с помощью программы «Нейросимулятор 5», результат отображен на рисунке 3.



**Рисунок 3. Значимость входных параметров нейросетевой модели**

Согласно рисунку 3 наиболее весомыми параметрами модели являются: заблаговременность покупки авиабилета (т.е количество дней между приобретением авиабилета и вылетом), количество пересадок, общее время в пути, вылет в «пик» и т.д. Дадим объяснения полученным результатам.

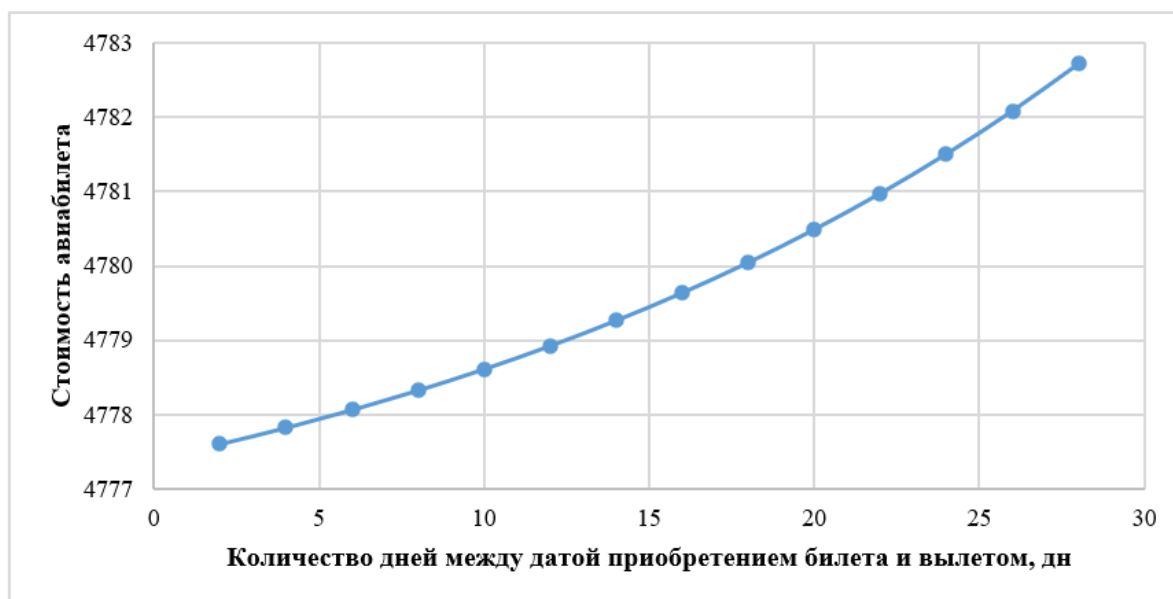
Как и ожидалось, количество дней между датой вылета и покупкой билета на рейс явилось самым статистически значимым фактором.

Д. Эскобари отмечено, что время покупки билета (количество дней до выполнения рейса) имеет корреляцию со стоимостью билета [9]. Он определил, что за семь и четырнадцать дней до дня фактической даты вылета, показатель цены на билеты увеличивается в силу намечающихся рабочих поездок.

Особая степень значимости данного фактора подтверждается анализом стратегий ценообразования на европейском авиарынке. У авиакомпаний British Airlines и KLM между 55 и 21 днями до вылета цена держится на одном уровне. По истечению этого периода тарифы резко растут в цене (М. Альдериги и другие) [10]. Стоимость не будет значительно повышаться (на 50%) при бронировании, сделанном за 20 дней (Р. Малигетти и другие) [11], а также за 18 дней (Д. Менчеро и другие) [12].

Проведено исследование полученных зависимостей между входными параметрами и вероятностью возможной задержки рейса. Исследование произведено методом «замораживания» [13], суть которого заключается в варьировании значения одного параметра и фиксации значений всех других параметров. С помощью данного способа возможно определить влияние рассматриваемого параметра на значение выходного.

Рассмотрены влияние наиболее и относительно значимых параметров на прогнозируемый показатель. На рисунке 4 показан график зависимости стоимости авиабилета от заблаговременности его покупки.



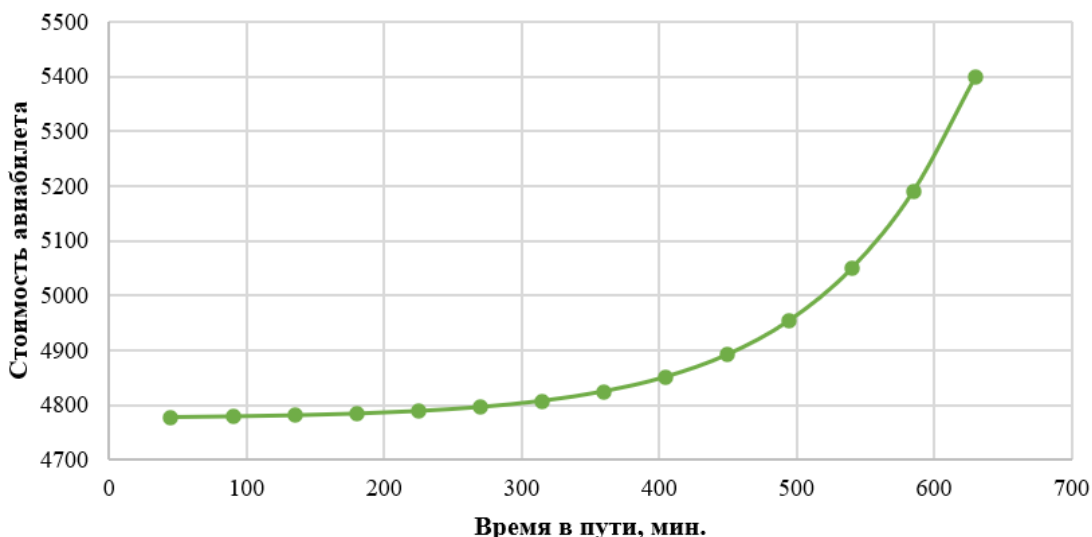
**Рисунок 4. Зависимость стоимости билета от заблаговременности его покупки**

Из графика следует, что между данными параметрами присутствует взаимосвязь. Соответственно можно сделать вывод о том, что при заблаговременной покупке авиабилета, его стоимость будет высока. Так, при приближении к дате вылета – цена будет наоборот снижаться, в особенности, за 4-8 дней. Обращаясь к научным источникам, ряд авторов определило причину возникновения данной тенденции в процессе изучения влияния рассматриваемого параметра на стоимость тарифа.

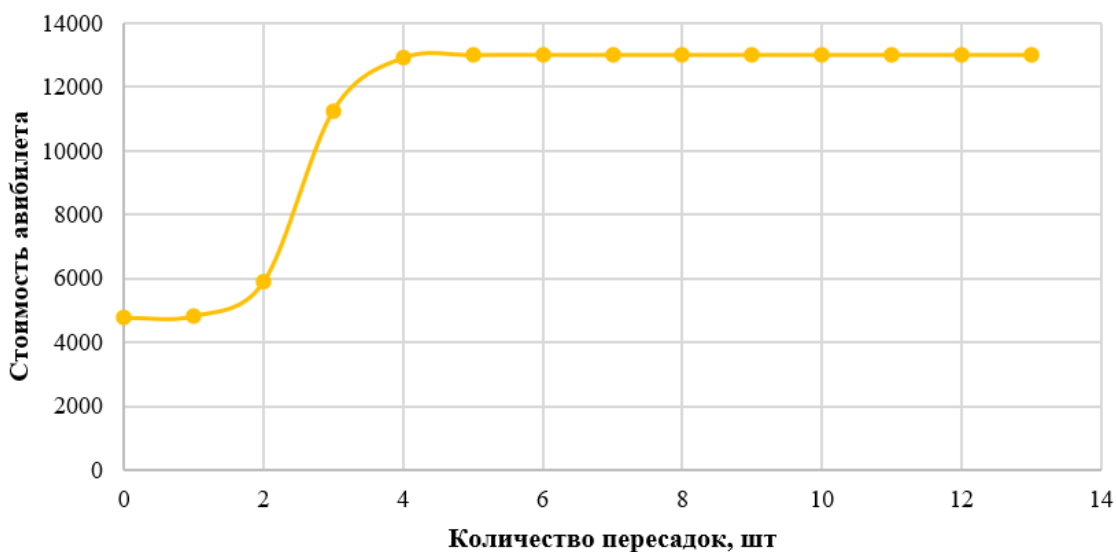
Например, Д. Эскобари утверждает, что стоимость тарифов в том числе снижается по мере уменьшения количества мест на рейс [9]. Правильность вывода автора подтверждается в работах У. Кинкейда и Д. Дарлинг [14], Г. Гальего и Г. ван Ризин [15], объясняющих тенденцию падения цен на тарифы по мере приближения к дате вылета.

П. Малигетти, С. Палеари и Р. Редонди проанализирована ценовая политика лоукостера Ryanair [11]. В ходе исследования выделена важность следующих характеристик: процент забронированных мест на рейсе. Факт роста цены за счет уменьшения количества доступных мест на рейс объясняется желанием авиаперевозчика заполнить все места в самолете как можно быстрее. Этим они стимулируют клиента купить тариф [16].

На рисунке 5 продемонстрирована зависимость стоимости авиабилета от общего времени в пути. С увеличением времени перелета на каждые 45 минут, цена тарифа постепенно увеличивается. Следовательно, продолжительность полета аналогично имеет тесную взаимосвязь с ценой [17].



**Рисунок 5. Зависимость цены билета от времени в пути**



**Рисунок 6. Зависимость стоимости билета от количества пересадок**



На рисунке 6 продемонстрирована зависимость стоимости авиатарифа на рейсы с пересадками от их количества. Из графика видно, что чем больше пересадок, тем больше итоговая стоимость билета.

Согласно графику определено, что одна пересадка снижает расходы на авиабилеты, а 2-4 — нет. Если пересадок более 4, то цена лежит на одном уровне и принимает максимальное значение.

В качестве эксперимента были применены методы машинного обучения – язык программирования Python и библиотека Keras для повторения действий, выполненные в программе «Нейросимулятор 5» [7].

В результате удалось спроектировать нейронную сеть с параметрами, ранее указанными в данной статье, но имеющая в качестве функции активации **relu**, в отличие от первого этапа работы, где был применен **tanh** (тангенс) по причине возникновения трудностей в ее обучении. Главными отличиями результатов, полученных с помощью обоих программных инструментов, было выделено следующее:

1. Вычисленное значение  $R^2$  в первом случае было получено 0.54, а во втором 0.36.

2. Графики зависимостей от ряда параметров, которые имеют эффект на рассматриваемый показатель стоимости, в программе Нейросимуляторе более точно отображают тенденцию, в отличие от примененного шаблона Keras, где графики линейны.

Полученные результаты исследования не противоречат реальности. Спроектированную нейронную сеть можно считать пригодной для прогнозирования стоимости билетов на авиаперелеты.

**Заключение.** Построена система нейросетевого прогнозирования стоимости авиабилета в программе «Нейросимулятор 5». Спроектированная нейросетевая модель учитывает 7 параметров: тип авиакомпании; вылет в вечер пятницы, выходные и утро понедельника; время вылета; вылет в часы «пик»; время в пути; количество пересадок; количество дней между приобретением авиабилета и вылетом.

Методом сценарного прогнозирования построены графики зависимостей прогнозируемого показателя цены авиабилета от изменения входных параметров. Применение данного набора параметров в модели позволяет с высокой точностью прогнозировать стоимость авиаперелета.

### Библиографический список

1. Россияне стали активнее пользоваться авиатранспортом, выяснили социологи [Электронный ресурс] / РИА Новости: ежедн. интернет-изд. – URL: <https://ria.ru/20180207/1514130948.html> (дата обращения: 21.03.2022).
2. Авиация бьет рекорды [Электронный ресурс] / Коммерсантъ: ежедн. интернет-изд. – URL: [https://www.kommersant.ru/doc/4220601?utm\\_source=uxnews&utm\\_medium=desktop](https://www.kommersant.ru/doc/4220601?utm_source=uxnews&utm_medium=desktop) (дата обращения: 21.03.2022).
3. Airfare Prices Prediction Using Machine Learning Techniques / Tziridis K., Kalampokas Th., & Papakostas, G.A. // European Signal Processing Conference, August. – Kos, 2017. DOI:10.23919/EUSIPCO.2017.8081365.

4. Comparing Price Dispersion on and off the Internet Using Airline Transaction Data / Sengupta A., & Wiggins, S. N. // *Economic Policy*. 2014. – Vol. 6, Iss 1. – P. 272– 307. DOI: 10.1515/1446-9022.1244.
5. Flight Price Prediction. – URL: <https://www.kaggle.com/datasets/shubhambathwal/flight-price-prediction> (дата обращения: 21.03.2022).
6. EaseMyTrip: сайт. – URL: <https://www.easemytrip.com> (дата обращения: 21.03.2022).
7. Нейросимулятор 5.0: свидетельство о государственной регистрации программы для № 2014618208 от 12.07.2014 г. / Черепанов Ф.М., Ясницкий В.Л. (РФ).
8. Обнаружение статистических выбросов в R [Электронный ресурс] / LEFTJOIN. – URL: <https://leftjoin.ru/all/outliers-detection-in-r/>? (дата обращения: 21.03.2022).
9. Dynamic pricing, advance sales and aggregate demand learning in airlines / Escobari D. // *The Journal of Industrial Economics*. 2012. – Vol. 4, Iss. 60. – P. 697-724. DOI: 10.1111/joie.2012.60.issue-4.
10. A case study of pricing strategies in European airline markets: The London Amsterdam route / Alderighi M., Cento A., & Piga, C.A. // *Journal of Air Transport Management*. 2011. – Vol. 17, Iss. 6. – P. 369-373. DOI: 10.1016/j.jairtraman.2011.02.009.
11. Pricing strategies of low-cost airlines: The Ryanair case study / Malighetti P., Paleari S., Redondi R. // *Journal of Air Transport Management*. 2009. – Vol. 15, Iss. 4. – P. 195-203. DOI: 10.1016/j.jairtraman.2008.09.017.
12. Optimal purchase timing in the airline market / Domínguez-Menchero J.S., Rivera J., Torres-Manzanera E. // *Journal of Air Transport Management*. 2014. – Vol. 40. – P. 137-143. DOI: 10.1016/j.jairtraman.2014.06.010.
13. Ясницкий Л.Н. Введение в искусственный интеллект. М.: Издательский центр «Академия», 2005. – 176 с.
14. An inventory pricing problem / Kincaid W.M., & Darling D.A. // *Journal of Mathematical Analysis and Applications*. 1963. – Vol. 7, Iss. 2. – P. 183-208.
15. Optimal dynamic pricing of inventories with stochastic demand over Finite horizons / Gallego G., & van Ryzin G.J. // *Management Science*. 1994. – Vol. 40, Iss. 8. – P. 999-1020. DOI: 10.1287/mnsc.40.8.999.
16. Efficiency of the natural rate / Prescott E. // *Journal of Political Economy*. 1975. – Vol. 83, Iss. 6. – P. 1229-1236. DOI: 10.1086/260391.
17. International arrivals to Australia: Determinants and the role of air transport policy / Zhang Y. // *Journal of Air Transport Management*. 2015. – Vol. 44. – P. 21-24. DOI: 10.1016/j.jairtraman.2015.02.004.

# NEURAL NETWORK SYSTEM FOR PREDICTING OF AIR FLIGHT CANCELLATION ON THE BASE OF WEATHER AND HISTORY DATA

*Cheremnykh Aleksandr A.*

National Research University “Higher School of Economics”  
Studentskaya St., 38, Perm, Russia, 614070, cherema049@yandex.ru

The article describes the development of a neural network system for predicting the price of air travel. The system allows to accurately estimate the cost of an air ticket based on such indicators as the number of days between purchase and departure date, total travel time, departure time and date, air carrier category and others. With the help of the developed intellectual system, a study of the subject area was carried out, patterns of practical importance were identified.

**Keywords:** artificial intelligence, neural network, prediction, airflights, airfare price.

## ПРОГНОЗИРОВАНИЕ ТЕХНИЧЕСКОГО СОСТОЯНИЯ АВИАЦИОННЫХ ДВИГАТЕЛЕЙ ПРИ ПОМОЩИ НЕЙРОННЫХ СЕТЕЙ

*Кононов Евгений Алексеевич, Ключев Андрей Владимирович*

Пермский национальный исследовательский политехнический университет,  
614990, Россия, г. Пермь, Комсомольский пр-т, 29,  
kononov@pstu.ru, kav-1@bk.ru

В работе представлено описание разработки механизма принятия рациональных решений по обслуживанию авиационных двигателей на основе многомерных временных рядов. Представлен подход по интерпретации обученной модели нейронной сети.

**Ключевые слова:** искусственный интеллект, нейронные сети, прогнозирование, predictive maintenance

**Введение.** В настоящий момент в связи с развитием технологий поступает огромное количество данных с датчиков различных систем. Для того, чтобы они приносили пользу, необходимо извлекать из них знания, которые способствуют принятию решения в будущем или помогут решить проблемы, которые есть прямо сейчас, а ответа на них нет.

В данной работе рассматриваются данные модели авиационного двигателя с наложенной на него моделью повреждения.

Для обслуживания имеющихся систем необходим подход, позволяющий организациям повысить их производительность, эффективность и безопасность. Существует несколько подходов по обслуживанию активов:

- Reactive maintenance (RM, реактивное обслуживание) – метод обслуживания, при котором детали заменяются по мере их выхода из строя. Он обеспечивает полное использование деталей, но требует от бизнеса простоев, и, следовательно, имеет тенденцию вызывать серьезные задержки и приводит к высоким затратам на внеплановый ремонт.

- Preventive maintenance (PM, превентивное обслуживание) – метод обслуживания, при котором предприятия заранее определяют полезный срок службы детали и поддерживают или заменяют ее до выхода из строя. Превентивное обслуживание позволяет избежать внеплановых отказов, но влечет за собой затраты на запланированные простои и недоиспользование компонента.

- Predictive maintenance (PdM, предиктивное обслуживание) – метод обслуживания, целью которого является оптимизация баланса между RM и PM за счет своевременной замены компонентов. При таком подходе активы обслуживаются только тогда, когда они близки к отказу, и, соответственно, увеличивается срок службы компонентов (по сравнению с PM) и сокращается внеплановое обслуживание и затраты на рабочую силу (по сравнению с RM). В конечном счете, решение задач PdM приводит к оптимизации периодического технического обслуживания, минимизации простоев и как следствие – предотвращение материального и физического ущерба.

Концепция PdM существует уже много лет, но только недавно появляющиеся технологии становятся достаточно эффективными и недорогими, чтобы сделать PdM широкодоступным [1]. В связи с этим и с тем, какие возможности предоставляет PdM по сравнению с RM и PM, возникает актуальность выбранной темы.

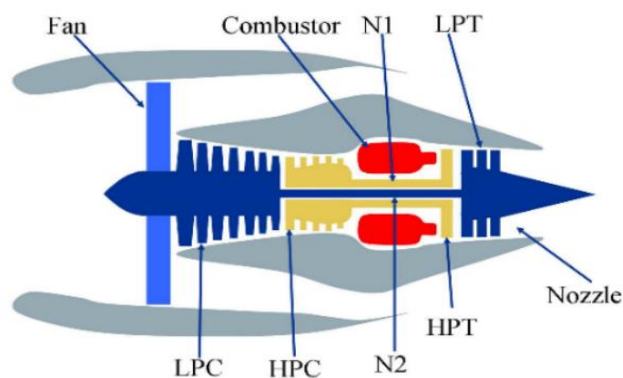
Цель работы: разработать механизм принятия рациональных решений по обслуживанию механических систем на основе интеллектуального анализа временных рядов показаний измерительных устройств.

Для решения задачи оценивания технического состояния авиационных двигателей предлагается два способа:

- 1) Прогнозирование оставшегося полезного срока службы (RUL);
- 2) Бинарная классификация выхода из строя двигателей на заданном горизонте прогноза.

Упрощенная схема двигателя представлена на рисунке 1. Его основные компоненты: вентилятор (Fan), компрессор низкого давления (LPC), компрессор высокого давления (HPC), камера сгорания (Combustor), турбина высокого давления (HPT), турбина низкого давления (LPT), сопло (Nozzle).

Исходными данными являются значения измеряемых признаков модели симуляции повреждений ста авиационных двигателей (рисунок 2). Обучающие данные имеют 20631 образец, а тестовые 13096.



**Рисунок 1. Упрощенная схема двигателя [2]**

UnitNumber	Cycle	Op_Setting_1	Op_Setting_2	Op_Setting_3	Sensor_1	Sensor_2	Sensor_3	Sensor_4	Sensor_5	...
0	1	1	-0.0007	-0.0004	100.0	518.67	641.82	1589.70	1400.60	14.62 ...
1	1	2	0.0019	-0.0003	100.0	518.67	642.15	1591.82	1403.14	14.62 ...
2	1	3	-0.0043	0.0003	100.0	518.67	642.35	1587.99	1404.20	14.62 ...
3	1	4	0.0007	0.0000	100.0	518.67	642.35	1582.79	1401.87	14.62 ...
4	1	5	-0.0019	-0.0002	100.0	518.67	642.37	1582.85	1406.22	14.62 ...
...	...	...	...	...	...	...	...	...	...	...
20626	100	196	-0.0004	-0.0003	100.0	518.67	643.49	1597.98	1428.63	14.62 ...
20627	100	197	-0.0016	-0.0005	100.0	518.67	643.54	1604.50	1433.58	14.62 ...
20628	100	198	0.0004	0.0000	100.0	518.67	643.42	1602.46	1428.18	14.62 ...
20629	100	199	-0.0011	0.0003	100.0	518.67	643.23	1605.26	1426.53	14.62 ...
20630	100	200	-0.0032	-0.0005	100.0	518.67	643.85	1600.38	1432.14	14.62 ...

**Рисунок 2. Пример данных симуляции повреждений авиационного двигателя**

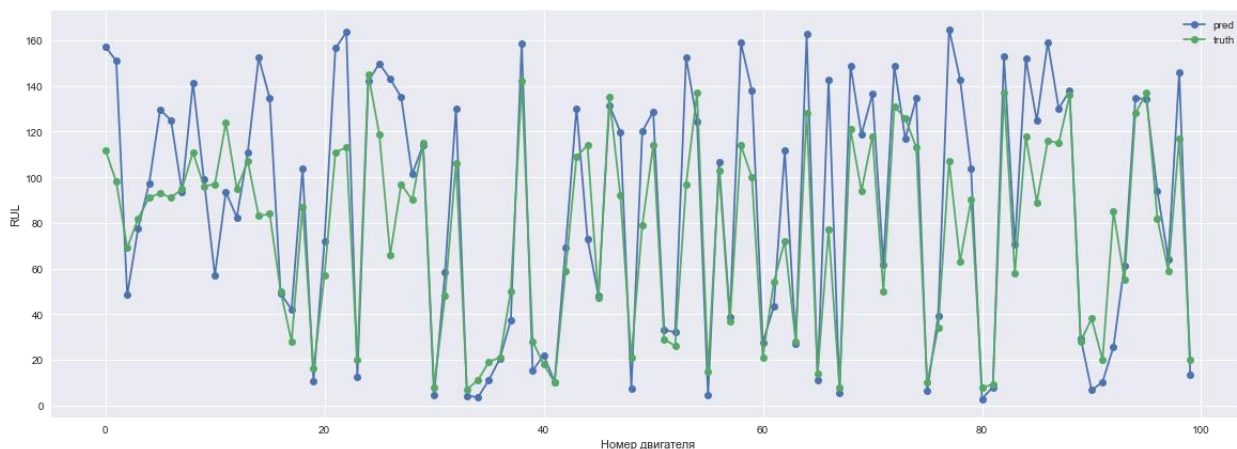
На рисунке 3 видно, что модель для решения задачи прогнозирования RUL обучилась высокой прогнозирующей способностью. Чтобы как-то интерпретировать полученный результат, удобно воспользоваться средней абсолютной ошибкой (1):

$$MAE = \frac{1}{n} \sum_{i=1}^n |Y_i - \hat{Y}_i|, \quad (1)$$

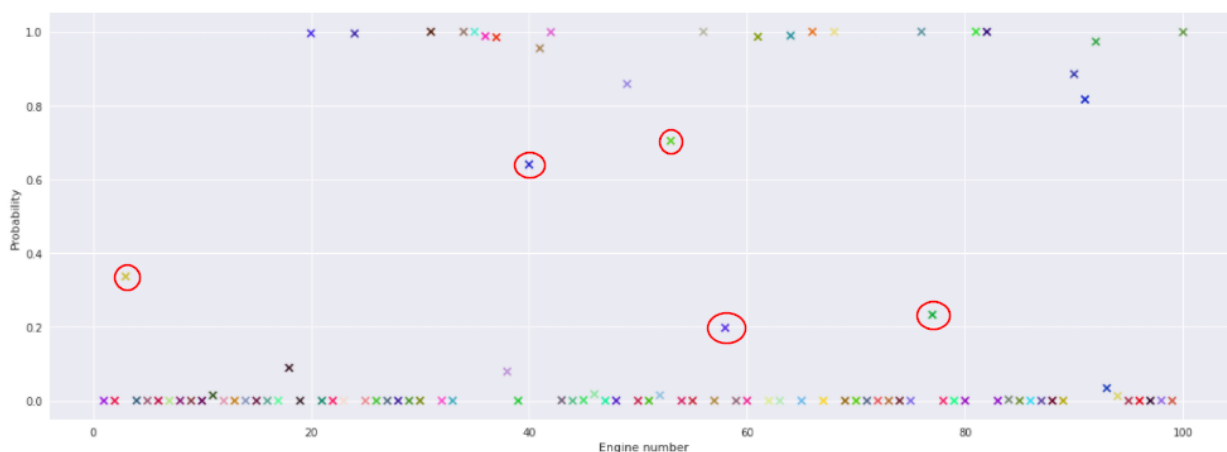
где  $Y_i$  – истинные значения,  $\hat{Y}_i$  – предсказанные значения,  $n$  – общее количество наблюдений. В нашем случае MAE составляет около 20.035, т.е. в среднем модель ошибается на 20 циклов. Важно отметить, что большую часть этой ошибки составляет около 10% двигателей, предсказания RUL которых сильно выделяются на фоне общей картины. То есть можно сделать вывод, что для 90 из 100 двигателей получены удовлетворительные прогнозы по их оставшемуся полезному сроку службы.

Применяя обученную модель к тестовым данным, получаем вероятности поломки двигателей (рисунок 4), как решение задачи классификации состояния двигателей. Красным цветом обведены те значения двигателей (можно обвести и некоторые другие), которые нельзя с хорошей уверенно-

стью отнести к какому-либо классу. В связи с этим, чтобы обеспечить четкое разделение на классы, было найдено оптимальное пороговое значения вероятности (threshold), ниже которого двигатель будет отнесен к классу «0», а иначе – «1».



**Рисунок 3. Спрогнозированные и истинные значения RUL каждого двигателя из тестовой выборки**



**Рисунок 4. Вероятности поломки двигателей в следующие 30 циклов**

Очень важным этапом при использовании нейронных сетей или алгоритмов машинного обучения является их интерпретация, поскольку понимание того, почему модель делает определенный прогноз, может быть столь же важным, как и точность модели. К сожалению, чем сложнее модель и используемая на этапе обучения форма данных, тем хуже ее интерпретируемость. Для объяснения полученных вероятностей поломки авиационных двигателей использовался метод Deep SHAP [3], который оценивает влияние каждого из имеющихся признаков на прогноз модели.

**Заключение.** Для решения задачи оценки технического состояния механических систем предложено и успешно реализовано на практике две постановки: прогнозирование оставшегося полезного срока службы и бинарная классификация выхода из строя. Полученные результаты указывают на достаточно высокий уровень оценки технического состояния предложенными

методами, а их интерпретация позволила выявить ключевые факторы, влияющие на поломку.

Таким образом, задачи решены в полном объеме, цель достигнута – на основе интеллектуального анализа временных рядов показаний измерительных устройств разработан механизм, позволяющий предоставить помощь в принятии рациональных решений по обслуживанию механических систем, а его возможности показаны на практике.

### **Библиографический список**

1. Ran Y. и др. A Survey of Predictive Maintenance: Systems, Purposes and Approaches // IEEE COMMUNICATIONS SURVEYS & TUTORIALS. 2019. Т. XX.

2. Li, Jialin. A directed acyclic graph network combined with CNN and LSTM for remaining useful life prediction / Jialin Li, Xueyi Li, David He // IEEE Access. – 2019. – Vol.7. – P. 75464 – 75475.

3. A Unified Approach to Interpreting Model Predictions: 31st Conference on Neural Information Processing Systems. Long Beach, CA, USA, 2017. / Scott M. Lundberg, Su-In Lee. – 11 P.

## **FORECASTING THE TECHNICAL CONDITION OF AIRCRAFT ENGINES USING NEURAL NETWORKS**

*Kononov Evgeniy A., Klyuev Andrey V.*

Perm National Research Polytechnic University

Komsomolsky Ave, 29, Perm, Russia, 614990, kononov@pstu.ru, kav-1@bk.ru

The paper describes the development of rational decision-making mechanism for aircraft engine maintenance on the basis of multidimensional time series. An approach to interpret the trained model of the neural network is shown.

**Keywords:** artificial intelligence, neural networks, forecasting, predictive maintenance.

УДК 004.661

## **НЕЙРОСЕТЕВАЯ СИСТЕМА ПРОГНОЗИРОВАНИЯ КЛАССА ВЯЗКОСТИ МОТОРНОГО АВТОМОБИЛЬНОГО МАСЛА**

*Гордеев Максим Витальевич*

Пермский государственный национальный исследовательский университет,  
ФИТ. 614990, Россия, г. Пермь, ул. Букирева, 15, max9991995@mail.ru

В работе представлено описание разработки нейросетевой системы для прогнозирования класса вязкости моторного автомобильного масла.

Система позволяет с большой точностью предсказать класс вязкости масла на основании данных его химического анализа.

**Ключевые слова:** искусственный интеллект, нейросетевые технологии, прогнозирование, моторное масло, автомасла, вязкость.

**Введение.** В настоящее время рынок моторных автомобильных масел очень вырос, по сравнению с 2000-ми. На рынке представлено много брендов с различными линейками масел. Вместе с выросшим рынком начали расти и требования к маслам, моторы стали современнее, экологичнее и, как следствие, стали более требовательны к качеству смазывающих жидкости. Нормы токсичности так же наложили свой отпечаток на классификацию масел. И теперь каждый автомобилист, любящий и следящий за своей машиной, задаётся вопросом какое масло заливать?

API, ACEA, ILSAC — это международные стандарты качества, разработанные крупными экспертными организациями. Производители автомобилей всегда указывают определенные спецификации в руководстве по эксплуатации автомобиля, поэтому они являются важным параметром при выборе моторного масла. Рекомендуемые спецификации зависят от конструктивных особенностей и типа двигателя конкретного автомобиля.

Кто же эти экспертные сообщества, занимающиеся разработкой стандартов:

- **API** (от англ. – American Petrol Institute) – Американский институт нефти;
- **ACEA** (от фр. – Association des Constructeurs Européens d'Automobiles) – Ассоциация европейских производителей автомобилей;
- **ILSAC** (от англ. – International Lubricant Standardization and Approval Committee) – Международный комитет по стандартизации и одобрению смазочных материалов;

В данной работе будет рассмотрено две указанные группы стандартов.

Классификация ACEA выделяет масла, которые ориентированы на требования европейских автопроизводителей и имеют развернутую систему одобрений. Ее требования более жесткие исходя из общеевропейских стандартов качества.

Из классификации нас интересуют лишь 2 спецификации, так как они являются наиболее актуальными на данный момент.

Спецификация **A3/B4** включает в себя спецификацию A3/B3, а также масла для бензинового двигателя с прямым впрыском и дизельного двигателя с системой инжекции. Масла этой спецификации подходят для увеличенного интервала замены масла и соответствуют повышенным требованиям автопроизводителей.

Спецификации **A5/B5** соответствуют энергоэффективные моторные масла низкой вязкости, ориентированные на экономию топлива. Они применимы только для двигателей определенных моделей.

Классификация ILSAC создана для масел, используемых в американских и японских автомобилях. Она имеет пять категорий качества, первая из



которых уже устарела, а наиболее актуальные в настоящее время ILSAC GF-4 и ILSAC GF-5.

Мы же будем рассматривать самую новую категорию **ILSAC GF-5** – стандарт соответствует уровню требований API SN.

Так же, важно отметить, я буду прогнозировать самую популярную степень вязкости масел **5W-30** [1].

Основная *цель* настоящей работы заключается в сборе множества публичных данных об анализах масел, а также создание и обучение нейросетевой модели на этих данных. Конечным результатом должна быть нейросетевая система, способная прогнозировать по химическому составу масла принадлежность его к тому или иному стандарту допуска, больше, чем 80% случаев.

Для создания нейросетевой системы были выбраны следующие параметры см. Табл. 1, Табл. 2:

**Таблица 1**

**Основные параметры**

Описание параметра	Параметр
Вязкость кинематическая при 40 °С	X1
Вязкость кинематическая при 100 °С	X2
Индекс вязкости	X3
Щелочное число	X4
Кислотное число	X5
Зола сульфатная	X6
Температура застывания	X7
Температура вспышки	X8
Массовая доля серы	X9

**Таблица 2**

**Дополнительные параметры**

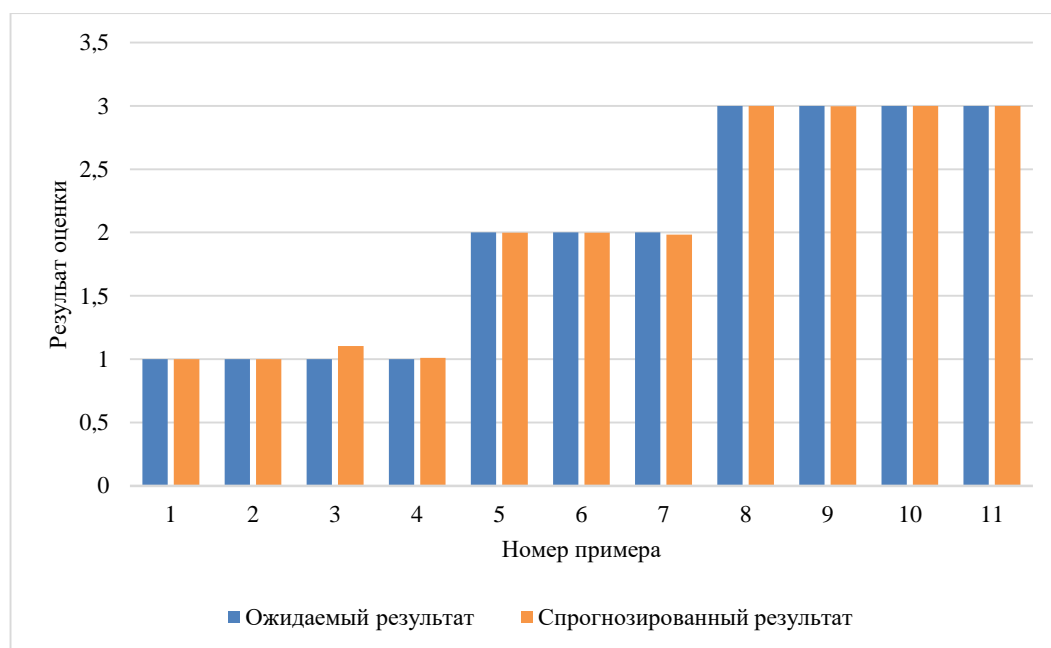
	Описание параметра	Параметр
<i>Массовая доля элементов</i>	Молибден (Mo)	X10
	Фосфор (P)	X11
	Цинк (Zn)	X12
	Бор (B)	X13
	Магний (Mg)	X14
	Кальций (Ca)	X15
	Алюминий (Al)	X16
	Железо (Fe)	X17
	Кремний (Si)	X18
	Натрий (Na)	X19
	Калий (K)	X20

Обучающее множество было собрано вручную. Перед переходом к проектированию и обучению нейросетевой модели, была выполнена очистка исходного множества от противоречивых примеров, выбросов, дубликатов. Чистка проводилась параллельно со сбором данных ( т.к. сбор осуществлялся в ручную, то не брались данные подвергающиеся сомнениям или данные не

имеющие анализа по тому или иному параметру. После сбора данных, они были подгружены «Нейросимулятор 5» и проверены его алгоритмом на вбросы, которые не были выявлены, что может говорить о хорошей подготовке множества. Таким образом, объем итогового множества включает в себя 111 примеров. Данное множество было разделено на обучающее и тестирующее в соотношении 90% и 10% соответственно. Большая часть данных была собрана с интернет-ресурсов вручную [2].

Проектирование, обучение, тестирование нейронной сети выполнялись с помощью программы «Нейросимулятор 5» [3] по методике [4]. В результате оптимизации спроектированная нейронная сеть представляла собой персептрон, который имеет двадцать входных нейронов, один выходной и один скрытый слой с четырьмя нейронами.

Для оценки корректной работы спроектированной нейронной сети использовалось тестирующее множество, состоящее из 11 примеров. Средняя относительная ошибка тестирования составила 9.6%, что можно считать приемлемым результатом. На рисунке 1 представлена гистограмма, демонстрирующая разницу между фактическим и прогнозируемым нейросетью победителями матчей.

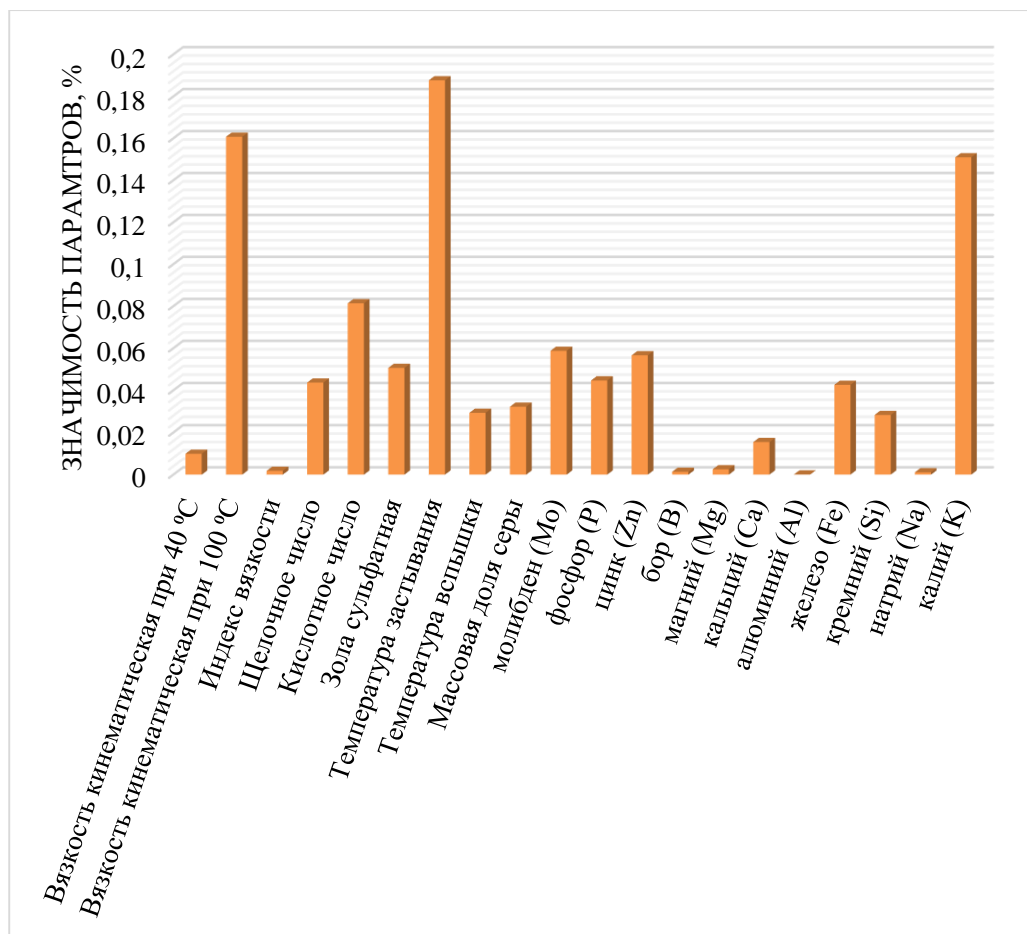


**Рисунок 1. Результат тестирования нейронной сети**

Из результатов, изображенных на рисунке 1, можно сделать вывод об адекватной работе спроектированной нейронной сети.

Оценка значимости входных параметров была выполнена с помощью программы «Нейросимулятор 5», результат отображен на рисунке 2.

Как видно из рисунка 2, наиболее значимыми являются *Температура застывания*, *Вязкость кинематическая при 100 °C* и т.д. Наиболее влиятельными параметрами являются *Вязкость кинематическая при 40 °C* и *Температура застывания*.



**Рисунок 2. Значимость входных параметров нейросетевой модели**

Параметр *калий (K)* скорее всего определён ошибочно, т.к. относится к дополнительным параметрам и лишь косвенно может влиять на классификацию масла.

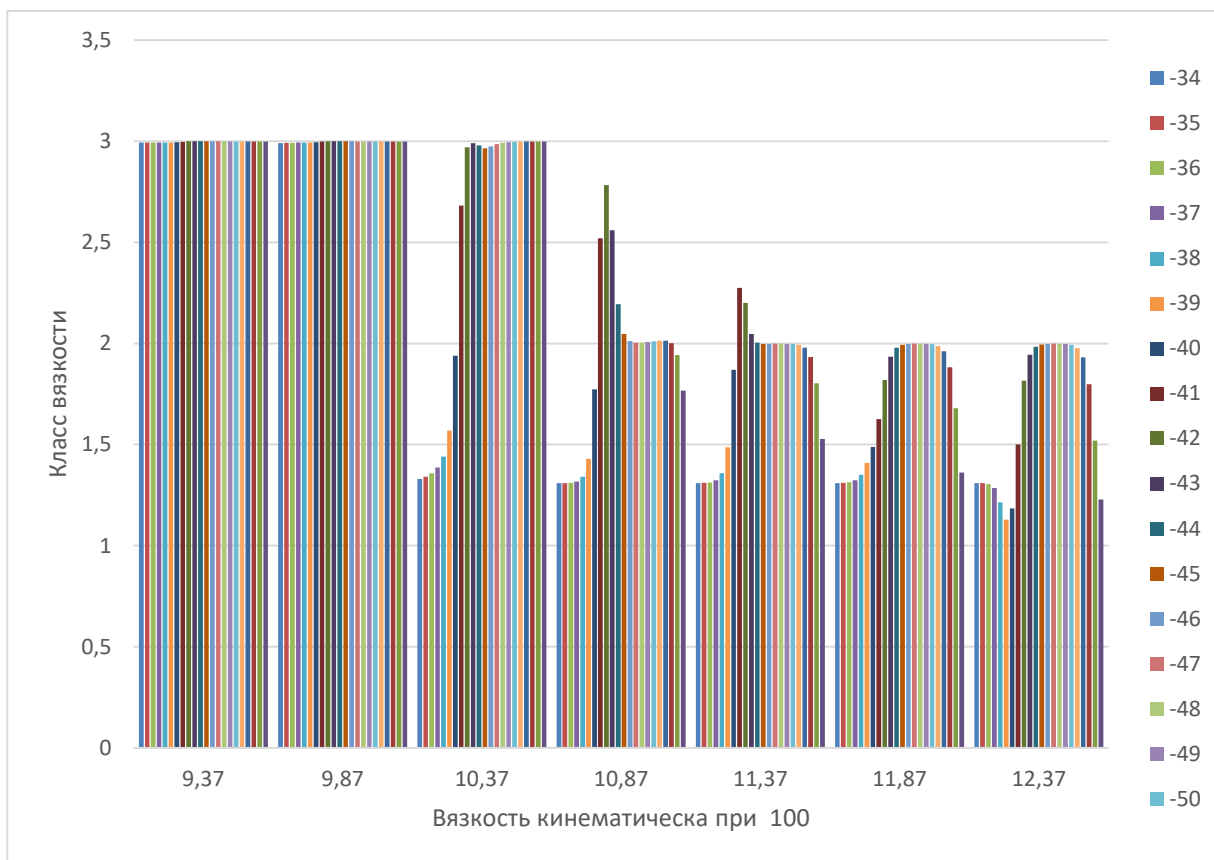
Далее было проведено исследование полученных зависимостей между входными параметрами и значения класса вязкости моторного автомобильного масла. Исследование проводилось с помощью метода «замораживания» [4], суть которого заключается в варьировании значения одного параметра и фиксировании значений всех других параметров. Для этой цели был отобран «нейтральный пример», про который нейросеть не может с уверенностью сказать какой класса вязкости моторного автомобильного масла она определяет.

Как результат мы не получили ожидаемой зависимости не по одному из параметров. Поэтому сделаем фиксацию всех параметров, кроме X2 и X7 и построим сводную таблицу, тем самым рассмотрим зависимость и значимость этих параметров вместе.

На рисунке 3 показан сводный график зависимости прогнозируемого значения класса вязкости моторного автомобильного масла от Температуры застывания и Вязкости кинематической при 100 °С. Заметна тенденция, подтверждающая значимость этих параметра вместе для спроектированной нейросетевой модели:

1. При вязкости от 9 до 10 и не зависимо от температуры мы имеем масло класса ILSAC – GF-5

2. При вязкости от 10 до 12 мы имеем зависимость температуры застывания и вязкости про 100 °С: которая обусловлена нормальным распределением.



**Рисунок 3. Сводная диаграмма зависимости температуры застывания и вязкости при 100 °С**

**Заключение.** Построена нейросетевая система для прогнозирования класса вязкости моторного автомобильного масла. Спроектированная нейросетевая модель учитывает 20 параметров: Вязкость кинематическая при 40 °С, Вязкость кинематическая при 100 °С, Индекс вязкости, Щелочное число, Кислотное число, Зола сульфатная, Температура застывания, Температура вспышки, Массовая доля серы, молибден (Mo), фосфор (P), цинк (Zn), бор (B), магний (Mg), кальций (Ca), алюминий (Al), железо (Fe), кремний (Si), натрий (Na), калий (K), и верно прогнозирует в 99% случаев. Самыми значимыми параметрами были выявлены Вязкость кинематическая при 100 °С и Температура застывания, причём не по отдельности, а вместе.

### Библиографический список

1. Mobil.ru – О чем говорят спецификации моторных масел API, ACEA, ILSAC, AAE? [Электронный ресурс]. Режим доступа: <https://www.mobil.ru/ru-ru/obslužhivanie-avtomobilya/what-do-abbreviations-acea-api-etc-mean>

2. Oil-Club.ru – Выбор моторных масел ... – Форум oil-club.ru [Электронный ресурс]. Режим доступа: <https://www.oil-club.ru/forum/>

3. Черепанов Ф.М., Ясницкий Л.Н. Нейросимулятор 4.0. Свидетельство о регистрации программы для ЭВМ RUS 2014612546. Заявка № 2014610341 от 15.01.2014.

4. Ясницкий Л.Н. Интеллектуальные системы: учебник. М.: Лаборатория знаний, 2016. 221 с.

## NEURAL NETWORK SYSTEM FOR FORECASTING VISCOSITY CLASS OF MOTOR AUTOMOTIVE OIL

*Gordeev Maxim V.*

Perm State University

Str. Bukireva, 15, Perm, Russia, 614990, max9991995@mail.ru

The paper presents a description of the development of a neural network system for predicting the viscosity class of motor vehicle oil. The system allows you to predict the viscosity class of the oil with high accuracy based on the data of its chemical analysis.

**Keywords:** artificial intelligence, neural network technologies, forecasting, engine oil, car oils, viscosity.

УДК 330.46

## ПРИМЕНЕНИЕ МНОГОФАКТОРНОЙ НЕЙРОСЕТЕВОЙ МОДЕЛИ ДЛЯ РАЦИОНАЛИЗАЦИИ БЮДЖЕТНЫХ РАСХОДОВ ДЕЯТЕЛЬНОСТИ ЛЕЧЕБНЫХ УЧРЕЖДЕНИЙ

*Бирюков Александр Николаевич*

доктор экономических наук, доцент

Стерлитамакский филиал Башкирского государственного университета, 453103, Россия, г. Стерлитамак, проспект Ленина, 49, biryukov\_str@mail.ru

Развитый в данной статье подход к многокритериальной оценке в экономических объектах в условиях неопределенности в комбинации с байесовским подходом к предобработке данных и обучению нейросетей открывает путь к созданию эффективных инструментариев моделирования. Разработан подход к оценке эффективности деятельности учреждений здравоохранения, относящимся к трудноформализуемым системам, как по множеству статистических показателей, оценивающих медицинские параметры, так и экономические.

**Ключевые слова:** фонд развития здравоохранения, учреждения здравоохранения, нейросетевая модель, критерий.

**Введение.** При планировании сметных расчетов на любом бюджетном уровне, в частности при распределении фонда развития здравоохранения (ФРЗ) между учреждениями здравоохранения (УЗ), возникает сложная и трудноформализуемая проблема количественной оценки эффективности работы УЗ в основных её аспектах. В медицинской статистике имеются множество показателей, характеризующих качество оказания медицинских услуг: дефицитность финансирования по территориальной программе государственных гарантий (ТПГГ) оказания бесплатной медицинской помощи, включая систему обязательного медицинского страхования (ОМС), эффективность управления кадровым потенциалом, коечным фондом стационара, стоимостью медицинских услуг и др. Однако все эти показатели разрознены и не приведены к обобщённому критерию эффективности, который можно было бы прогнозировать при планировании бюджета любого уровня и, тем более, оптимально управлять УЗ на основе этого критерия [1].

В данной статье предлагается обобщённый критерий  $\Phi$ , отражающий соревновательный принцип распределения ФРЗ: больше получает то УЗ, которое эффективнее работало в предшествующий период (3...5 лет) при прочих равных условиях. Для моделирования критерия  $\Phi$  в виде совокупности ряда аддитивных и мультипликативных свёрток предлагается нейросетевая динамическая модель (НСМ) вида:

$$\hat{\Phi} = F(X_1, X_2, \dots, X_n, t, W), \quad (1)$$

где  $\hat{\Phi}$  – расчётное (оцениваемое нейросетью (значение обобщённого критерия эффективности;  $\vec{X} = (X_1, X_2, \dots, X_n)$  – вектор экзогенных (объясняющих) переменных;  $t$  – время, отсчитываемое в месяцах;  $F(\cdot)$  – функция нейросетевого отображения, которая определяется выбранной архитектурой сети, видом активационных функций, алгоритмом и параметрами обучения;  $W$  – матрица параметров модели (синаптических весов нейронов).

Модель строится на основе реальных данных панельного типа по четырём УЗ, наблюдаемым в течении 5 лет ежемесячно (всего 240 наблюдений).

При построении нейросетевой модели (НСМ) нетривиальным моментом является спецификация переменных, подробно изложенная ниже.

### **1. Спецификация зависимых (объясняемых) переменных**

При спецификации зависимых (эндогенных) переменных будем следовать идеологии монографии [2], согласно которой вначале производится агрегирование в виде линейных свёрток  $\{\Phi_v\}$  отдельных выходных показателей  $\{Y_r\}$ ,  $r = \overline{1, R}$  в каждой  $v$ -той группе показателей, кластеризованных в соответствии с их содержательным смыслом, а затем получение результирующего мультипликативного критерия эффективности  $\Phi$ :

$$\Phi = \prod_v^8 \Phi_v, \quad (2)$$

где  $\Phi_v$  – частные критерии эффективности, характеризующие деятельность УЗ.

Совокупность критериев  $\Phi_1 \dots \Phi_8$  позволяет количественно оценить основные стороны деятельности УЗ:

- критерий  $\Phi_1$  характеризует уровень достижения основной цели работы УЗ;
- критерий  $\Phi_2$  оценивает дефицит финансирования ТППГ по данному ЛУ бесплатной медицинской помощи, в том числе по ОМС путём сопоставления с нормативным значением  $P_H$ ;
- критерий  $\Phi_3$  оценивает эффективность управления кадровыми ресурсами;
- критерий  $\Phi_4$  определяет объём эффективных расходов на управление коечным фондом в УЗ;
- критерий  $\Phi_6$ , оценивает уровень госпитализации;
- критерий  $\Phi_7$  определяет эффективность расходов в УЗ на управление длительностью пребывания больного на койке и уровнем госпитализации;
- критерий  $\Phi_8$  определяет эффективность расходов на управление объёмами медицинской помощи: стационарной, амбулаторной, скорой.

В целом, высокие значения показателей кадровых ресурсов, коечной мощности, уровня госпитализации, длительности лечения при низком значении продолжительности занятости койки в стационарных муниципальных УЗ, а также высокие значения объёмов по видам медицинской помощи и её стоимости отражают неэффективное управление медицинской помощью, диспропорции в оказании различных видов медицинских услуг.

При конструировании мультипликативного критерия  $\Phi$  по (1) и входящих в него сомножителей  $\Phi_v$  будем исходить из правила: *чем меньше критерии  $\Phi_v$  и  $\Phi$ , тем эффективнее работа данного УЗ*. Другими словами, в различных оптимизационных процедурах целевую функцию  $\Phi(\bar{X})$ , следует минимизировать ( $\bar{X}$  – вектор независимых переменных (входных факторов)).

Таким образом, система выходных характеристик, оценивающих эффективность работы УЗ, полностью формализована. Мультипликативный обобщённый критерий  $\Phi$  приобретает вид:

$$\Phi = \Phi_1 \cdot \Phi_2 \cdot \tilde{\Phi}_3 \cdot \tilde{\Phi}_4 \cdot \Phi_5 \cdot \Phi_6 \cdot \Phi_7 \cdot \tilde{\Phi}_8. \quad (3)$$

Чем меньше этот показатель, тем лучше оценивается работа УЗ. После образования критерия  $\Phi$  следует проверить наличие его корреляции со всеми компонентами вектора независимых переменных  $X_j, j = \overline{1, N}$ .

## 2. Спецификация объясняющих переменных

Следуя [2], выберем независимые переменные  $\{X_j\}$  так, чтобы их число было в несколько раз меньше объёма выборки  $N=240$ :

$$N \geq \xi \cdot n, \quad \xi = 2 \dots 10, \quad (4)$$

где  $\xi$  – коэффициент запаса по репрезентативности. Рекомендация (4) получена эмпирическим путём и согласуется с известными монографиями по нейросетевому моделированию сильнозашумлённых экономических объектов [2].

Нормировке подлежат факторы  $X_1, X_2, X_3, X_8, X_9, X_{10}, X_{11}, X_{12}, X_{13}, X_{14}, X_{15}$ .

Перечислим все входные факторы:

$X_1$  – среднемесячная номинально начисленная заработная плата работников УЗ (включая средний персонал, млн. руб.);  $X_2$  – среднемесячная номинально начисленная заработная плата врачей, (млн. руб.);  $X_3$  – заработная плата среднего персонала, (млн. руб.);  $X_4$  – доля отделений УЗ, переведённых преимущественно на одноканальное финансирование через систему обязательного медицинского страхования, (%);  $X_5$  – доля отделений УЗ, применяющих экономические стандарты оказания медицинской помощи, (%);  $X_6$  – доля УЗ переведённых на новую (отраслевую) систему оплаты труда, ориентированную на результат, (%);  $X_7$  – доля населённых пунктов, обеспеченных питьевой водой надлежащего качества, (%);  $X_8$  – расходы консолидированного бюджета в данном лечебном учреждении, (руб./человека);  $X_9$  – расходы консолидированного бюджета в данном лечебном учреждении на капитальное строительство, (млн. руб.);  $X_{10}$  – число врачей в расчёте на 10 тыс. человек населения (физические лица, человек);  $X_{11}$  – число среднего медицинского персонала в расчёте на 10 тыс. человек населения (физические лица, человек);  $X_{12}$  – стоимость содержания одной койки в УЗ в сутки, (руб./сутки);  $X_{13}$  – стоимость единицы объёма оказания медицинской помощи в стационаре УЗ (руб./человека);  $X_{14}$  – стоимость единицы амбулаторной медицинской помощи в УЗ, (руб./человека);  $X_{15}$  – стоимость единицы скорой медицинской помощи в УЗ, (руб./вызов);  $\tilde{X}_{16}$  – относительное время  $\tilde{t}_i$ :  $\tilde{X}_{16} \equiv \tilde{t}_i = \frac{t_i}{T}$ , где  $\tilde{t}_i$  – текущий номер месяца;  $T$  – период наблюдения ( $T = 12 \cdot 5 = 60$  месяцев).

*Замечание.* При построении нейросетевой модели абсолютные значения входных факторов целесообразно преобразовать в относительные значения, т.е. нормировать по формуле:

$$\tilde{x}_{ij} = \frac{x_{ij} - x_{j.\min}}{x_{j.\max} - x_{j.\min}} \in [0;1], \quad j = \overline{1,15} \quad (5)$$

Здесь  $i$  – сквозной номер опыта в данных панельного типа, т.е. данных, где наблюдения упорядочены как по объектам (УЗ № 1,2,3,4), так и по времени  $i = \overline{1,240}$ ;  $j$  – номер фактора. Фактор  $X_{16} \equiv \tilde{t}$  (относительное время) является безразмерным и его перенормировка по формуле (5) нецелесообразна.



### 3. Методика оценки эффективности деятельности учреждений здравоохранения

1. На основе предложенной спецификации зависимых и независимых переменных по пунктам 1 и 2 строится таблица панельных данных панельного типа (табл. 1) с учетом [3].

2. Используя таблицу 1, разбиваем множество данных  $\Omega = 240$  на подмножества обучения, кросс-валидации, тестирования:

$$\Omega = \Omega^{train} \cup \Omega^{(CV)} \cup \Omega^{test}. \quad (6)$$

Здесь используются практические рекомендации из [2,3], полученные в результате вычислительных экспериментов. При этом множества  $\Omega^{(CV)}, \Omega^{test}$  составляют 10...20% от всего множества точек  $\Omega$ .

3. Обучаем НСМ на множестве  $\Omega^{train}$ , проверяем процедуру перекрёстного подтверждения (cross validation (CV)) на множестве  $\Omega^{(CV)}$  и тестируем на  $\Omega^{test}$ .

4. Для проверки адекватности полученной НСМ реализуем процедуру обобщённого перекрёстного подтверждения [1], т.е. на одном и том же множестве  $\Omega$  (табл. 1) строится несколько (обычно 10) сетей с различными архитектурами и видом активационных функций (гиперболический тангенс, сигмоид, радиальная базисная функция). Большая часть этих сетей (80...90% от общего количества сетей, участвующих в процедуре) должна дать близкие результаты расчёта обобщённого критерия  $\Phi$  по числовой мере разброса между различными сетями:

$$\delta_1 = \max_q \left| \frac{\hat{\Phi}_q(1) - \bar{\Phi}(1)}{\bar{\Phi}(1)} \right| \cdot 100\% \leq \eta_1, \quad \tilde{t} = 1,0; \quad q = \overline{1, Q}, \quad (7)$$

а также по мере отклонения от эксперимента:

$$\delta_2 = \max_q \left| \frac{\hat{\Phi}(1) - \Phi(1)}{\Phi(1)} \right| 100\% \leq \eta_2, \quad \tilde{t} = 1,0. \quad (8)$$

Здесь обозначено  $\hat{\Phi}_q(1)$  – расчётное значение обобщённого критерия эффективности по (2), полученное в  $q$ -той НСМ в процедуре обобщённого перекрёстного подтверждения в точке  $\tilde{t} = 1,0$ ;  $\bar{\Phi}$  – среднее по множеству  $Q$  сетей значение критерия  $\Phi$  в точке  $\tilde{t} = 1,0$ ;  $\eta_1, \eta_2$  – задаваемые уровни отклонения в %. На практике достаточно задать  $\eta_1 \approx 10\%$ ,  $\eta_2 \approx 5\%$ .

После установления адекватности НСМ, можно её «эксплуатировать», т.е. использовать для прогнозных оценок в точке горизонта прогноза  $\tilde{t} = t_0 > 1,0$ .

5. Для заданного горизонта планирования  $\tilde{t}_0$ , используя методы прогнозирования временных рядов, например методы авторегрессии [4], находим прогнозные значения входных факторов  $\tilde{x}_1(\tilde{t}_0), \tilde{x}_2(\tilde{t}_0), \dots, \tilde{x}_{15}(\tilde{t}_0), \tilde{x}_{16}(\tilde{t}_0) \equiv \tilde{t}_0$  отдельно для каждого учреждения здравоохранения ((УЗ) = 1, 2, 3, 4).

## Данные для построения НСМ

Номер учреждения здравоохранения (УЗ)	Сквозной номер наблюдения, $i$	Безразмерные входные факторы $\{\tilde{x}_j\}$			Частные критерии эффективности $\{\Phi_v\}$			Обобщённый критерий эффективности $\Phi$
		$\tilde{x}_1$	$\tilde{x}_j$	$\tilde{x}_{16}$	$\Phi_1$	$\Phi_v$	$\Phi_8$	
1	1			0,0166				
	2			0,0333				
	...							
	$i$							
	...							
	60			1,0				
...								
4	181			0,0166				
	182			0,0333				
	...			$\vdots$				
	240			1,0				

6. Подставляя в НСМ прогнозные значения входных факторов  $\tilde{x}_1(\tilde{t}_0), \tilde{x}_2(\tilde{t}_0), \dots, \tilde{x}_{15}(\tilde{t}_0)$  и время горизонта прогноза  $\tilde{t}_0 > 1,0$ , получаем прогнозную оценку  $\hat{\Phi}$  для каждого учреждения здравоохранения ((УЗ) = 1,2,3,4).

7. Можно повторить процедуры по пунктам 1...6, выстраивая отдельные НСМ для каждого учреждения здравоохранения отдельно, и сравнить результаты оценок  $\hat{\Phi}$  с единой НСМ, реализуемой в пунктах 1...6. Однако множество данных здесь сузится во столько раз, сколько УЗ в панельных данных (в нашем примере оно составит  $240/4 = 60$  точек). Если на множества  $\Omega^{(CV)}$  и  $\Omega^{test}$  отвести по 10%, т.е. 12 точек, то в множестве  $\Omega^{train}$  останется всего 48 точек и может возникнуть проблема дефицита данных при столь большом количестве входных факторов ( $n=16$ ).

В этом случае можно рекомендовать не уменьшать множество  $\Omega$ , разбивая НСМ на подмодели для каждого объекта, а ввести в общую НСМ бинарные (булевы) переменные и т.д. Число таких бинарных независимых переменных будет на единицу меньше числа моделируемых объектов (УЗ).

**Заключение.** Можно сделать вывод, что на основе предложенного в качестве инструментария моделирования «собрание» нейросетевых моделей, разработана методика спецификации зависимых и независимых переменных выходных показателей.

В этой связи применение предложенной практической методики нейросетевого моделирования показателей и критерия оценки эффективности деятельности учреждений здравоохранения при ограниченности ресурсов государства в сочетании с широким спектром задач в этой социальной области, требуют усиления контроля за эффективностью их использования и придает необходимость рационализации бюджетных расходов.

## Библиографический список

1. Горбатков С.А., Бирюков А.Н. и др. Методологические основы разработки нейросетевых моделей экономических объектов в условиях неопределенности // Монография. М.: Издательский дом «Экономическая газета», 2012. – 494 с.

2. Бирюков А.Н. Многокритериальная оценка эффективности работы учреждений здравоохранения в условиях неопределенности с помощью нейросетевой модели // Монография: Проблемы разработки и реализации стратегии предприятия. Книга 2; Новосибирск: Центр развития научного сотрудничества, издательство «СИБПРИНТ», 2010.- С.-188-212

3. Об утверждении методик расчета показателей для оценки эффективности деятельности высших должностных лиц (руководителей высших исполнительных органов государственной власти) субъектов Российской Федерации и деятельности органов исполнительной власти субъектов Российской Федерации. Постановление Правительства РФ от от 3 апреля 2021 г. № 542

## APPLICATION OF A MULTIFACTOR NEURAL NETWORK MODEL FOR RATIONALIZATION OF BUDGET EXPENDITURES OF MEDICAL INSTITUTIONS

*Biryukov Alexander N.*

Doctor of Economics, Associate Professor  
Sterlitamak Branch of Bashkir State University,  
49 Lenin Avenue, Sterlitamak, 453103, Russia,  
e-mail: biryukov\_str@mail.ru

The approach developed in this article to multi-criteria assessment in economic objects under uncertainty in combination with the Bayesian approach to data preprocessing and neural network training opens the way to the creation of effective modeling tools. An approach has been developed to assess the effectiveness of health care institutions related to difficult-to-formalize systems, both by a variety of statistical indicators evaluating medical parameters and economic ones.

**Keywords:** healthcare development fund, healthcare institutions, neural network model, criterion.

## ПРОГНОЗИРОВАНИЕ КОЛИЧЕСТВА АРЕНДОВАННЫХ ВЕЛОСИПЕДОВ

*Осмоловский Всеволод Андреевич*

Пермский государственный национальный исследовательский университет,  
ПМИ. 614990, Россия, г. Пермь, ул. Букирева, 15,  
sevaosm01@mail.ru

В статье представлено описание разработки нейросетевой системы для прогнозирования количества арендованных велосипедов. Система позволяет оценить, сколько велосипедов арендуют в определенный час. Это поможет компаниям, предоставляющим данные транспортные средства, выгоднее и удобнее для себя управлять процессом сдачи в аренду, устанавливать стоимость в зависимости от спроса, отправлять велосипеды к определенному месту перед «час пиком» использования.

**Ключевые слова:** искусственный интеллект, нейросетевые технологии, прогнозирование, велосипед, аренда, арендаторы.

**Введение.** В определенное время в течение дня человеку может понадобиться добраться куда-либо при помощи велосипеда. Различные компании предоставляют аренду данных транспортных средств по всему населенному пункту, чтобы охватить максимальную площадь территории потенциальных клиентов. Данный проект показывает, какие факторы коррелируют с количеством арендованных велосипедов в определенное время. Одним из наиболее важных аспектов проекта машинного обучения является понимание основных функций, которые соотносятся с атрибутом, который мы пытаемся предсказать [1]. Основная цель заключается в поиске множества данных, на котором будет обучаться и тестироваться нейросеть. Найденное готовое множество включает в себя 10 входных параметра и 1 выходной. Оно собиралось на основе сбора статистики с пользователей данной системы.

Входные параметры: X1 – месяц, X2 – час, X3 – наличие праздника, X4 – день недели, X5 – наличие рабочего дня, X6 – погода, X7 – температура, X8 – температура по ощущению, X9 – влажность, X10 – скорость ветра. Выходной параметр D1 – количество арендованных велосипедов. Некоторые параметры были непрерывно закодированы: месяц (1 – январь, 2 – февраль, 3 – март, 4 – апрель, 5 – май, 6 – июнь, 7 – июль, 8 – август, 9 – сентябрь, 10 – октябрь, 11 – ноябрь, 12 – декабрь); наличие праздника (0 – не праздник, 1 – праздник); день недели (1 – понедельник, 2 – вторник, 3 – среда, 4 – четверг, 5 – пятница, 6 – суббота, 7 – воскресенье); наличие выходного дня (0 – не выходной, 1 – выходной); погода (1 – ясно, малооблачно, облачно; 2 – туман, переменная облачность, дождь; 3 – дождь, снег, гроза, рассеянные облака; 4 – сильный дождь, снегопад, град).

Температура измеряется по формуле:  $(t - t_{\min}) / (t_{\max} - t_{\min})$ , где  $t$  – актуальная температура,  $t_{\min} = -8$ ,  $t_{\max} = +39$ ; температура по ощущению измеряется по формуле:  $(t - t_{\min}) / (t_{\max} - t_{\min})$ , где  $t$  – актуальная температура,  $t_{\min} = -16$ ,  $t_{\max} = +50$ ; влажность выражается через актуальное значение влажности, деленное на 100; скорость ветра выражается через актуальное значение скорости ветра, деленное на 67. Шапка таблицы с данными по описанным выше входным параметрам:

X1	X2	X3	X4	X5	X6	X7	X8	X9	X10
Месяц	Час	Праздник	День недели	Рабочий день	Погодные условия	Температура	Ощущаемая темп.	Влажность	Скорость ветра

Исходное множество было взято с UCI Machine Learning Repository [2] и содержало около 17 тыс. строк, из которых было отобрано 6900 строк. Дальнейшая обработка данных производилась с учетом деления отобранных строк в процентном соотношении 70:20:10. Для обучения нейросети было использовано 4800 строк, валидирующее множество состояло из 1500 строк и для прогноза использовалось 700 строк данных.

Проектирование, обучение, тестирование нейронной сети выполнялись с помощью программы «Нейросимулятор 5» [3] по методике [1]. В результате оптимизации спроектированная нейронная сеть представляла собой персептрон, который имеет десять входных нейронов, один выходной и 3 скрытых слоя с 14, 11 и 8 нейронами соответственно, количество которых оказалось оптимальнее всего, учитывая оценку количества нейронов в скрытых слоях по теореме Арнольда-Колмогорова.

Для оценки корректной работы спроектированной нейронной сети использовалось валидирующее множество, состоящее из 1500. На графике использовано 30 случайных примеров из общего количества.

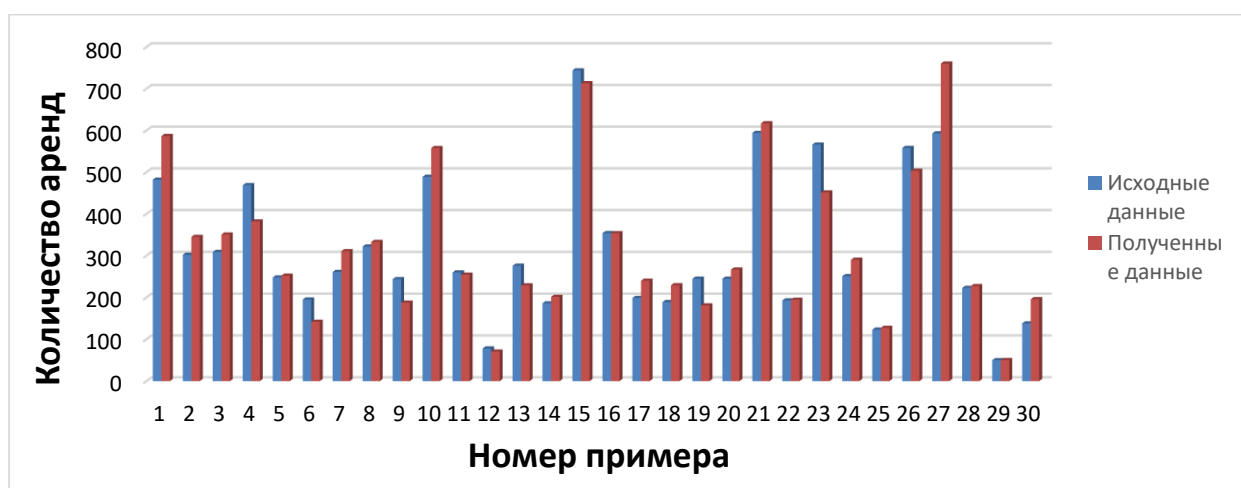


Рисунок 1. Результат тестирования нейронной сети

Средняя относительная ошибка тестирования составила 6%, что можно считать приемлемым результатом. На рисунке 1 представлена гистограмма,

демонстрирующая разницу между фактическим и прогнозируемым нейросетью количеством аренд.

Из результатов, изображенных на рисунке 1, можно сделать вывод об адекватной работе спроектированной нейронной сети.

Оценка значимости входных параметров была выполнена с помощью программы «Нейросимулятор 5», результат отображен на рисунке 2.

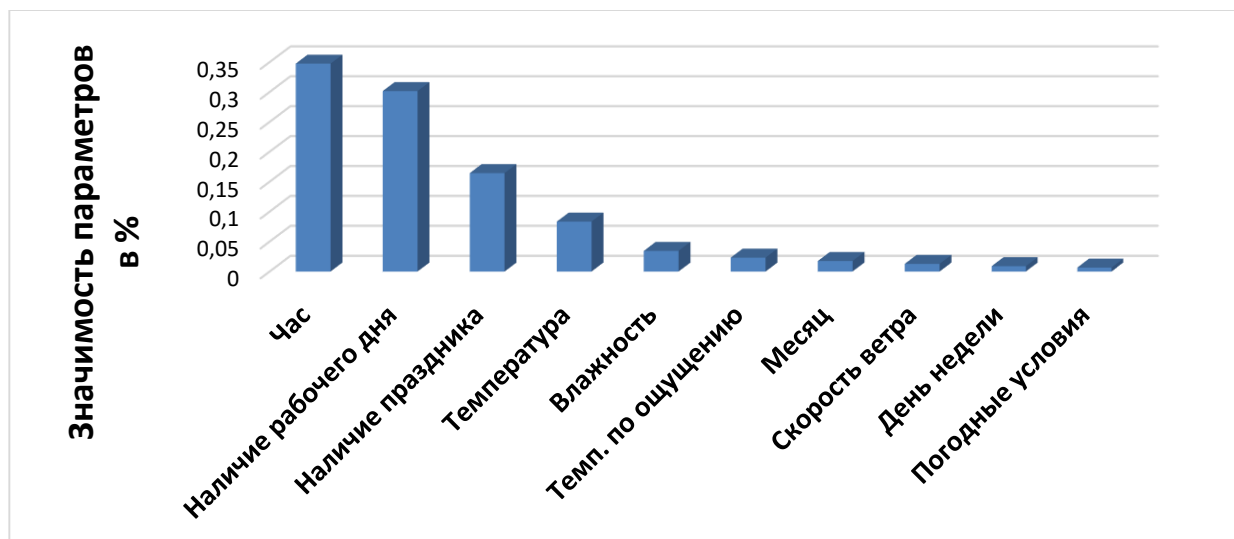


Рисунок 2. Значимость параметров нейросетевой модели

Как видно из рисунка 2, наиболее значимыми являются определенный час в сутках, наличие рабочего дня, наличие праздника и температура. Как и ожидалось, наиболее влиятельным параметром является определенный час, так как велосипеды в большинстве случаев используют как транспортные средства, чтобы добраться в определенное время в нужное место (например, ехать из дома на работу или наоборот).

Далее было проведено исследование полученных зависимостей между входными параметрами и количеством аренд. Исследование производилось с помощью метода «замораживания» [1], суть которого заключается в варьировании значения одного параметра и фиксировании значений всех других параметров. Для этой цели был отобран «нейтральный пример», про который нейросеть не может с уверенностью сказать, какой показатель аренд будет в это время.

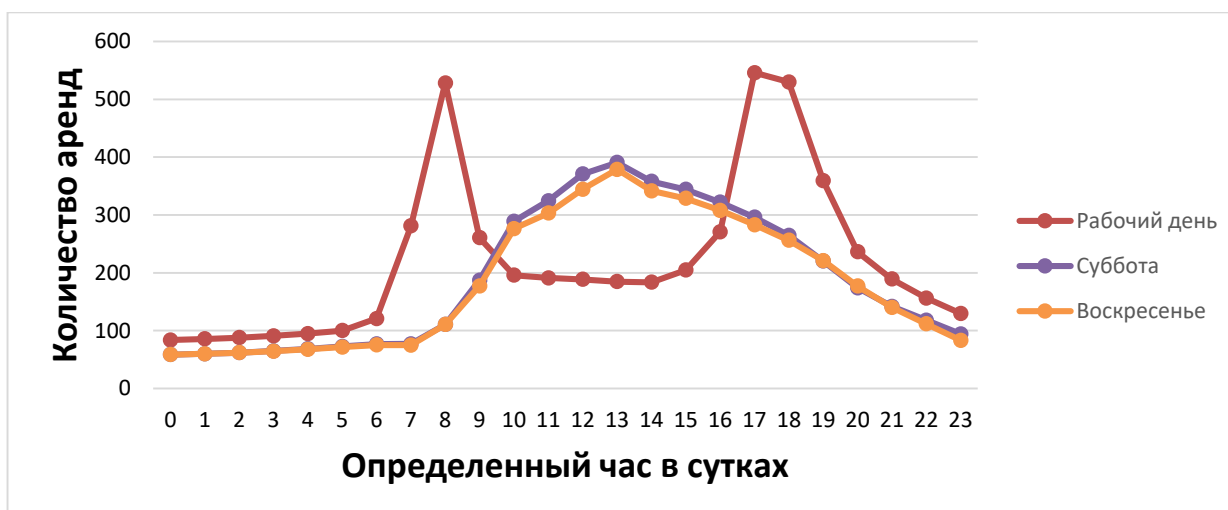
На рисунке 3 показан график зависимости количества аренд велосипедов от определенного часа в сутках. Можно по нему провести небольшой анализ. Люди начинают активно арендовать транспортные средства к утру (6:00 – 8:00), чтобы доехать до работы, учебы. Далее происходит резкий спад после 9:00 для предположительной трудовой деятельности до 12:00 – 13:00, когда необходимо сделать перерыв для обеда. Очередной спад, хоть и небольшой, для продолжения трудовой деятельности наблюдается до 17:00 – конца рабочего, учебного дня. Люди активно берут велосипеды, для того чтобы добраться домой или на вечерние встречи. После 20:00 происходит резкий спад аренды, люди, предположительно, приезжают домой (конечная точка) и до утра велосипедом не пользуются. Был выбран и изменен по пара-

метрам часа рабочий день (четверг) в июле без наличия праздника, с погодными условиями со значением 1, с температурой +28 градусов по Цельсию, с влажностью 29% и силой ветра 18 м/с.



**Рисунок 3. Зависимость количества аренд от определенного часа**

На рисунке 4 представлена зависимость количества аренд велосипедов от выходного (суббота и воскресенье) или рабочего дня.

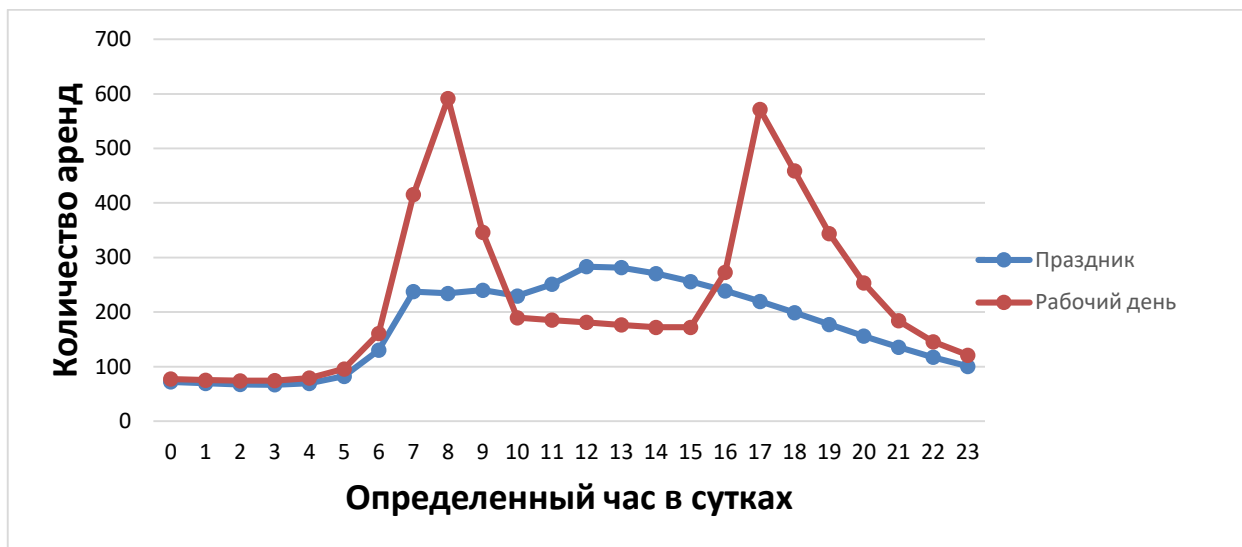


**Рисунок 4. Зависимость количества аренд велосипедов от наличия выходного дня (суббота и воскресенье) и рабочего дня**

Анализируя ее, можно сказать, что в выходной день люди катаются, преимущественно, днем (активно с 8:00 до 18:00), нет никаких перепадов в связи с отсутствием рабочего графика. Возможно, велосипедом пользуются для активного отдыха или для передвижения до необходимых мест в городе. Был выбран и изменен по параметрам часа день (вторник, суббота и воскресенье) в мае без наличия праздника, с погодными условиями со значением 3, с температурой +20 градусов по Цельсию, с влажностью 70% и силой ветра 10 м/с.

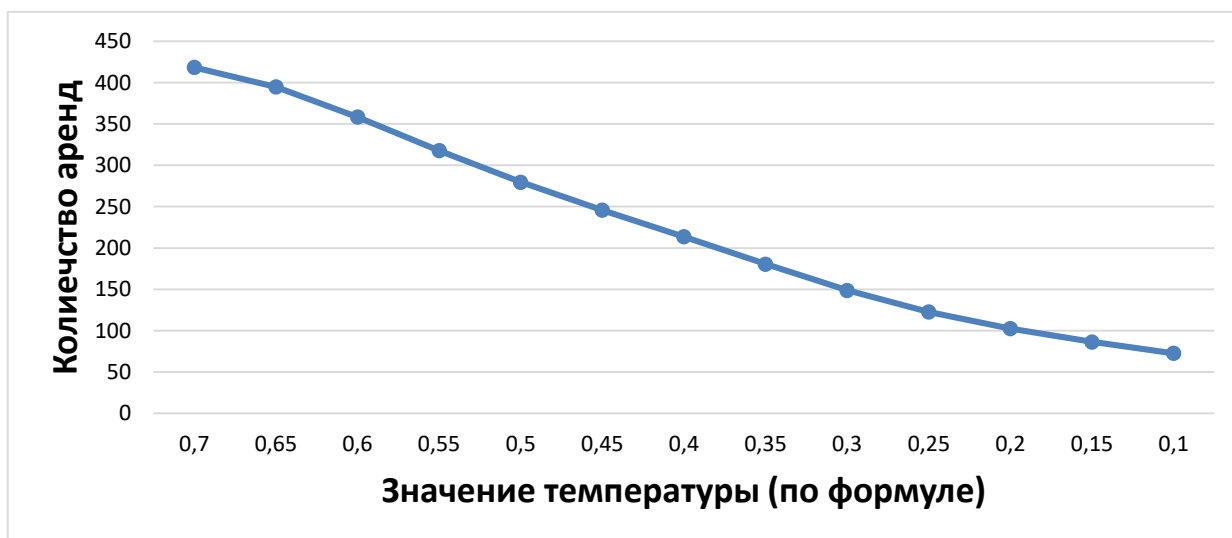
На рисунке 5 представлена зависимость количества аренд велосипедов в праздник (выходной день) и рабочий день. Исследуя ее, можно сказать, что в праздник люди катаются на протяжении всего дня (активно с 7:00 до 19:00), нет никаких перепадов в связи с отсутствием рабочего графика. Возможно,

велосипедом пользуются для активного отдыха или для передвижения до необходимых мест в городе аналогично обычному выходному дню. Был выбран и изменен по параметрам часа день (среда) в сентябре с наличием праздника и без, с погодными условиями со значением 1, с температурой +17,5 градусов по Цельсию, с влажностью 77% и силой ветра 13,4 м/с.



**Рисунок 5. Зависимость количества аренд велосипедов от наличия праздника (выходной день) и рабочего дня**

На рисунке 6 представлена зависимость количества аренд велосипедов от температуры на улице.



**Рисунок 6. Зависимость количества аренд велосипедов от температуры**

Данная зависимость показывает, что с понижением температуры (вычисленной и ощущаемой) люди меньше арендуют велосипеды, так как в более холодные погодные условия кататься менее приятно. Был выбран и изменен по параметрам температуры выходной день (воскресенье) в ноябре без наличия праздника в 18 часов, с погодными условиями со значением 2, с



влажностью 63% и силой ветра 6,7 м/с. В данном случае значение температуры разнится от значения 0,7 (примерно +30,5 градусов по Цельсию) до значения 0,1 (примерно -3 градусов по Цельсию), исходя из формулы, описанной выше.

На рисунке 7 представлена зависимость количества аренд велосипедов от влажности.



**Рисунок 7. Зависимость количества аренд велосипедов от влажности**

По данному графику можно сделать вывод, что с уменьшением влажности количество аренд увеличивается. Был выбран и изменен по параметрам влажности выходной день (суббота) в феврале без наличия праздника в 12 часов, с погодными условиями со значением 1, с температурой +7 градусов по Цельсию и силой ветра 10 м/с. В данном случае значение влажности разнится от 0,89 (89%) до 0,09 (9%).

На рисунке 8 представлена зависимость количества аренд велосипедов от силы ветра.



**Рисунок 8. Зависимость количества аренд велосипедов от силы ветра**

По графику из рисунка 8 можно сказать, что с уменьшением силы ветра количество аренд увеличивается. Был выбран и изменен по параметрам силы ветра рабочий день (четверг) в апреле без наличия праздника в 8 часов, с по-

годными условиями со значением 3, с температурой +20 градусов по Цельсию и влажностью 20%. Значение силы ветра изменяется от 0,45 (около 30 м/с) до значения 0,05 (около 3 м/с).

Полученные результаты исследований не противоречат реальности, что подтверждает, что спроектированную нейронную сеть можно считать пригодной для прогнозирования аренды количества велосипедов.

**Заключение.** Построена система нейросетевого прогнозирования количества арендованных велосипедов. Спроектированная нейросетевая модель учитывает 10 параметров: месяц, определенный час в сутках, наличие праздника, день недели, наличие выходного дня, погодные условия, вычисленная температура, ощущаемая температура, уровень влажности, сила ветра. Методом сценарного прогнозирования построены графики зависимостей предсказанного количества аренд от изменения входных параметров. Применение такого набора параметров в модели позволяет с высокой точностью прогнозировать количество аренды. Данный набор параметров может быть изменен для прогнозирования с условием добавление различных факторов, которые будут кардинально менять значения аренды, если, например, в городе будет проходить концерт известной музыкальной группы или будут проводиться всемирные спортивные мероприятия.

### **Библиографический список**

1. Ясницкий Л.Н. Интеллектуальные системы. М.: Лаборатория знаний, 2016. – 221с.
2. UCI Machine Learning Repository – сайт по предоставлению баз данных для машинного обучения [Электронный ресурс]. Режим доступа: <https://archive.ics.uci.edu/ml/index.php>
3. Черепанов Ф.М., Ясницкий Л.Н. Нейросимулятор 4.0. Свидетельство о регистрации программы для ЭВМ RUS 2014612546. Заявка № 2014610341 от 15.01.2014.

## **NEURAL NETWORK FOR PREDICTING THE NUMBER OF BIKE RENTALS**

*Osmolovsky Vsevolod A.*

Perm State University

Str. Bukireva, 15, Perm, Russia, 614990, sevaosm01@mail.ru

The system allows you to accurately predict the number of rented bicycles depending on a certain time in the day, the availability of a day off and the weather. With the help of the developed intellectual system, a study of the subject area was carried out, patterns of practical importance were identified.

**Keywords:** artificial intelligence, neural network, forecasting, rent, bicycle.

## НЕЙРОСЕТЕВАЯ СИСТЕМА ПРОГНОЗИРОВАНИЯ КАЧЕСТВА ВИНА НА ОСНОВЕ ДАННЫХ О ЕГО ХИМИЧЕСКОМ СОСТАВЕ

*Епишина Наталья Валерьевна*  
*Семёнов Сергей Павлович*

Пермский государственный национальный исследовательский университет,  
ПМИ. 614990, Россия, г. Пермь, ул. Букирева, 15, epishina.nata@bk.ru,  
rpadvans@yandex.ru.

В статье представлено описание разработки нейросетевой системы для прогнозирования качества вина. Система позволяет с большой точностью предсказать оценку качества вина на основе данных о его химическом составе. С помощью разработанной интеллектуальной системы проведено исследование предметной области, выявлены закономерности, имеющие практической значение.

**Ключевые слова:** искусственный интеллект, нейросетевые технологии, прогнозирование, алкоголь, вино.

**Введение.** Вино – древнейший алкогольный напиток с многовековой историей, который по настоящее время считается одним из самых популярных среди потребителей. Цена вина зависит от довольно абстрактной концепции оценки вина дегустаторами, мнение среди которых может иметь высокую степень изменчивости. Так же еще одним ключевым фактором в сертификации и оценке качества вина являются физико-химические испытания, которые проводятся в лабораторных условиях и учитывают такие факторы, как кислотность, уровень pH, наличие сахара и другие химические свойства. Знание оценки качества может помочь в выборе вина, сопоставив предполагаемую оценку качества и предлагаемую цену изделия в магазине, покупатель сможет выбрать подходящее вино для себя.

При анализе литературных источников выяснилось, что работ на тему прогнозирования оценки качества вина было проведено мало. В работах рассматривают множества с 3-5 тыс. значений, в основном погрешность нейросетевых моделей составляет 10 – 15 %, работы выполнены в Deductor Studio Academic и в “Эйдос”. В основном исследования качества вина проводятся в специальных лабораториях и могут быть выполнены только обученными людьми. Самостоятельно получить данных химического состава определенной марки вина не просто [2, 3, 7, 8].

Основная цель настоящей работы заключается в сборе множества данных химического состава вин, а также создание и обучение нейросетевой модели на этих данных. Конечным результатом должна быть нейросетевая система, способная прогнозировать оценку качества вина по ее химическому составу.

Для создания нейросетевой системы были выбраны следующие параметры: X1 – фиксированная кислотность, X2 – летучая кислотность, X3 – лимонная кислота, X4 – остаточный сахар, X5 – хлориды, X6 – свободный диоксид серы, X7 – общий диоксид серы, X8 – плотность, X9 – рН, X10 – сульфаты, X11 – алкоголь. Выходной параметр – оценка качества вина.

Обучающее множество было собрано с помощью программы-парсера с интернет-ресурсов. Перед переходом к проектированию и обучению нейросетевой модели, была выполнена очистка исходного множества от избыточности и дубликатов. Например, данные были подобраны так, чтобы было задействовано как можно больше оценок качества (от 4 до 8). Таким образом, объем итогового множества включает в себя 352 примера. Данное множество было разделено на обучающее, валидирующее и тестирующее в соотношении 80%, 14% и 6% соответственно. Большая часть данных была собрана с интернет ресурса [1].

Проектирование, обучение, тестирование нейронной сети выполнялись с помощью программы «Нейросимулятор 5» [5] по методике [4, 6]. Выбросы статистической информации обнаруживались и удалялись с помощью метода [9]. В результате оптимизации спроектированная нейронная сеть представляла собой персептрон, который имеет одиннадцать входных нейронов, один выходной и один скрытый слой с одним нейроном.

Для оценки корректной работы спроектированной нейронной сети использовалось валидирующее множество, состоящее из 47 примеров. Средняя относительная ошибка тестирования составила 12,6%, что можно считать приемлемым результатом. На рисунке 1 представлена гистограмма, демонстрирующая разницу между фактической и прогнозируемой нейросетью оценкой качества.

Из результатов, изображенных на рисунке 1, можно сделать вывод об адекватной работе спроектированной нейронной сети.

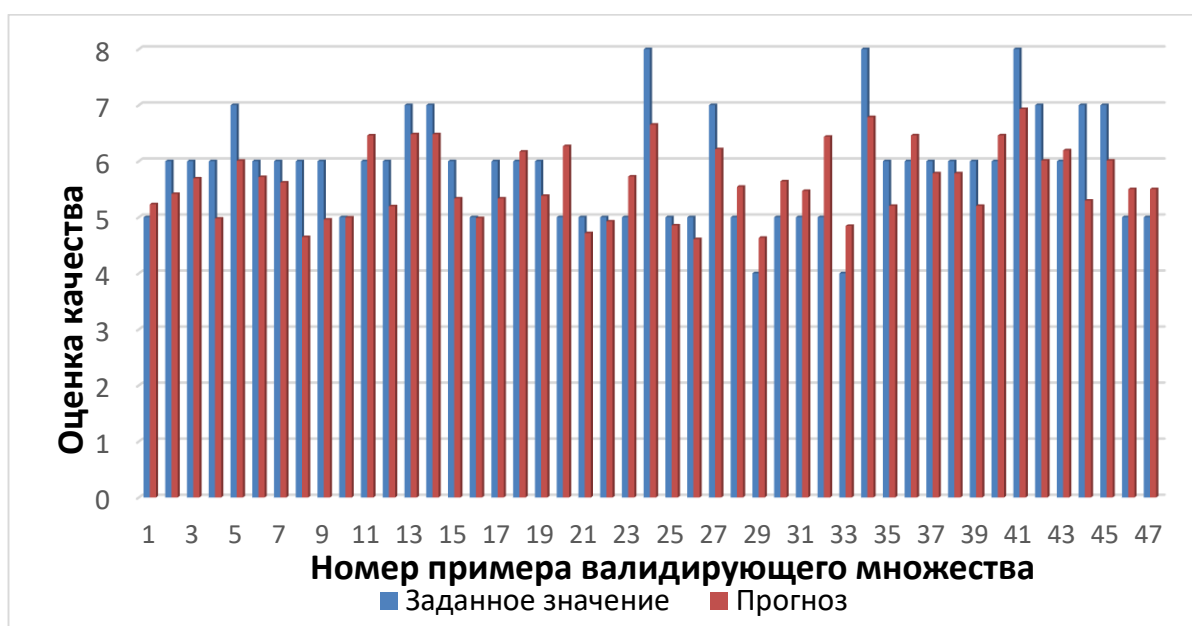
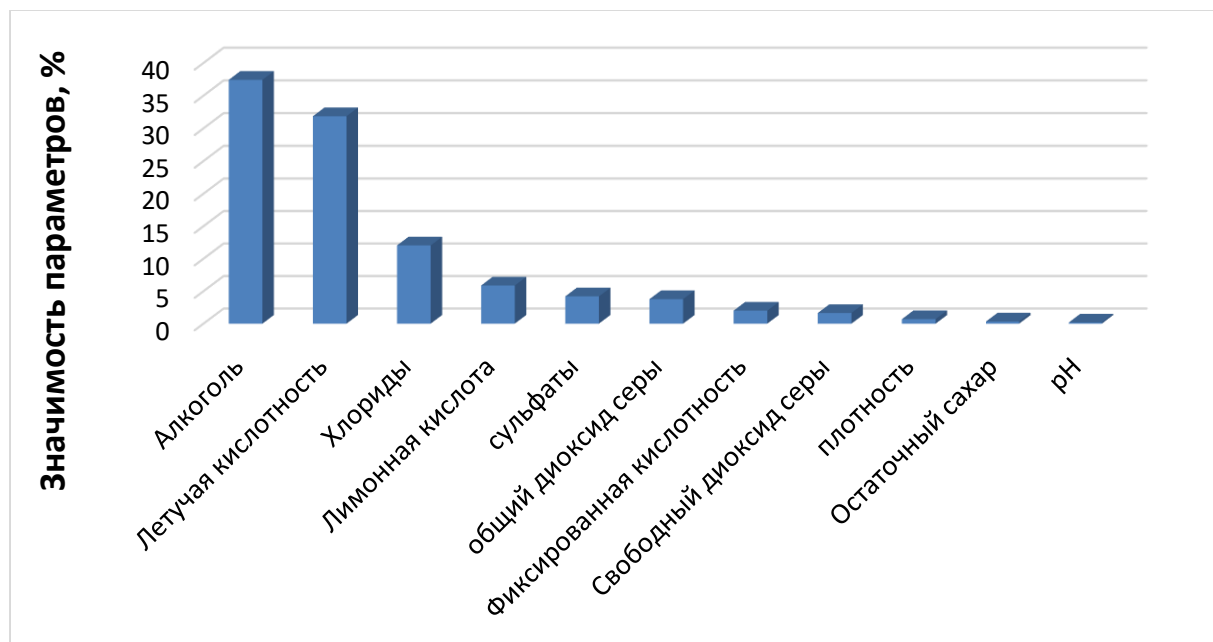


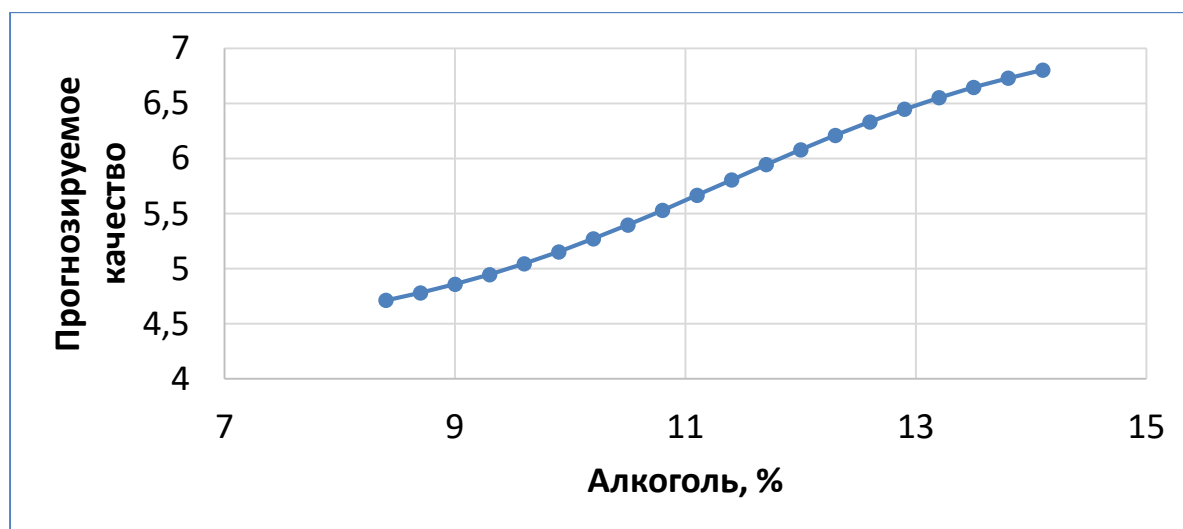
Рисунок 1. Результат тестирования нейронной сети

Оценка значимости входных параметров была выполнена с помощью программы «Нейросимулятор 5», было найдено среднее значение обучающего и валидирующего множества, результат отображен на рисунке 2.



**Рисунок 2. Значимость входных параметров нейросетевой модели**

Как видно из рисунка 2, наиболее значимыми являются алкоголь, летучая кислотность, хлориды и т.д. Наиболее влиятельным параметром является процент алкоголя в напитке. Вино – сложное химическое соединение. На его характеристики очень существенное влияние оказывает уровень алкоголя.



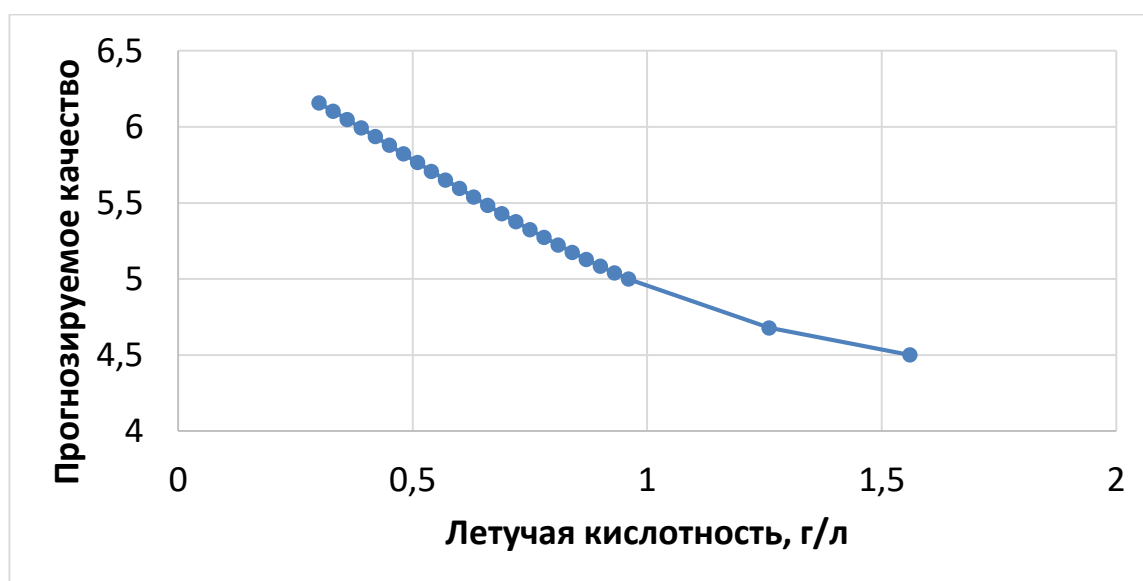
**Рисунок 3. Зависимость прогнозируемой оценки качества вина от процента алкоголя**

Далее было проведено исследование полученных зависимостей между входными параметрами и оценкой качества. Исследование производилось с помощью метода «замораживания», суть которого заключается в варьирова-

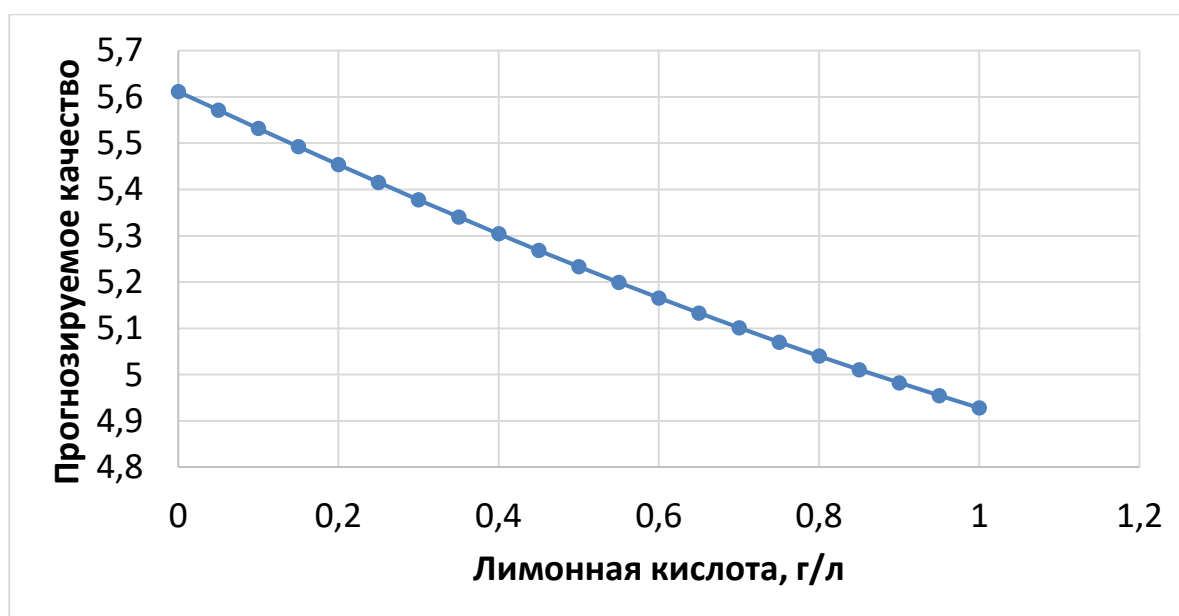
нии значения одного параметра и фиксировании значений всех других параметров. Для этой цели был отобран «нейтральный пример», про который нейросеть не может с уверенностью сказать, какой оценкой качества обладает алкогольный напиток.

На рисунке 3 показан график зависимости прогнозируемой оценки качества вина от процента алкоголя. В том случае, при более высоком содержании алкоголя, нейросеть прогнозирует увеличение качества вина.

На рисунке 4 продемонстрирована зависимость качества от летучей кислотности. Можно заметить, что содержание летучей кислотности оказывает сильное влияние на качество вина, при увеличении показателя кислотности, понижается качество, причем сильнее спад идет после 1 г/л.



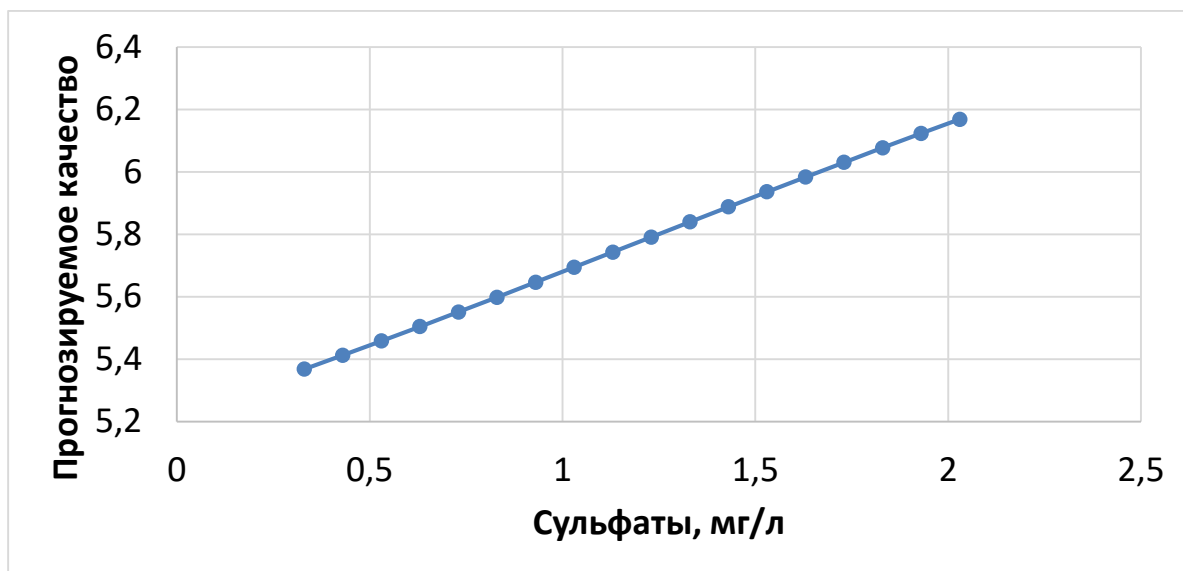
**Рисунок 4. Зависимость прогнозируемой оценки качества вина от летучей кислотности**



**Рисунок 5. Зависимость прогнозируемой оценки качества вина от лимонной кислоты**

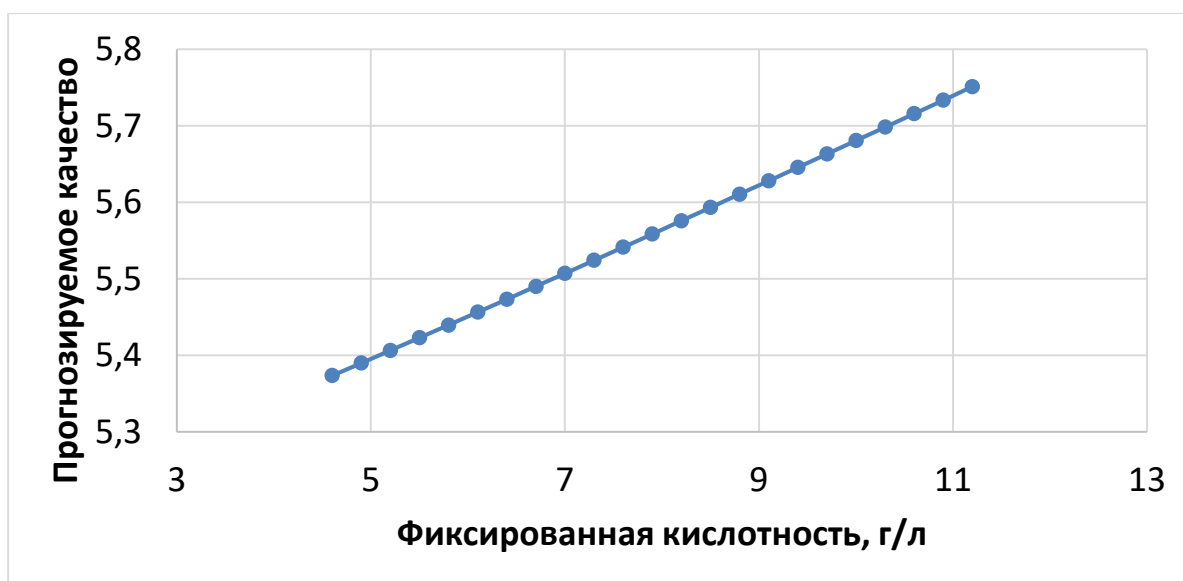
На рисунке 5 изображен график зависимости оценки качества от лимонной кислоты. Как видно из графика, качество вина резко уменьшается при повышении содержания лимонной кислоты.

На рисунке 6 продемонстрирована зависимость прогнозируемой оценки качества от количества сульфатов. Как видно из графика, количество сульфатов очень сильно влияет на качество вина, с повышением содержания сульфатов, повышается качество.



**Рисунок 6. Зависимость прогнозируемой оценки качества вина от сульфатов**

На рисунке 7 продемонстрирована зависимость прогнозируемой оценки качества от фиксированной кислотности. Содержание фиксированной кислотности не особо сильно влияет на качество, но, не смотря на это, при повышении фиксированной кислотности, повышается качество.



**Рисунок 7. Зависимость прогнозируемой оценки качества вина от лимонной кислоты**

Выше были продемонстрированы диаграммы, когда показатель вино-материала оказывал видимое влияние на качество вина, в остальных случаях особого изменения качества вина не замечено.

Полученные результаты исследований не противоречат реальности, что подтверждает, что спроектированную нейронную сеть можно считать пригодной для прогнозирования оценки качества вина.

**Заключение.** Построена система нейросетевого прогнозирования оценки качества вина. Спроектированная нейросетевая модель учитывает 11 параметров: фиксированная кислотность, летучая кислотность, лимонная кислота, остаточный сахар, хлориды, свободный диоксид серы, общий диоксид серы, плотность, рН, сульфаты, алкоголь. Методом сценарного прогнозирования построены графики зависимостей прогнозируемой оценки качества от изменения входных параметров. Применение такого набора параметров в модели позволяет с высокой точностью прогнозировать оценку качества вина.

### Библиографический список

1. Исходные данные. Набор данных о качестве вина. [Электронный ресурс]. Режим доступа: <https://www.kaggle.com/yasserh/wine-quality-dataset>
2. Анализ данных DataSet UCI – данные вина. [Электронный ресурс]. Режим доступа: <https://russianblogs.com/article/70032027666/>
3. Сырьё для производства виноградных вин. Химический состав виноградных вин. [Электронный ресурс]. Режим доступа: <https://www.bestreferat.ru/referat-119245.html>
4. Ясницкий Л.Н. Интеллектуальные системы: учебник. М.: Лаборатория знаний, 2016. 221 с.
5. Черепанов Ф.М., Ясницкий Л.Н. Нейросимулятор 4.0. Свидетельство о регистрации программы для ЭВМ RUS 2014612546. Заявка № 2014610341 от 15.01.2014.
6. Ясницкий Л.Н. Введение в искусственный интеллект. М.: Издательский центр «Академия», 2005. – 176 с.
7. Научный журнал КубГАУ, №149(05), 2019 года. [Электронный ресурс]. Режим доступа: <http://ej.kubagro.ru/2019/05/pdf/15.pdf>
8. PennState Eberly College of Science, Analysis of Wine Quality Data Режим доступа: <https://online.stat.psu.edu/stat857/node/223/>
9. Yasnitsky L.N. Algorithm for searching and analyzing abnormal observations of statistical information based on the Arnold – Kolmogorov – Hecht-Nielsen theorem // International Journal of Advanced Trends in Computer Science and Engineering. 2020. 9(2). Pp. 1814-1819.



# NEURAL NETWORK SYSTEM FOR PREDICTING THE WINE QUALITY BASED ON DATA ON ITS CHEMICAL COMPOSITION

*Epishina Natalya V., Semenov Sergey P.*

Perm State University

Str. Bukireva, 15, Perm, Russia, 614990, epishina.nata@bk.ru,  
rpadvans@yandex.ru

The article describes the development of a neural network system for predicting quality. The system allows you to predict with great accuracy the assessment of the quality of wine based on data on its chemical composition. With the help of the developed intellectual system, a study of the subject area was carried out, patterns of practical importance were identified.

**Keywords:** artificial intelligence, neural network technologies, forecasting, alcohol, wine.

УДК 004.032.26

## НЕЙРОСЕТЕВАЯ СИСТЕМА ПРОГНОЗИРОВАНИЯ КАЧЕСТВА КРАСНОГО ВИНА

*Тарасова Валентина Николаевна*

Национальный исследовательский университет «Высшая школа экономики»,  
614070, Россия, г. Пермь, ул. Студенческая, 38, valyatar146@yandex.ru

В статье представлено описание разработки нейросетевой системы для прогнозирования оценки качества красного вина. Система позволяет с большой точностью предсказать оценку качества красного вина на основании данных о химическом составе вина. С помощью разработанной интеллектуальной системы проведено исследование предметной области, выявлены закономерности, имеющие практического значения.

**Ключевые слова:** искусственный интеллект, нейросетевые технологии, прогнозирование, вино, оценка качества, качество вина.

**Введение.** На протяжении многих лет ведутся дискуссии о пользе и вреде красного вина для здоровья человека. Некоторые ученые были убеждены, что вино оказывает благоприятное воздействие на здоровье человека. Так, профессор Н.Н. Простосердов утверждал, что "вино усиливает защитные функции организма, многие болезнетворные микробы погибают в вине в течение 15 минут, при добавлении в воду 1/3 вина можно считать ее обеззараженной", а профессор медицины А.Ф. Голубев писал: "...вино, вовремя и в

надлежащем количестве данное больному, является лечебным фактором высокого значения"[1]. В настоящее время вопрос о полезных свойствах красного вина остается актуальным, однако ученые сходятся во мнении, что лишь качественное вино способно оказывать благоприятное влияние.

Исследование литературы, относящейся к предметной области [2-4], показало, что качество вина зависит от физико-химических показателей. Для выделения зависимостей между составом красного вина и оценкой его качества целесообразно использование нейросетевых технологий для автоматизации процесса и исключения человеческого фактора.

Целью данной работы является создание и обучение нейросетевой модели на основе набора данных о физико-химических параметрах красных вин. Для создания модели был взят набор данных с электронного ресурса [5].

Полученный набор данных содержит информацию о следующем наборе из 11 параметров:

- X1 – Фиксированная кислотность (4,6-11,5)
- X2 – Летучая кислотность (0,22-1,33)
- X3 – Лимонная кислота (0, 1)
- X4 – Остаточный сахар (1,2-10,7)
- X5 – Хлориды (0,045-0,61)
- X6 – Свободный диоксид серы (3-52)
- X7 – Общий диоксид серы (8-153)
- X8 – Плотность (0,9916-0,9996)
- X9 – pH (2,74-3,9)
- X10 – Сульфаты (0,33-2)
- X11 – Спирт (9-14)

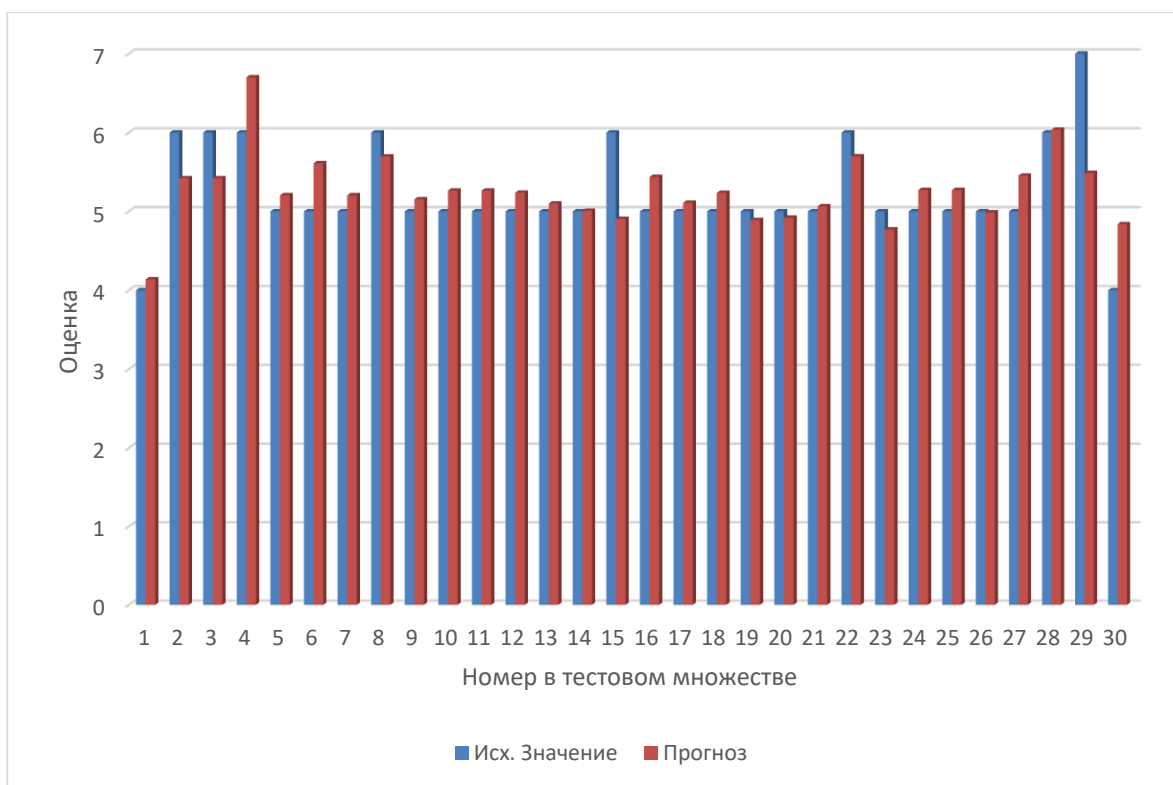
Для создания модели использовалась часть данных из набора для ускорения и стабильной работы нейросимулятора. Таким образом, вся выборка состоит из 200 записей о параметрах и оценках красных вин; обучающее множество содержит 130 записей (65%); валидирующее множество содержит 40 записей (~20%) и тестовое множество содержит 30 записей (~15%).

Итоговая оценка качества вина в наборе данных варьируется от 3 до 8. Перед началом проектирования модели была выполнена очистка исходного набора данных от выбросов и дубликатов.

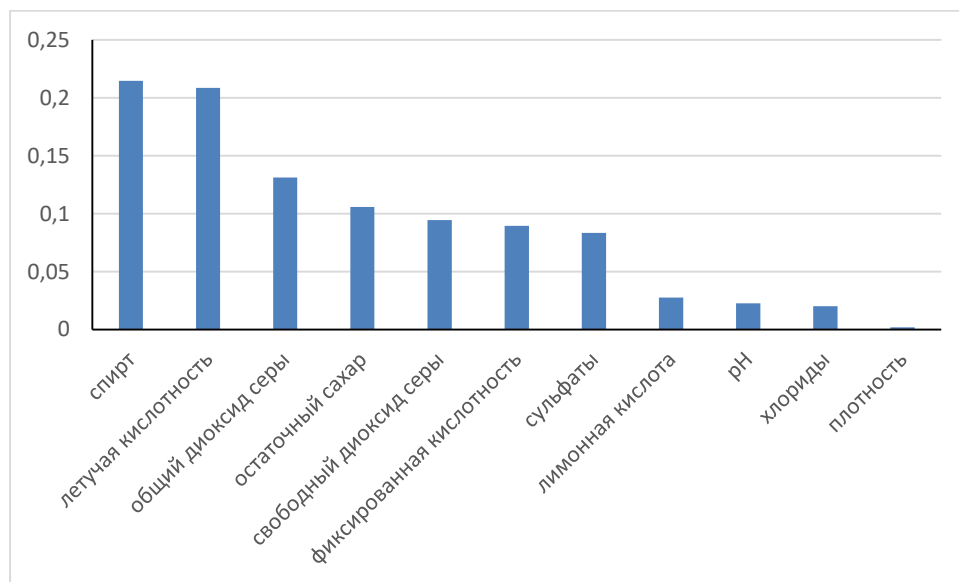
Проектирование, обучение и тестирование созданной нейронной сети проходило с использованием программы «Нейросимулятор 5» [6] по методике [7]. Полученная нейронная сеть представляет собой персептрон, обладающий 11 входными нейронами для каждого из параметров, 3 скрытыми слоями, состоящими из 8, 6 и 4 нейронов соответственно, и одним нейроном на выходном слое, предназначенном для итоговой оценки.

В результате обучения модели кв. ошибка обучения составила 1,27, а кв. ошибка обобщения 1,89, что можно считать хорошим результатом. На рисунке 1 приведена гистограмма, отображающая разницу между фактической и прогнозируемой оценкой качества красного вина для тестового множества.

По результатам, представленным на Рисунке 1, можно сделать вывод об адекватности спроектированной нейронной сети.



**Рисунок 1. Результат тестирования нейронной сети**



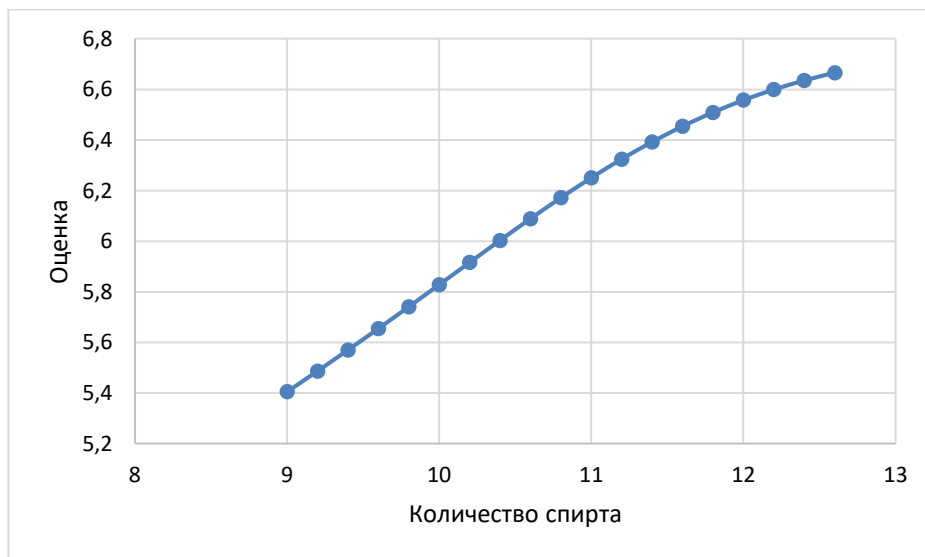
**Рисунок 2. Значимость входных параметров нейросетевой модели**

Далее с помощью приложения «Нейросимулятор 5» была выполнена оценка значимости входных параметров для обучающей выборки. Результат оценки приведен на Рисунке 2.

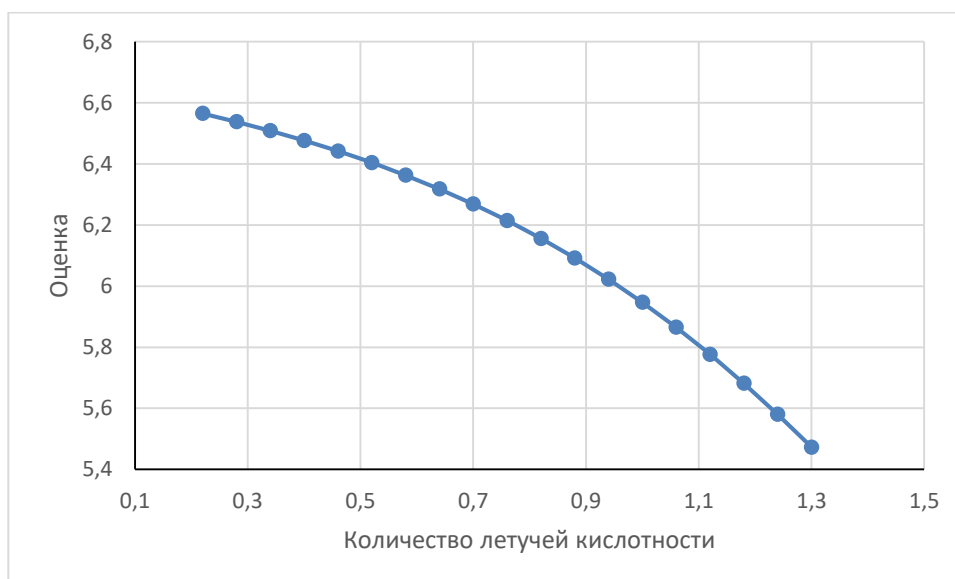
По результатам оценки значимости было выявлено, что наибольшее влияние оказывают 4 (ост. сахар), 6 (св. диоксид серы), 7 (общ. диоксид серы), 11 (спирт) и 2 (летучая кислотность) параметры.

На следующем этапе было проведено исследование с целью поиска зависимости между наиболее значимыми параметрами и качеством вина. Исследование производилось с помощью метода «замораживания» [7], суть которого заключается в изменении значения одного параметра (в нашем случае – наиболее значимых параметров) и фиксировании значений всех других параметров. Далее приведены графики, отражающие зависимость полученной оценки качества красного вина в соответствии с изменением каждого из 5 наиболее значимых параметров.

На рисунке 3 представлена зависимость оценки качества красного вина от количества спирта. Как видно по графику, с увеличением количества спирта в красном вине, увеличивается оценка его качества.



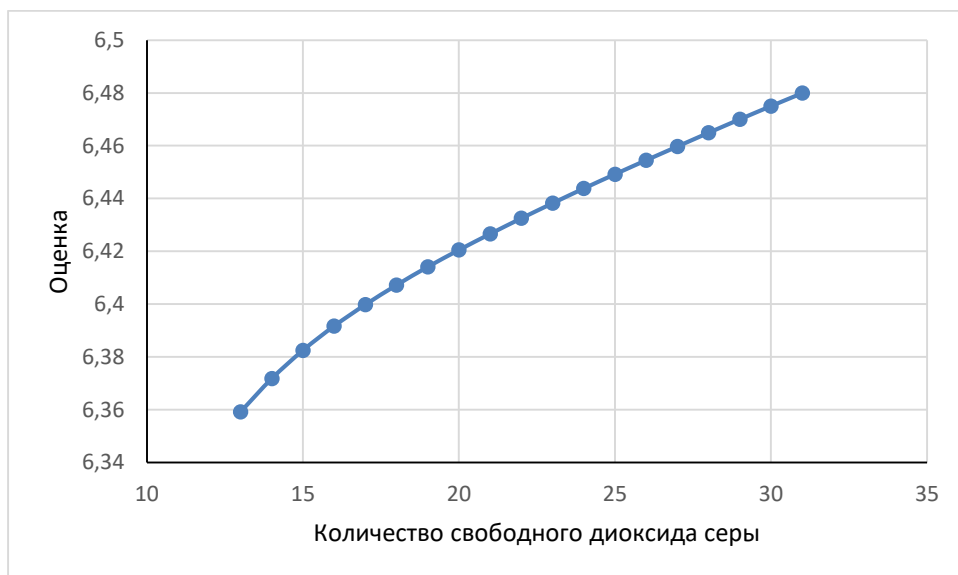
**Рисунок 3. Зависимость оценки от количества спирта**



**Рисунок 4. Зависимость оценки от количества летучей кислотности**

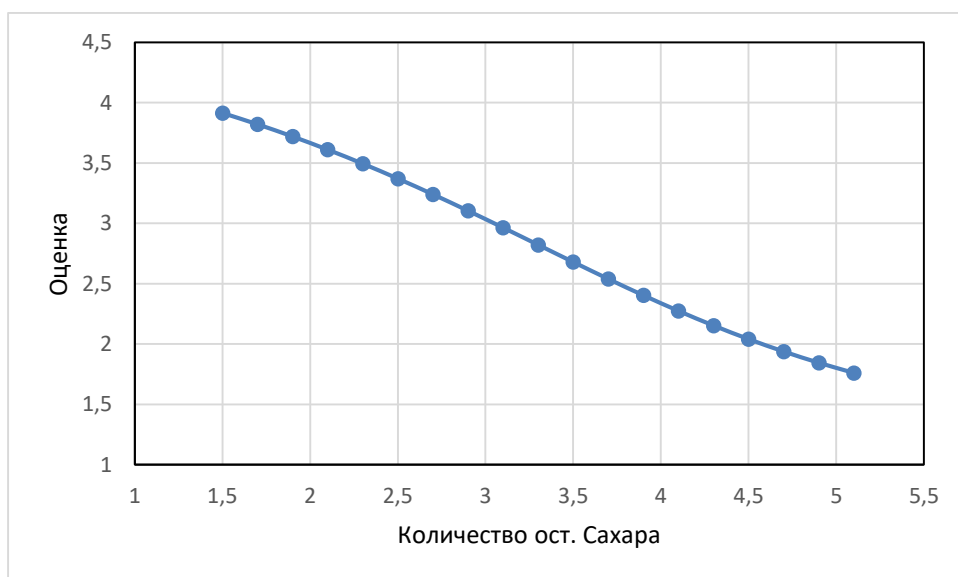
На рисунке 4 представлена зависимость оценки качества красного вина от количества летучей кислотности. Как видно по графику, с увеличением количества летучей кислотности в красном вине, уменьшается оценка его качества.

На рисунке 5 представлена зависимость оценки качества красного вина от количества свободного диоксида серы. Как видно по графику, с увеличением количества свободного диоксида серы в красном вине, увеличивается оценка его качества.



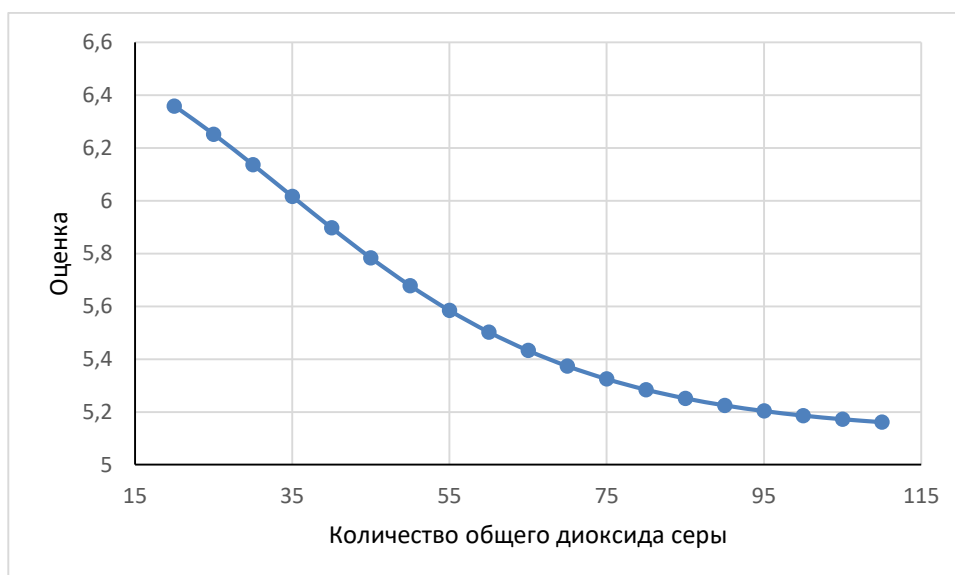
**Рисунок 5. Зависимость оценки от количества свободного диоксида серы**

На рисунке 6 представлена зависимость оценки качества красного вина от количества остаточного сахара. Как видно по графику, с увеличением количества остаточного сахара в красном вине, уменьшается оценка его качества.



**Рисунок 6. Зависимость оценки от количества остаточного сахара**

На рисунке 7 представлена зависимость оценки качества красного вина от количества общего диоксида серы. Как видно по графику, с увеличением количества общего диоксида серы в красном вине, уменьшается оценка его качества.



**Рисунок 7. Зависимость оценки от количества общего диоксида серы**

Исследование с помощью «Нейросимулятора» показало, что вина, обладающие большим количеством спирта и свободного диоксида серы, и меньшим количеством летучей кислотности, остаточного сахара и общего диоксида серы, обладают более высокой оценкой качества.

Далее был проведен анализ данных с использованием языка python. Для более справедливого сравнения решений на python и с использованием «Нейросимулятора» выборка для обучения модели на python сохранила процентное соотношение выборки для «Нейросимулятора».

На основе входных данных из набора [5] была построена и обучена модель. Код обучения модели приведен на рисунке 8.

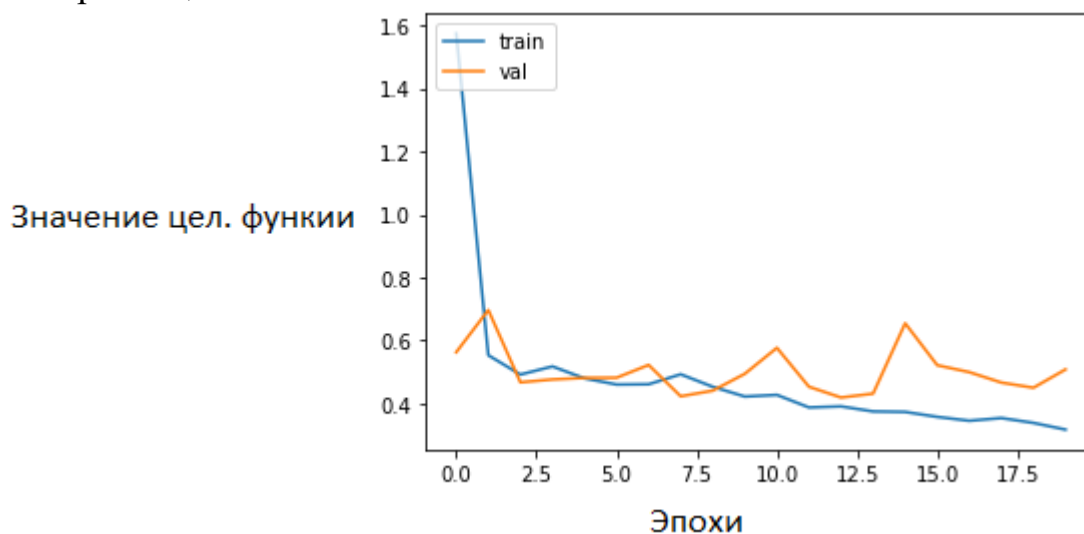
```
[27]: # Создаем модель
model = Sequential()
model.add(Dense(164, input_dim= 11, activation = 'relu'))
model.add(Dense(64, activation = 'relu'))
model.add(Dense(1))

[28]: # Создаем обработчик для сохранения лучшей
best_model_save = ModelCheckpoint('./model_weights.pt', save_best_only=True, monitor='val_mean_squared_error', mode='min')
# Создаем модель
model.compile(loss='mse', optimizer='sgd', metrics=['mse'])
# Обучаем модель
history = model.fit(X_train, y_train, validation_data = (X_val, y_val), epochs=20, callbacks=[best_model_save], batch_size=1, verbose=1)
```

**Рисунок 8. Код обучения модели**

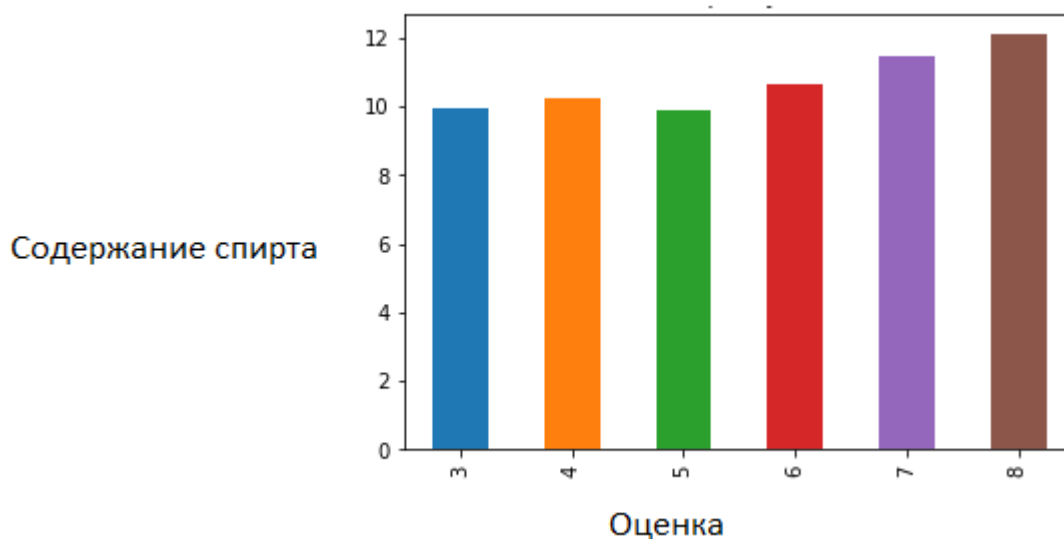
График обучения полученной модели приведен далее на рисунке 9. Полученное среднеквадратичное отклонение для данной модели на тестовой выборке составило 0.423, в то время как среднеквадратичное отклоне-

ние для «Нейросимулятора» составило 0,58, что является достаточно небольшой разницей.



**Рисунок 9. График обучения модели с использованием python**

Были построены графики, позволяющие отслеживать тенденции изменения оценки качества вина в зависимости от количества содержащихся в составе веществ. В целом тенденции, обнаруженные с помощью «Нейросимулятора» подтвердились. Так, исследование данных с использованием языка python показало, что чем больше спирта в вине, тем выше его оценка, данные подтверждаются графиком, представленным на рисунке 10.



**Рисунок 10. Влияние количества спирта на качество вина**

В результате проведенных исследований можно сделать вывод о том, что созданная с помощью «Нейросимулятора» адекватна для данной предметной области, и показывает результаты, сравнимые по точности с результатами, полученными с использованием более современных средств анализа данных.

**Заключение.** В ходе данной работы была построена система нейросетевого прогнозирования оценки качества красного вина в зависимости от физико-химических параметров состава вин. Были проведены исследования в рамках предметной области. В созданных условиях была доказана адекватная работа модели. В результате прогнозирования были выявлены зависимости между наиболее значимыми параметрами и оценкой качества красного вина.

### **Библиографический список**

1. Воронина Т. А. Заздравная чаша: к истории русского застолья (X–XVII века) // Хмельное и иное: Напитки народов мира. – 2008. – С. 72–101.
2. Павлова И. С., Пешкова К. Е., Дороганов В. С. Классификация и качественный анализ вин с использованием искусственных нейронных сетей // Информационно-телекоммуникационные системы и технологии. – 2015. – С. 201–201.
3. Сидорина С. А. Воронова Л.И. Применение методов интеллектуального анализа данных в задаче классификации экспертных оценок качества винных изделий. Научное обозрение. Педагогические науки. – 2019. – № 4 (часть 3) – С. 76–78.
4. Щелкунов А.А. Прогнозирование оценки качества вина методом нейросетевого моделирования [Электронный ресурс]. Пермь, 2018.
5. Открытые наборы данных для машинного обучения. URL: <https://www.kaggle.com/uciml/red-wine-quality-cortez-et-al-2009> (дата обращения: 20.01.2022).
6. Ясницкий Л.Н., Черепанов Ф. М. Искусственный интеллект: лабораторный практикум. URL: <http://www.lbai.ru/> (дата обращения: 20.03.2022).
7. Ясницкий Л.Н. Введение в искусственный интеллект. М.: Издательский центр «Академия», 2005. – 176 с.

## **NEURAL NETWORK SYSTEM FOR PREDICTING THE QUALITY OF RED WINE**

*Tarasova Valentina N.*

National Research University “Higher School of Economics”  
614070, Russia, Perm, st. Studentcheskaya, 38, valyatar146@yandex.ru

The article describes the development of a neural network system for predicting the quality of red wine. The system allows you to predict with great accuracy the assessment of the quality of red wine based on data on the chemical composition of the wine. With the help of the developed intellectual system, a study of the subject area was carried out, patterns of practical importance were identified.

**Keywords:** artificial intelligence, neural network technologies, forecasting, wine, quality assessment, wine quality.



## НЕЙРОСЕТЕВОЕ ПРОГНОЗИРОВАНИЕ СТОИМОСТИ КВАРТИР В ГОРОДЕ ПЕРМЬ

*Багин Максим Николаевич*

Пермский государственный национальный исследовательский университет,  
ПМИ. 614990, Россия, г. Пермь, ул. Букирева, 15, baginpmi3@gmail.com

В статье представлено описание разработки и применения нейросетевых технологий для прогнозирования стоимости квартир в городе Пермь. Система позволяет примерно оценить стоимость квартиры по некоторым показателям, таким как кол-во комнат, этаж, площадь, район, состояние квартиры, и др.

**Ключевые слова:** нейросетевые технологии, оценка, прогнозирование недвижимости, стоимость.

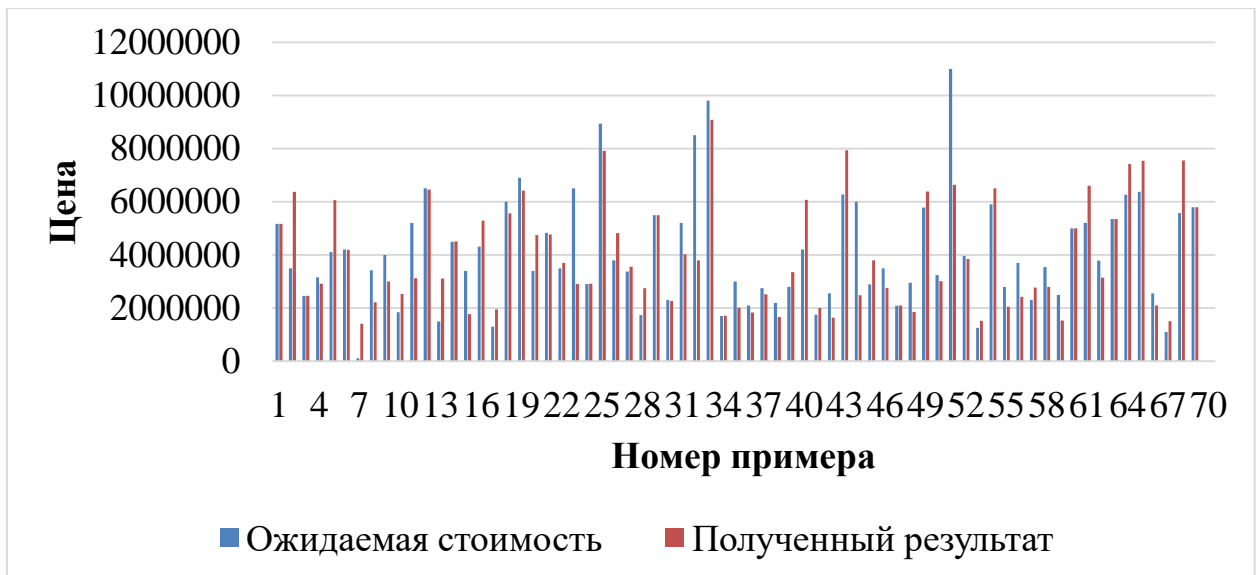
**Введение.** Недвижимость является довольно дорогим ресурсом и требует при ее выборе огромного внимания к каждой детали. Нейронные сети позволяют учитывать неявные факторы формирования стоимости, адаптироваться к специфике территориальных рынков недвижимости. Как показал опыт Пермской научной школы искусственного интеллекта, например, [1], правильно натренированная нейросетевая система может добиться успешности большей части прогнозов. При продаже или покупке недвижимости, следует выполнить глубокую оценку его стоимости. Обычно такие услуги предоставляют агентства по недвижимости или риэлторы, и зачастую эта оценка очень субъективна и может быть не точна.

Данная работа направлена помочь людям, которым требуется купить себе квартиру или оценить квартиру для ее продажи, но огромного количества времени для разбора этого вопроса они не имеют или не очень знакомы с ценами на недвижимость в Перми.

Входные параметры, с помощью которых прогнозируется стоимость квартиры: X1 – Район, X2 – Новостройка, X3 – Этаж, X4 – Количество комнат, X5 – Площадь, X6 – Состояние квартиры, X7 – Наличие балкона (лоджии), X8 – Наличие мебели, X9 – Санузел, X10 – Кухня, X11 – Курс доллара, D1 – ожидаемая стоимость квартиры.

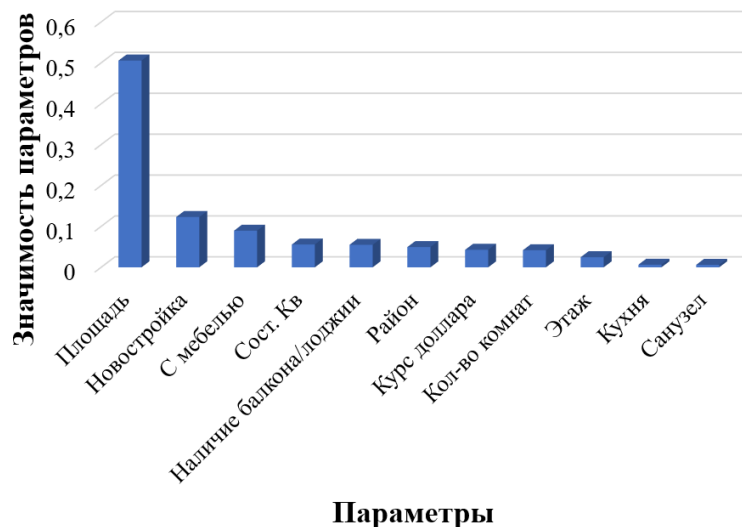
Обучающее множество было создано вручную с интернет-ресурсов [2].

Проектирование, обучение, тестирование нейронной сети выполнялись с помощью программы «Нейросимулятор 5» [3] по методике [4]. В результате оптимизации спроектированная нейронная сеть представляла собой персептрон, который имеет одиннадцать входных нейронов, один выходной и один скрытый слой с четырьмя нейронами.



**Рисунок 1. Результат тестирования нейронной сети**

Средняя относительная ошибка тестирования составила 8,439%, что можно считать приемлемым результатом. На рисунке 1 представлена гистограмма, демонстрирующая разницу между фактическим и прогнозируемым нейросетью ценой квартиры.



**Рисунок 2. Значимость параметров**

Как видно из рисунка 2, наиболее значимыми являются площадь, является ли дом новостройкой, продается с мебелью или нет, состояние квартиры и т.д. Как и ожидалось, наиболее влиятельным параметром является площадь.

Изучая гистограмму, можно заметить, что огромный вклад в цену квартиры вносит площадь, 50,5%. Также можно заметить, что важность очень мала таких параметров: этаж, количество комнат, наличие балкона/лоджии, наличие кухни и санузла. Курс доллара даже при изменении от 100 рублей до 120 внес вклад около 6%, скорее всего цены еще не изменились. А вот нали-

чие кухни и санузла зависит от площади квартиры, следовательно не сильно влияет на конечную цену.

Далее было проведено исследование жилищного рынка г. Пермь. Исследование производилось с помощью метода «замораживания» [1, 4], суть которого заключается в варьировании значения одного параметра и фиксации значений всех других параметров. Данный метод позволяет выявить влияние исследуемого параметра на выходной. Для этой цели было отобрано 2 квартиры, информация о параметрах которых отражена в таблице 1.

Таблица 1

### Характеристики квартир, выбранных для исследования

№ квартиры	1	2
X1	5	6
X2	2	2
X3	5	4
X4	1	2
X5	36,0	43,1
X6	2	1
X7	1	2
X8	2	1
X9	2	2
X10	2	2
X11	116,08	116,08

Для начала рассмотрим самый значимый параметр, и убедимся, что при его изменении значение выходного параметра сильно меняется. Рассмотрим график зависимости стоимости квартиры от параметра X15 – Площадь на рис 3.

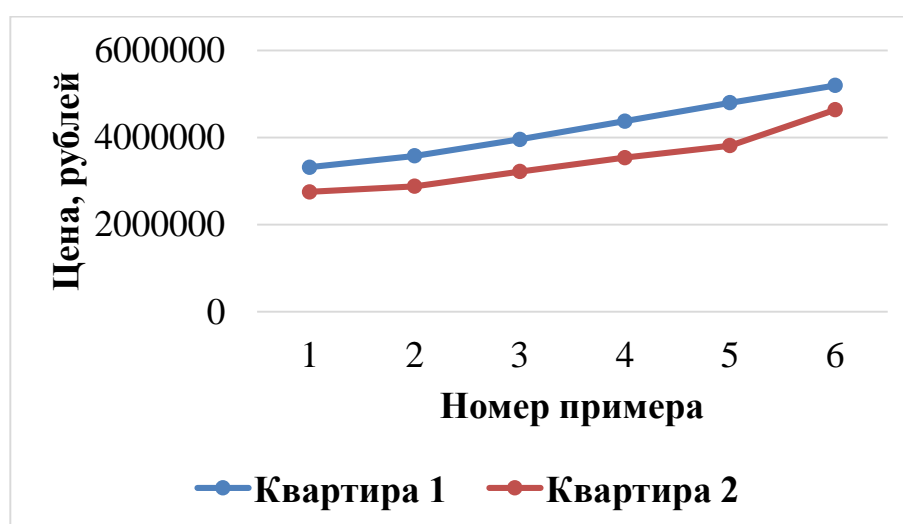
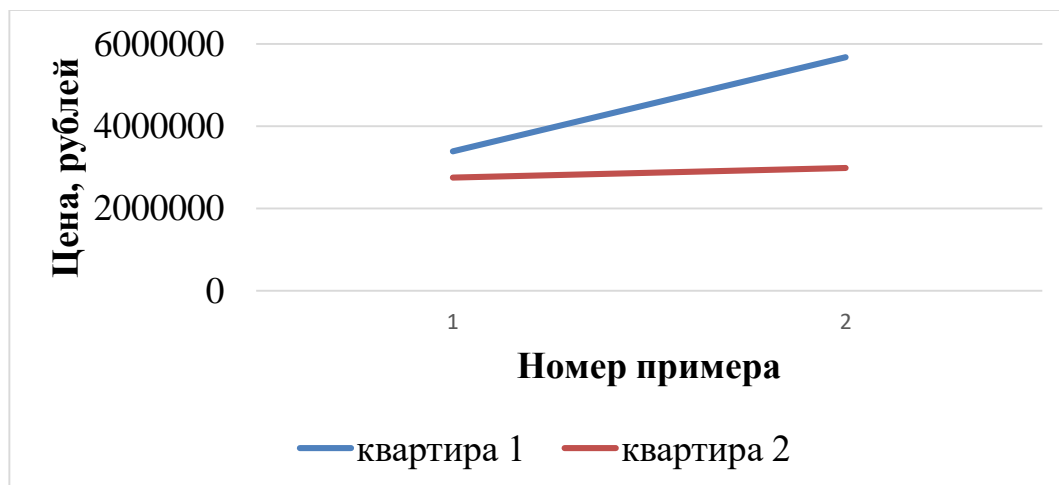


Рисунок 3. Зависимость стоимости квартир Перми от площади

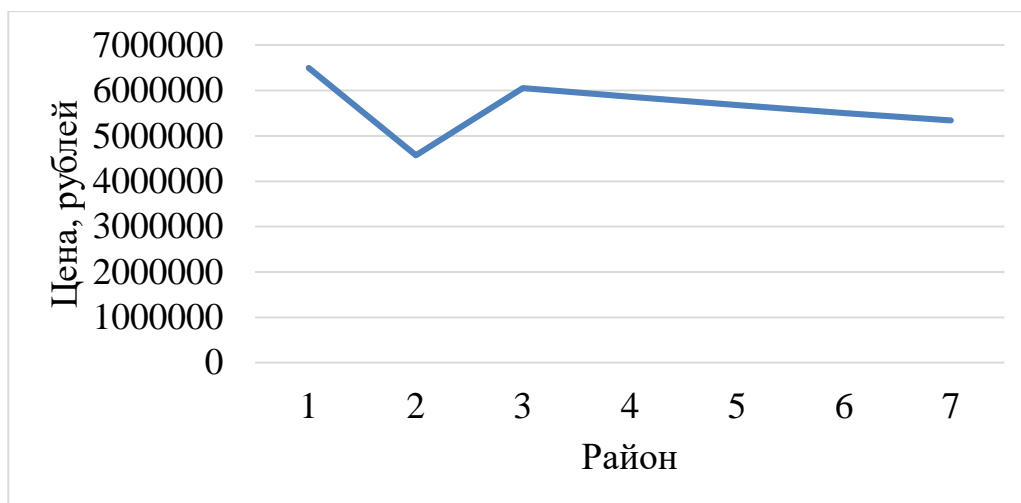
Так как уровень значимости параметра около 50%, увеличение площади даже на небольшое число, дает резкий рост в цене. Посчитав, можно определить, что 1 квадратный метр стоит около 80 тыс. рублей.

Рассмотрим чуть менее значимый параметр X2 – новостройка, он второй по значимости, его вклад 12,2%. Ниже построен зависимости от параметра X2.



**Рисунок 4. Зависимость от новостройки**

Из рисунка 4 видно, что в новостройке цены выше, что неудивительно. Но цена новостройки сильно зависит от района. Это можно увидеть на примере 1 квартиры на Рисунке 5.



**Рисунок 5. Зависимость цены новостройки от района**

Тут видно, что в данный момент цены квартир в новостройках в Дзержинском и Кировской районах сейчас самые высокие

Следующий параметр, который мы будем рассматривать, это X3 – Этаж, он является одним из самых незначимых параметров. Ниже приведен график зависимости стоимости от параметра этаж.

По рисунку 6 можно заметить, что цена квартиры на 1 этаже ниже, по сравнению с точно такой же квартирой на этаже повыше.

**Заключение.** Построена система нейросетевого прогнозирования стоимости квартир в городе Пермь. Спроектированная нейросетевая модель

учитывает 11 параметров: район, новостройка, этаж, количество комнат, площадь, состояние квартиры, наличие балкона (лоджии), наличие мебели, санузел, кухня, курс доллара. Методом сценарного прогнозирования построены графики зависимостей цен квартир от изменения входных параметров. Применение такого набора параметров в модели позволяет с 8,4% погрешностью определить цену какой-либо квартиры в Перми. Полученная модель работает адекватно и пригодна для дальнейшего использования.

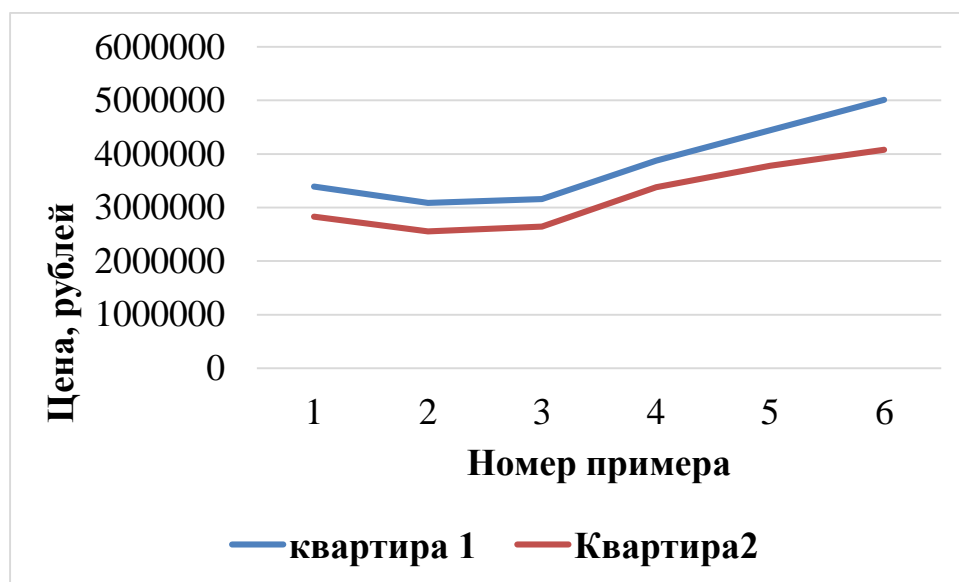


Рисунок 6. Зависимость от этажа

### Библиографический список

1. Ясницкий Л.Н., Ясницкий В.Л. Разработка и применение комплексных нейросетевых моделей массовой оценки и прогнозирования стоимости жилых объектов на примере рынков недвижимости Екатеринбурга и Перми // Имущественные отношения в Российской Федерации. 2017. № 3 (186). С. 68-84.
2. Режим доступа: <https://cyberleninka.ru/article/v/razrabotka-i-primenenie-kompleksnyh-neyrosetevyh-modeley-massovoy-otsenki-i-prognozirovaniya-stoimosti-zhilyh-obektov-na-primere-rynkov>.
3. avito.ru – Интернет-сервис для размещения объявлений о товарах, недвижимости и тд. [Электронный ресурс]. Режим доступа: <https://www.avito.ru/>
4. Черепанов Ф.М., Ясницкий Л.Н. Нейросимулятор 4.0. Свидетельство о регистрации программы для ЭВМ RUS 2014612546. Заявка № 2014610341 от 15.01.2014.
5. Ясницкий Л.Н. Введение в искусственный интеллект. М.: Издательский центр «Академия», 2005. – 176 с.

# NEURAL NETWORK FORECASTING OF THE COST OF APARTMENTS IN THE CITY OF PERM

*Bagin Maxim N.*

Perm State University

Str. Bukireva, 15, Perm, Russia, 614990, baginpmi3@gmail.com

The article describes the development and application of neural network technologies for predicting the cost of apartments in the city of Perm. The system allows you to roughly estimate the cost of an apartment by some indicators, such as the number of rooms, floor, area, area, condition of the apartment, etc.

**Keywords:** neural network technologies, valuation, forecasting real estate, cost.

УДК 004.032.26

## НЕЙРОСЕТЕВАЯ СИСТЕМА ОЦЕНКИ СТОИМОСТИ КВАДРАТНОГО МЕТРА ЖИЛОЙ НЕДВИЖИМОСТИ ГОРОДА ПЕРМИ

*Макшаков Антон Аркадьевич*

Пермский государственный национальный исследовательский университет,  
ПМИ. 614990, Россия, г. Пермь, ул. Букирева, 15, ruinmath@bk.ru

В статье представлено описание разработки нейросетевой системы для оценки стоимости квадратного метра жилья в городе Перми. Система позволяет с достаточной точностью оценить стоимость на основании ряда параметров, непосредственно связанных с недвижимостью и общей экономической ситуацией.

**Ключевые слова:** искусственный интеллект, нейросетевые технологии, прогнозирование, недвижимость, стоимость, оценка, Пермь.

**Введение.** Задача оценки стоимости недвижимости как никогда актуальна в наше время. Рынок жилья поддается прогнозированию с большим успехом, нежели наиболее волатильные виды рынков (например – акции). При грамотном составлении обучающего множества, обоснованном выборе параметров можно довольно точно предсказывать стоимость жилой недвижимости. Конечно, нужно делать поправку на то, что рынок подчиняется не только своим известным законам, возможны события, которые учитывать заранее не получится.

Сама задача прогнозирования стоимости жилья не нова, например, [1], но зачастую не учитывается специфика конкретного города, в котором необ-

ходимо провести исследование. Так, зачастую административное деление города на районы довольно специфично (центральный район города может включать в себя места с плохой транспортной доступностью), или количество объектов инфраструктуры (больниц, школ) уступает их качеству. Тем не менее, так как задача популярная, то мы будем ориентироваться на уже собранную информацию по данной теме.

Основная цель настоящей работы заключается в сборе множества объявлений о продаже недвижимости, на базе которой будет обучаться спроектированная нейросеть. Конечный результат – обученная нейросетевая система, способная адекватно оценить стоимость жилой недвижимости в городе Перми в нормальных рыночных условиях.

Для создания нейросетевой системы были выбраны следующие параметры, указанные в таблице 1.

**Таблица 1**

**Входные параметры нейросети**

Параметр	Назначение
X1	Жилплощадь
X2	Кол-во комнат
X3	Этаж
X4	Тип дома
X5	Отделка
X6	Год постройки
X7	Курс доллара
X8	Ключевая ставка ЦБ
X9	Удаленность от центра
X10	Район

Выходной параметр – стоимость квадратного метра жилья в рублях.

X4 – тип дома, наиболее часто встречающийся в жилом фонде, указанный в сопроводительной документации, где (1- панельный, 2- монолитный, 3- кирпичный, 4 – монолитно-кирпичный).

X5 – отделка определена как дискретная величина, где (1- черновая, 2 – чистовая, 3 – готовый ремонт).

X7 – курс доллара к рублю [2] и X8 – ключевая ставка ЦБ [3] учитывалась на момент выставления объявления в публичный доступ.

Значение параметра X9 рассчитывалось на основании транспортной доступности [4] конкретного дома на личном авто в субботу на момент 20:00 (для снижения влияния нештатных ситуаций) от здания правительства Пермского края (как конкретная точка центра города).

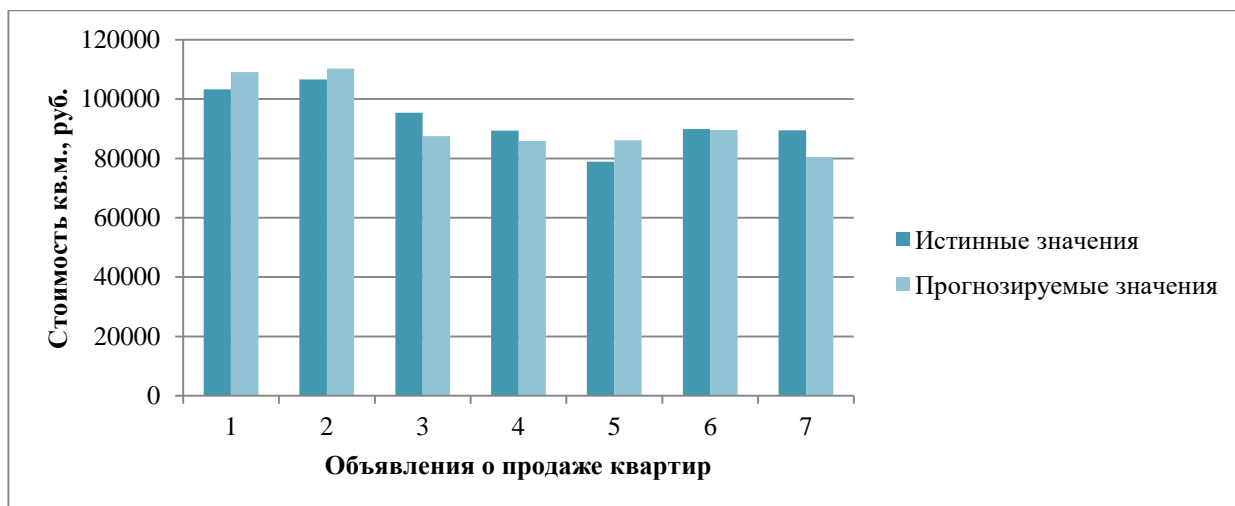
X10 – район определялся дискретной величиной, где 1- Дзержинский, 2 – Индустриальный, 3 – Кировский, 4 – Ленинский, 5 – Мотовилихинский, 6 – Орджоникидзевский, 7 – Свердловский.

Обучающее множество было собрано вручную с помощью анализа досок объявлений [5, 6]. Были отобраны случайным образом объявления в различных районах города за прошедший год. Перед обучением нейросетевой модели были проанализированы собранное множество на предмет противоречивых при-

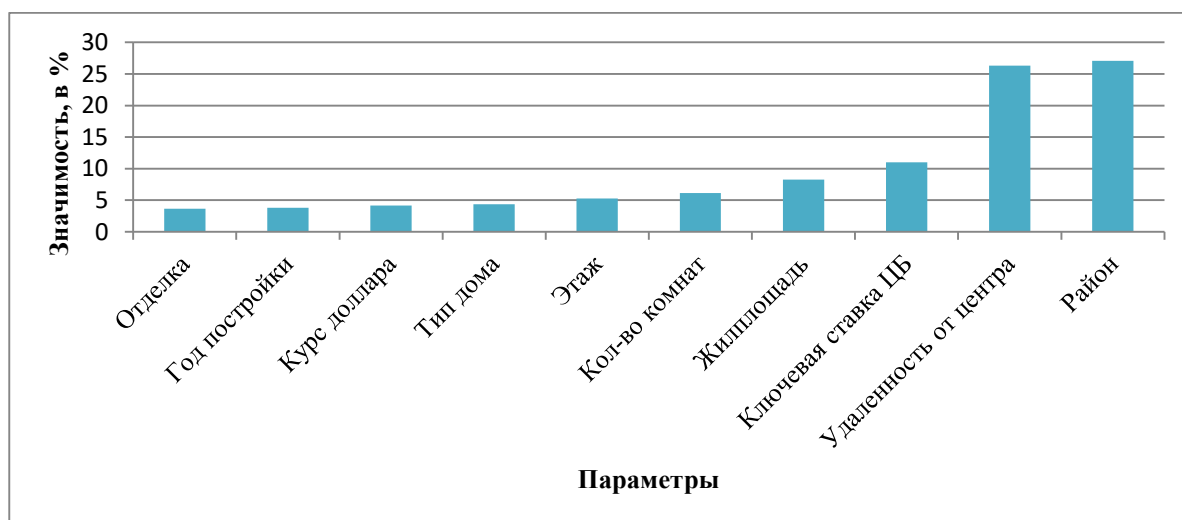
меров, нетипичных результатов (выбросов). Например, объявление могло не попасть в выборку, если продавец без оснований завышал (или занижал) стоимость, при наличии ограничений на переоформление квартиры, аварийного состояния жилья, сомнительной репутации продавца и так далее. Таким образом, были отобраны 46 примеров, среди которых большая часть (80%) использованы в качестве обучающего множества, а остальные – тестирующего.

Проектирование, обучение, тестирование нейронной сети выполнялось на базе программы «Нейросимулятор 5» [7] по методике [8]. В результате был спроектирован персептрон с десятью входными нейронами, одним скрытым слоем с одним нейроном, и один выходной нейрон. В качестве активационной функции была отобрана функция гиперболического тангенса.

Для оценки корректности итоговых результатов использовалось тестирующее множество из 7 примеров. Средняя относительная ошибка составила порядка 10%, что примерно соответствует действительности. На рисунке 1 представлена гистограмма, демонстрирующая разницу между фактическим и прогнозируемым нейросетью результатом.



**Рисунок 1. Результат тестирования нейросети**



**Рисунок 2. Значимость входных параметров**



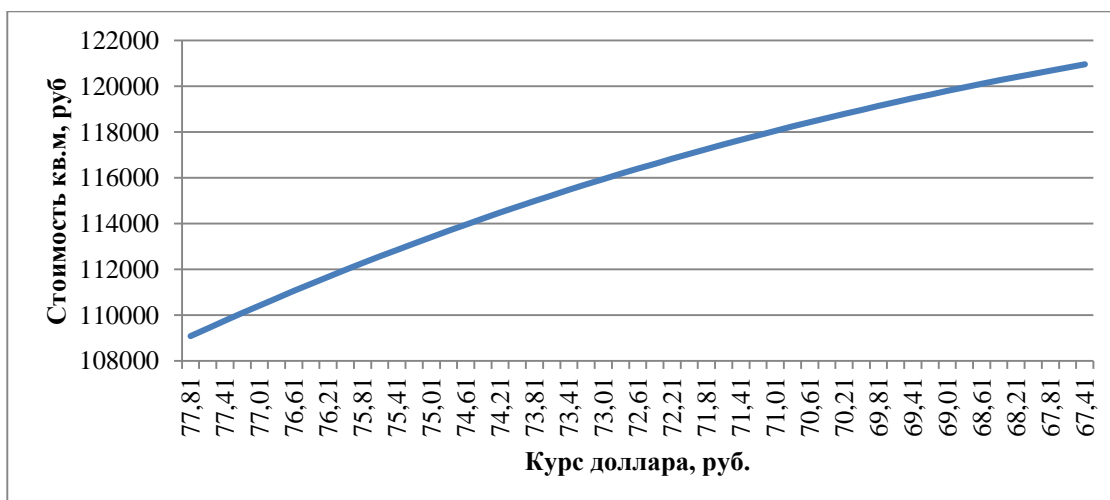
Оценка значимости входных параметров была получена с помощью стандартного инструментария программы, вклад каждой из них отображен на рисунке .

Как видно из диаграммы, наиболее значимыми являются параметры, связанные с территориальным положением квартиры (удаленностью от центра) и ключевой ставки ЦБ (как параметр, который косвенно способен продемонстрировать ситуацию на рынке). При низком значении ставки кредиты, в том числе и ипотечные кредиты, становятся более доступны для населения, как следствие растет и спрос на недвижимость. Район вносит ощутимый вклад только в связке с транспортной доступностью.

Далее было проведено исследование полученных зависимостей. Исследование производилось с помощью метода «замораживания» [9], суть которого заключается в варьировании значения одного параметра и фиксации значений всех других параметров. Для этого был взят пример из исходного множества и были изменены параметры курса валюты, ставки ЦБ, транспортной удаленности. Результаты представлены на рисунках 3, 4 и 5 соответственно.



**Рисунок 3. Зависимость от ставки ЦБ**



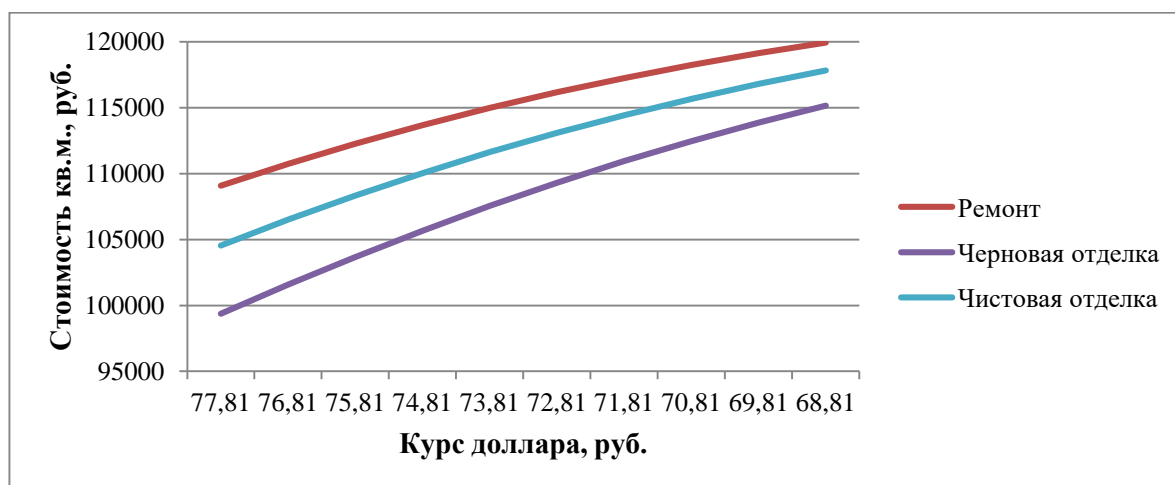
**Рисунок 4. Зависимость от курса доллара**



**Рисунок 5. Зависимость от удаленности**

Можно заключить несколько интересных закономерностей. Зависимость от ставки ЦБ сильная, это один из наиболее влиятельных параметров на стоимость – чем она выше, тем больше риски для выдачи кредитов банкам, а значит экономическая ситуация в неудовлетворительном состоянии (высокая инфляция) и цены на недвижимость (как, вообще говоря, и на множество других товаров) повышаются. При изменении курса доллара стоимость жилья в рублях колеблется довольно слабо, это объяснимо отсутствием прямой зависимости от внешней валюты, но чем курс ниже, тем больше свободных денег для покупки недвижимости, а значит, растет спрос, вместе с ним и цена. Транспортная доступность центра города влияет практически линейно на стоимость жилья, но в то же время это один из самых весомых параметров.

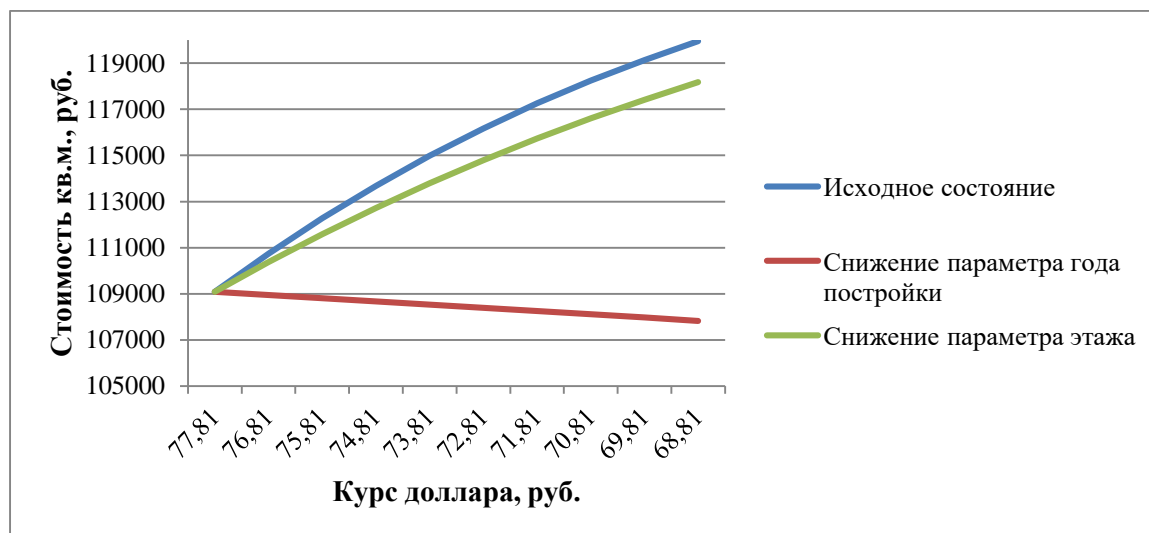
Также исследованы менее влиятельные параметры, например смена типа отделки, – на рисунке 6.



**Рисунок 6. Зависимость от смены типа отделки**

Вполне закономерно, что варианты без отделки потребуют вложений на ремонт, и их рыночная стоимость будет ниже. Вариант с чистовой отделкой (чаще всего от застройщика) занимает практически усредненную позицию.

Представляет интерес совмещенный график изменения параметров этажности (снижение с 24 до 6 этажа с шагом 2), года постройки дома (с исходного 2017 понижение до 1999 с шагом 2). Результат продемонстрирован на рисунке .



**Рисунок 7. Вариация различных параметров**

Как мы можем заметить, понижение этажности влияет крайне незначительно (как правило, стоит учитывать особенности конкретного дома, так как повышенная этажность может иметь как преимущества, так и недостатки, и наоборот, но данная система не предусматривает столь тонкой оценки). В противоположность можно изменить год постройки дома и получить сильную зависимость – более молодой жилой фонд ценится больше, нежели старый.

**Заключение.** Построена система нейросетевого прогнозирования стоимости квадратного метра жилья в городе Перми. Спроектированная нейросетевая модель учитывает 10 параметров: метраж, этаж, тип здания, год постройки, тип отделки, количество комнат, транспортную доступность, район дислоцирования, и ряд экономических показателей – ключевая ставка ЦБ и курс доллара. Разумеется, разумное увеличение количество параметров (например, добавление доступных инфраструктурных объектов), расширение обучающего множества может повысить точность оценки и применяться к более гибкому диапазону значений (например, адекватно оценивать старый жилой фонд, или квартиры в доме, который еще не сдан).

### Библиографический список

1. Realty.rbc.ru – новостной агрегатор по рынку недвижимости. [Электронный ресурс]. Режим доступа: <https://realty.rbc.ru/news/6185223c9a7947626cebabb7>

2. Banki.ru – динамика курсов валют. [Электронный ресурс]. Режим доступа: <https://www.banki.ru/products/currency/usd/>
3. Cbr.ru – сайт Центрального Банка России. [Электронный ресурс]. Режим доступа: [https://cbr.ru/hd\\_base/KeyRate/](https://cbr.ru/hd_base/KeyRate/)
4. Яндекс.Карты – ГИС с возможностью построить маршрут. [Электронный ресурс]. Режим доступа: <https://yandex.ru/maps>
5. Циан – портал с объявлениями о продаже недвижимости. [Электронный ресурс]. Режим доступа: <https://perm.cian.ru/>
6. Яндекс.Недвижимость – портал с архивом объявлений о продаже недвижимости. [Электронный ресурс]. Режим доступа: <https://realty.yandex.ru/otsenka-kvartiry-po-adresu-onlayn/>
7. Черепанов Ф.М., Ясницкий Л.Н. Нейросимулятор 4.0. Свидетельство о регистрации программы для ЭВМ RUS 2014612546. Заявка № 2014610341 от 15.01.2014.
8. Ясницкий Л.Н., Ясницкий В.Л. Разработка и применение комплексных нейросетевых моделей массовой оценки и прогнозирования стоимости жилых объектов на примере рынков недвижимости Екатеринбурга и Перми // Имущественные отношения в Российской Федерации. 2017. № 3 (186). С. 68-84.

## **NEURAL NETWORK SYSTEM FOR ESTIMATING THE COST PER SQUARE METER OF RESIDENTIAL REAL ESTATE IN PERM**

*Makshakov Anton A.*

Perm State University

Str. Bukireva, 15, Perm, Russia, 614990, ruinmath@bk.ru

The article describes the development of a neural network system for estimating the cost per square meter of housing in Perm. The system allows you to estimate the cost with sufficient accuracy based on a number of parameters directly related to real estate and the common economic situation.

**Keywords:** artificial intelligence, neural network technologies, forecasting, real estate, value, valuation, Perm.

## НЕЙРОСЕТЕВАЯ СИСТЕМА ПРОГНОЗИРОВАНИЯ СТОИМОСТИ ДОМОВ

*Тимофеев Артём Валерьевич*

*Ушакова Ольга Антоновна*

Пермский государственный национальный исследовательский университет,  
614990, Россия, г. Пермь, ул. Букирева, 15, smthneverchanges@vk.com

В статье представлено описание разработки нейросетевой системы для прогнозирования цены частного дома. Система позволяет предсказать примерную стоимость дома в зависимости от его характеристик. С помощью разработанной интеллектуальной системы проведено исследование предметной области, выявлены закономерности, имеющие практическое значение.

**Ключевые слова:** оценка стоимости, нейронная сеть, система прогнозирования.

**Введение.** Как владельцам, так и потенциальным владельцам недвижимости (речь пойдет именно о частных домах) в какой-то момент времени может потребоваться ее оценка. Поводы могут быть разные, в том числе оценка суммы, которую придется уплачивать в качестве налога, расчет кредитных платежей, решение различных имущественных споров (например, раздел совместно нажитого имущества при разводе).

Оценка жилого дома (оценка его рыночной стоимости) учитывает множество различных параметров и требует работы опытных экспертов, которая обычно является далеко не дешевой.

Рыночная стоимость – та, за которую с наибольшей вероятностью конкретный дом может быть продан на рынке в условиях конкуренции. При ее оценке не учитываются никакие чрезвычайные обстоятельства, как то: срочность продажи, наличие обременений или наложенных арестов на имущество. Зато других параметров, влияющих на рыночную стоимость жилого дома с участком, можно выделить массу. Кроме непосредственно площади дома (как жилой, так и площади участка и дополнительных строений, ему принадлежащих), может учитываться и его внутреннее устройство: планировка и дизайн, качество и новизна ремонта; расположение относительно объектов торговли, социальных и культурных объектов; инфраструктура местности, в которой построен дом и т.д.

Рынки недвижимости предоставляют возможность анализа и прогнозирования изменения цен частных домов в зависимости от их характеристик.

В статье [4] (Россия, 2017) опубликована разработка нейросетевых моделей для оценки и прогнозирования стоимости квартир (на примерах двух городов). Логично, что некоторые входные параметры при построении

нейросетевых моделей для оценки стоимости частных домов и квартир будут совпадать, но есть и индивидуальные показатели, применимые лишь к одному из исследований.

### Формулировка математической модели и ее тестирование

Для создания нейросетевой системы были выбраны следующие параметры: X1 – количество спальных комнат в доме, X2 – количество ванных комнат в доме, X3 – жилая площадь в м<sup>2</sup>, X4 – площадь участка в м<sup>2</sup>, X5 – количество этажей, X6 – состояние дома по шкале от 1 до 5, X7 – год постройки дома, X8 – год ремонта дома (если дом не ремонтировался, значение – 0). Выходной параметр D1 – цена дома в \$ США.

Обучающее множество было собрано из базы данных, найденной на платформе Kaggle[1]. Исходное множество включает в себя 2946 примеров, из которых для обучения было использовано 1769 примеров (~60%), для валидации использовалось 736 примеров (~25%), для тестирования – 441 пример (~15%).

Проектирование, обучение, тестирование нейронной сети выполнялись с помощью программы «Нейросимулятор 5» [2] по методике [3]. В результате оптимизации спроектированная нейронная сеть представляла собой персептрон, который имеет восемь входных нейронов и один выходной.

Средняя относительная ошибка тестирования составила 10,14%, что можно считать приемлемым результатом.

**Таблица 1**

#### Значения ошибок обучения нейронной сети

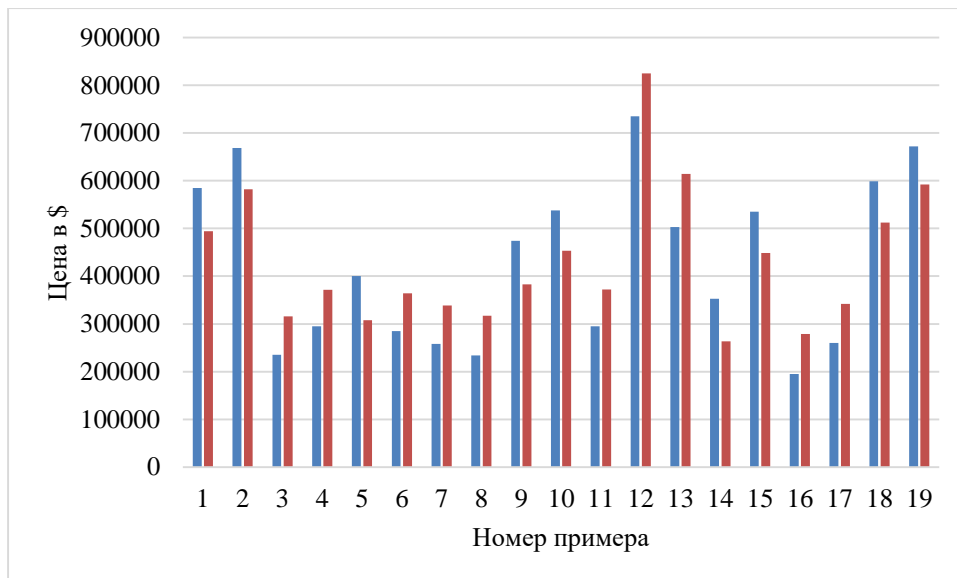
Ср. кв. отн. (%)	Средн. отн. погр.	Ср. кв. отн.(Т)(%)	Средн. отн. погр.
2,8869	2,4723	10,144	10,076

Ниже представлена гистограмма, иллюстрирующая результат тестирования нейронной сети. Вертикальная шкала соответствует стоимости дома, то есть показывает значение выходного параметра. Голубые столбцы соответствуют фактическим значениям цены дома, оранжевые – значениям, прогнозируемым нейросетью. Так как тестовых примеров большое количество, случайно взят отдельный фрагмент итоговых данных.

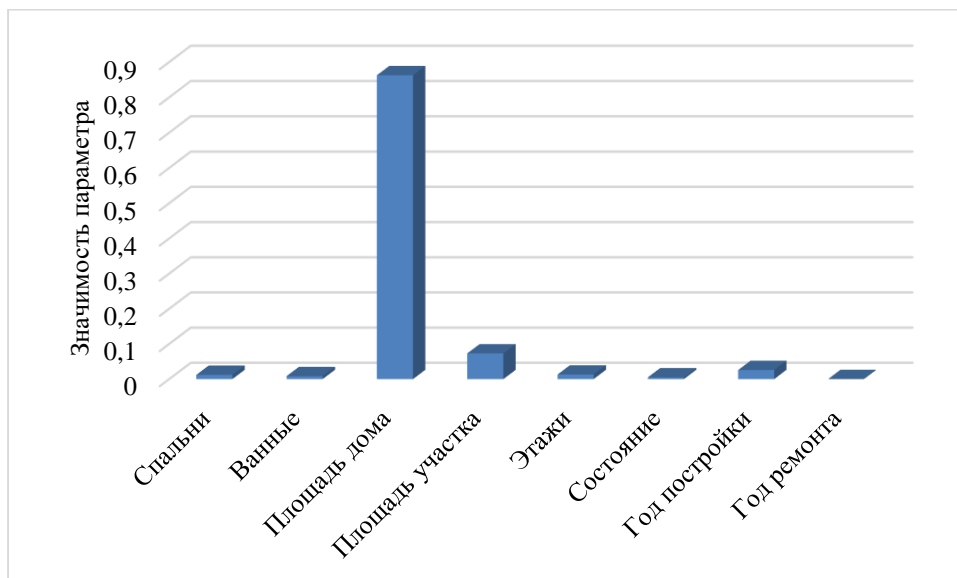
Видно, что разница между реальным значением, и тем, которое получено нейросетью, стабильно присутствует, но всё же значения достаточно близки. Особенно хорошо результат обучения сети виден на примере №11, так как среди данных примеров он сильно выделяется, но тем не менее нейросеть выдает значение, близкое к реальному.

После того, как нейросеть прошла проверку на тестовом множестве, и, как следствие, проверена адекватность построенной нейросетевой модели, можно перейти к ее анализу.

С помощью программы «Нейросимулятор 5» получены данные о значимости входных параметров X1-X8. Как видно, входной параметр X3 (жилая площадь в м<sup>2</sup>) является определяющим, что вполне соответствует естественной логике.



**Рисунок 1. Результаты тестирования нейронной сети**



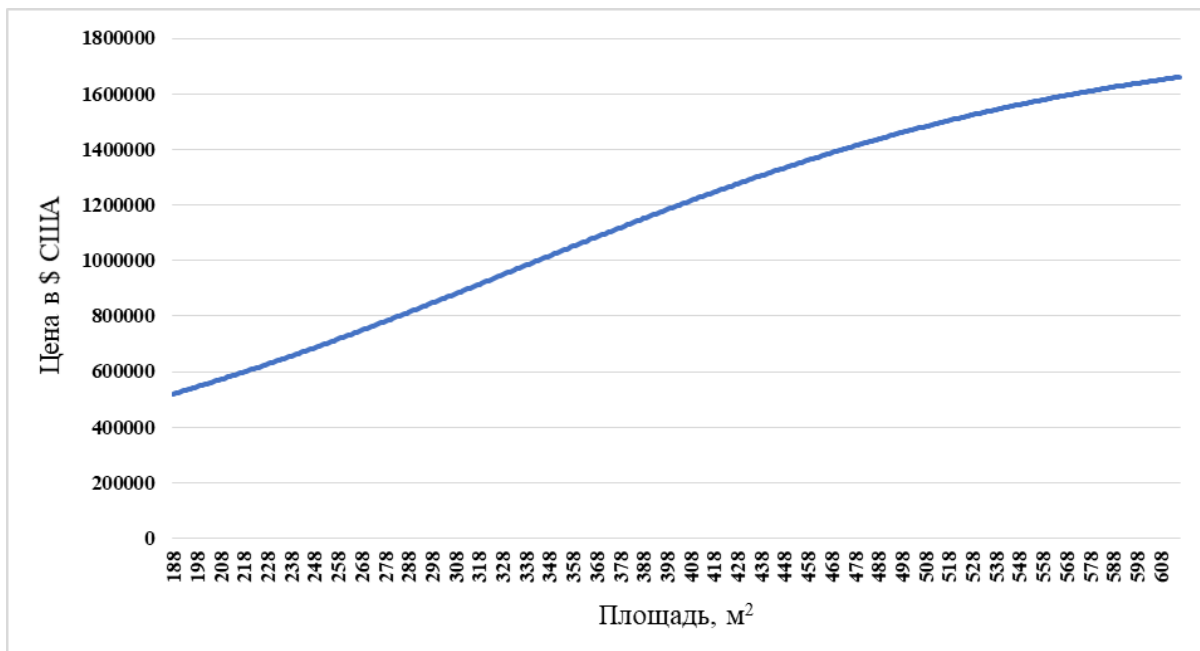
**Рисунок 2. Значимость параметров**

Для анализа сети взят некоторый пример дома со средним значением входного параметра  $X_3$  – жилой площади в  $m^2$ . Этот пример размножен с сохранением значений всех остальных входных параметров, а параметр  $X_3$  увеличен в каждой последующей копии примера. Нейросетью составлен прогноз стоимости таких домов – с отличающейся, возрастающей жилой площадью, но постоянными остальными характеристиками.

Построив график зависимости цены дома от увеличения жилой площади, можно увидеть, что цена монотонно возрастает с увеличением площади. Полученный результат соответствует реальности, следовательно, можно сделать вывод об адекватной работе нейросети и о целесообразности ее применения для оценки стоимости частных домов с участком.

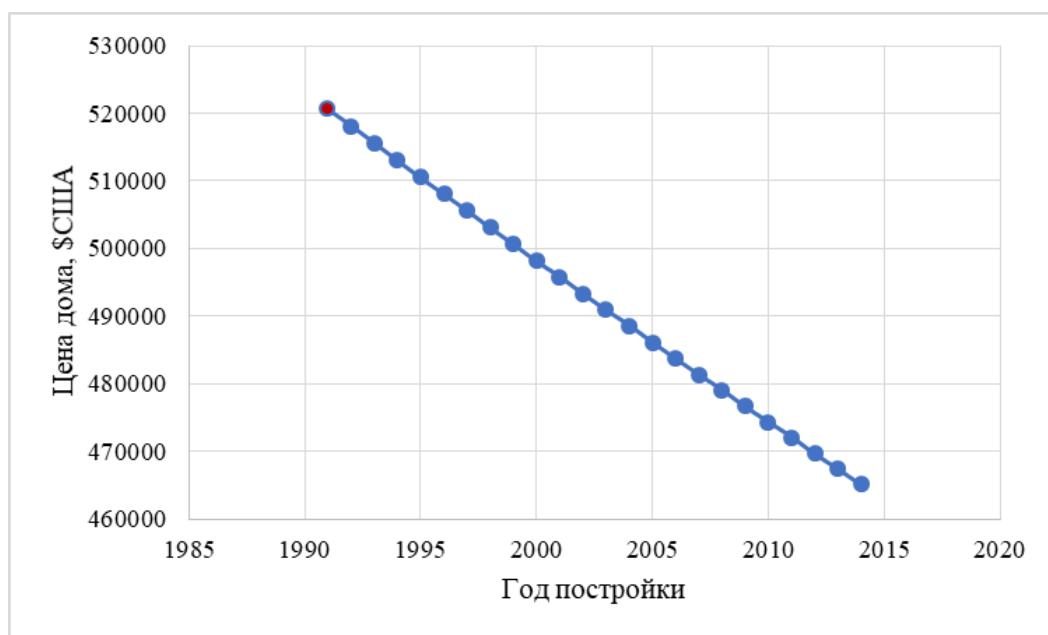
Для дальнейшего исследования взят тот же пример, но изменен параметр  $X_7$  – год постройки дома. Этот пример размножен с сохранением значе-

ний всех остальных входных параметров, а параметр  $X_7$  увеличен в каждой последующей копии.



**Рисунок 4. Зависимость цены от площади**

По результатам прогноза нейросети получилось, что чем раньше построен дом, тем он дороже (рис.5). Это может быть связано, например, с тем, что более старые дома могут быть построены из качественных строительных материалов, или же, что, приобретая дом постарше, вы приобретаете и многолетние труды предшествующих хозяев.



**Рисунок 5. Зависимость цены от года постройки**



## Заключение

Построена система нейросетевого прогнозирования цены дома с учетом таких параметров, как количество спальных комнат в доме, количество ванных комнат в доме, жилая площадь в м<sup>2</sup>, площадь участка в м<sup>2</sup>, количество этажей, состояние дома по шкале от 1 до 5, год постройки дома и год ремонта дома. Прогнозирование не обладает высокой точностью, но может дать примерную оценку стоимости. Разумеется, согласно исследованию предметной области, данный прогноз можно осуществлять, используя и другие параметры. Например, одним из достаточно важных показателей является то, в какой местности находится дом: это может быть столица либо отдаленный район; крупный город или деревня. В том числе и города могут отличаться своим географическим положением, или тем же престижем, и от этого цена на недвижимость также может варьироваться.

## Библиографический список

1. <https://www.kaggle.com/shree1992/housedata>
2. Черепанов Ф.М., Ясницкий Л.Н. Нейросимулятор 4.0. Свидетельство о регистрации программы для ЭВМ RUS 2014612546. Заявка № 2014610341 от 15.01.2014.
3. Ясницкий Л.Н. Введение в искусственный интеллект. М.: Издательский центр «Академия», 2005. – 176 с.
4. Ясницкий Л.Н., Ясницкий В.Л. Разработка и применение комплексных нейросетевых моделей массовой оценки и прогнозирования стоимости жилых объектов на примере рынков недвижимости Екатеринбурга и Перми // Имущественные отношения в Российской Федерации. 2017. № 3(186). С. 68-84.
5. Ясницкий Л.Н., Ясницкий В.Л. Методика создания комплексной экономико-математической модели массовой оценки стоимости объектов недвижимости на примере квартирного рынка города Перми // Вестник Пермского университета. Серия: Экономика. 2016. № 2 (29). С. 54-69.

## NEURAL NETWORK SYSTEM FOR PREDICTING THE COST OF HOUSES

*Timofeev Artem V., Ushakova Olga A.*

Perm State University, 15 Bukireva str., Perm,  
614990, Russia, smthneverchanges@vk.com

The article describes the development of a neural network system for predicting the price of a private house. The system allows you to predict the approximate cost of a house depending on its characteristics. With the help of the developed intellectual system, a study of the subject area was carried out, patterns of practical importance were identified.

**Keywords:** cost estimation, neural network, forecasting system.

## НЕЙРОСЕТЕВАЯ СИСТЕМА ПРОГНОЗИРОВАНИЯ ЦЕНЫ ЖИЛЬЯ ЗА КВАДРАТНЫЙ МЕТР

*Толочко Артём*

Национальный исследовательский университет «Высшая школа экономики»  
614070, Россия, г. Пермь, ул. Студенческая, 38, tolochko\_artem@bk.ru

Статья описывает разработку нейронной сети для прогнозирования цены за квадратный метр жилья, основываясь на некоторых его характеристиках. Система может быть использована риелторами и агрегаторами объявлений для выставления наиболее реальной цены при подаче объявления.

**Ключевые слова:** искусственный интеллект, нейронные сети, прогнозирование, недвижимость, жилье.

**Введение.** Для любого имущества, продающегося на вторичном рынке, не существует единой цены. Всё зависит от желания продавца. Однако есть способ объективизировать цену, если считать её производной функцией от некоторых характеристик самого жилья, а уже затем добавлять некоторую корректировку с учетом желания продавца или неформальных характеристик. Система, прогнозирующая цену за квадратный метр, основываясь на объективных показателях, смогла бы существенно упростить жизнь как риелторам, так и обычным покупателям и продавцам.

Пример такой системы – нейронная сеть. Её суть оптимизация некоторой функции большого числа переменных, то идеально подходит для условной задачи. Кроме того, практика Пермской школы искусственного интеллекта [1] показывает, что возможно обучить нейронную сеть таким образом, чтобы она давала точные прогнозы касательно цены на недвижимость.

При анализе литературы выявлено, что существуют исследования на тему прогнозирования цены на недвижимость. Все они описывают успех и получение достаточной точности прогнозирования. Например Э. Оливейра [2] в своей работе доказывает, что используя характеристики недвижимости для обучения нейронных сетей.

В качестве характеристик для нейронной системы в этой работе были выбраны следующие показатели:

X1 – возраст дома, лет

X2 – расстояние до ближайшей станции метро, м

X3 – количество магазинов поблизости, шт

Прогнозируемая переменная:

D1 – цена за квадратный метр, \$

Данные для построения нейронной сети были взяты с интернет ресурса Kaggle.com. Были выбраны 420 записей, обработаны, исключены пропуски.

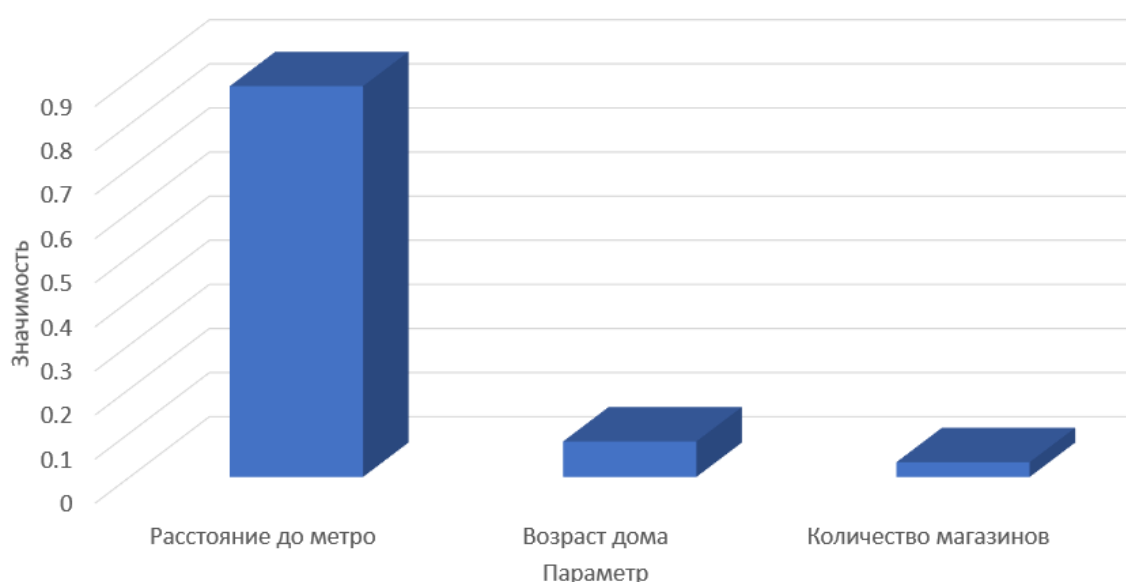
Полученный набор данных был разделен на выборки: обучающая – 340, валидирующая – 40, тестовая – 40.

Нейронная сеть была построена с помощью программы «Нейросимулятор 5» [3] по методике [4]. Полученная сеть имеет 3 входных нейронов, 2 скрытых слоя из 3 и 1 нейронов и 1 выходной нейрон.



**Рисунок 1. Результат работы нейросети**

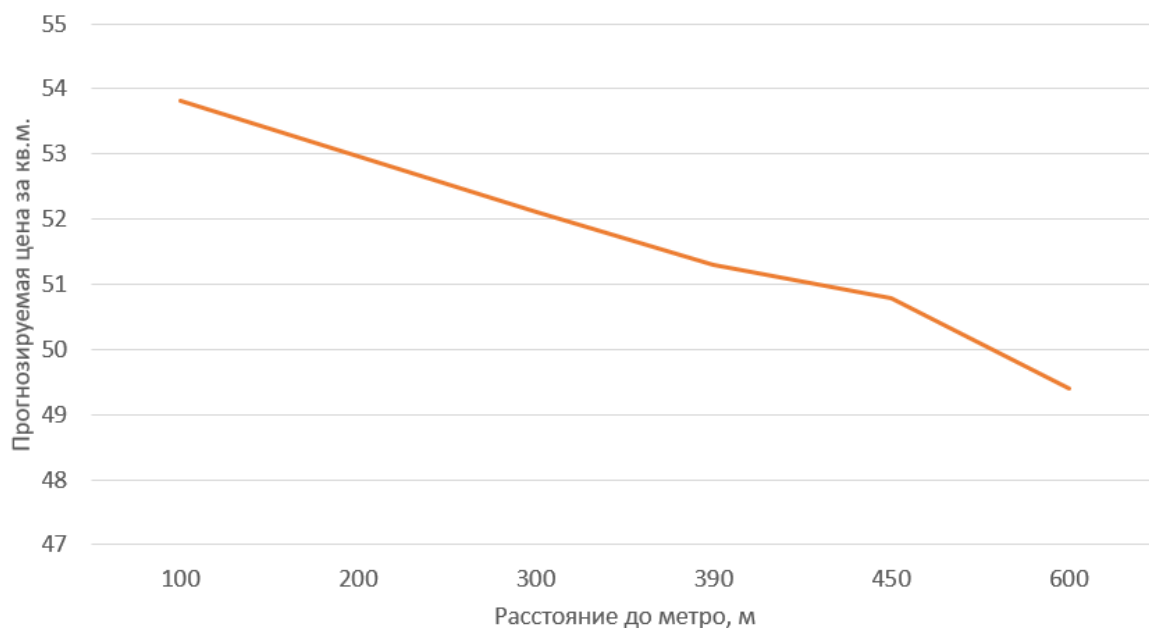
На тестовой выборке система показала отличный результат, средняя относительная ошибка составила 9%, что является хорошим показателем. Исходя из результатов проверки на тестовой выборке, можно сказать, что нейронная сеть обучилась и может предсказывать. Для достоверности исследуем значимость параметров.



**Рисунок 2. Значимость параметров**

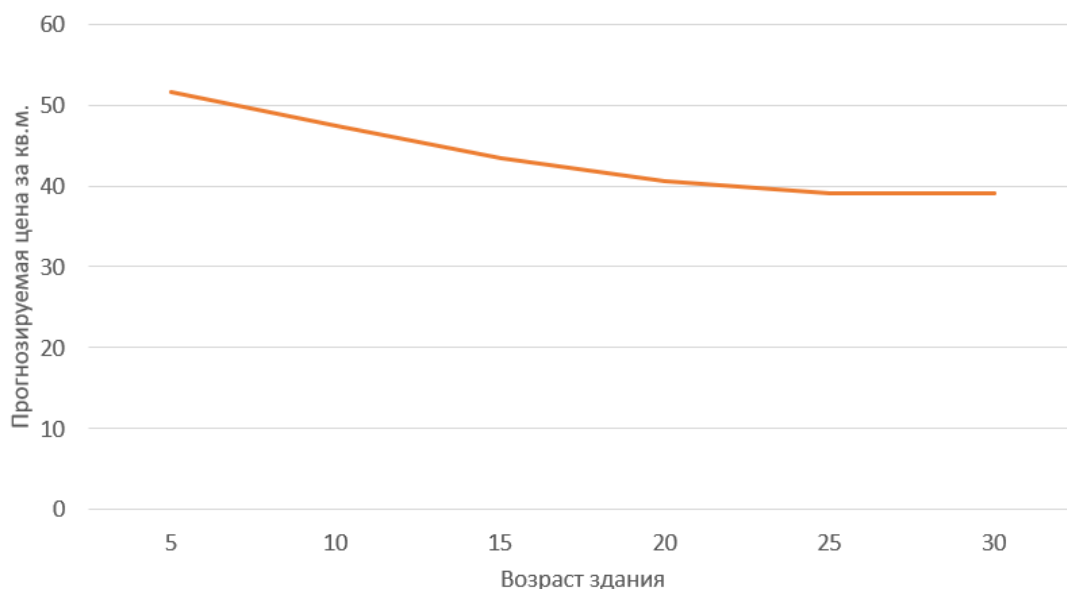
На рисунке 2 изображена значимость параметров. Видно, что наиболее сильно на результат влияет характеристика X2, «Расстояние до метро».

Остальные два параметра также имеют влияние, но куда меньшее, до 20%. Исследуем влияние каждого параметра на целевую переменную.



**Рисунок 3. Влияние расстояние до метро на прогнозируемую цену**

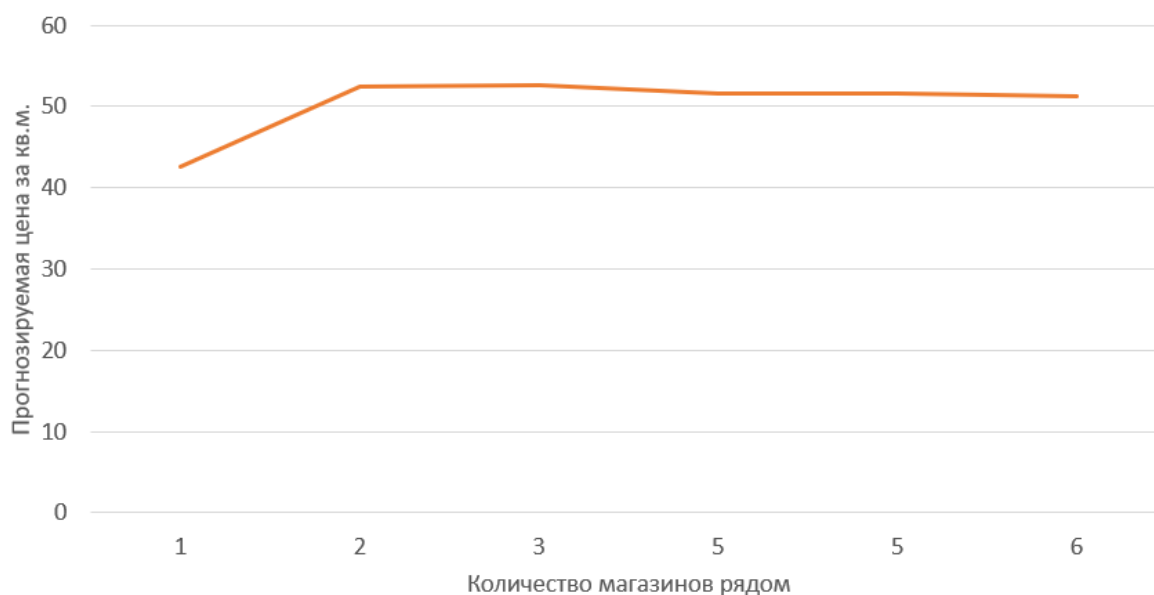
На рисунке 3 показан график зависимости между расстояние до метро и прогнозируемой ценой за квадратный метр. Наблюдается отчетливый отрицательный тренд, означающий, что жилье дальше от метро стоит дешевле. Данный показатель согласуется с реальностью, люди предпочитают жить ближе к транспортным узлам.



**Рисунок 4. Влияние возраста постройки на прогнозируемую цену**

На рисунке 4 представлена зависимость возраста жилья от прогнозируемой цены за квадратный метр. Также, как и на предыдущем графике, виден снижающийся, хотя и менее крутой, тренд. Это говорит о том, что цена на

квартиру в более современном доме выше, чем цена на такой же по характеристикам дом, но более старый.



**Рисунок 5. Влияние кол-ва магазинов рядом на прогнозируемую цену**

На рисунке 5 представлен график зависимости между количеством магазинов рядом с жильем и прогнозируемой ценой за квадратный метр. Интересно, что значение цены показывает рост в промежутке от 1 до 2х магазинов, а затем выходит на своеобразное плато. Это может быть объяснено тем, что людям в жизни достаточно 1-2 магазинов, в которых можно купить всё необходимое. Остальные не приносят добавочной ценности.

Разработка средствами keras на языке питон оказалась весьма трудозатратой задачей. Необходимо изучить большое количество модулей и библиотек, а также их взаимную работу. Приходилось искать значения каждой функции и процедуры. Однако, видно, что библиотека keras это очень мощный инструмент для работы с нейронными системами и данными. Нейросимулятор в свою очередь также является мощным инструментом и отличным программным продуктом, который способен выполнять практически те же функции, однако более дружелюбен к пользователю за счёт интерфейса, удобного импорта данных, автоматических визуализаций обучения и проектирования. Для меня именно нейросимулятор оказался удобнее

**Заключение.** Была построена нейронная система, успешно предсказывающая цену за квадратный метр жилья, на основе его характеристик. Система показала хорошие метрики качества. В дальнейшем её применение возможно в профессиональных областях торговли вторичной недвижимостью, для экономии времени риэлторов и клиентов.

### **Библиографический список**

1. Ясницкий В.Л. Алексеев А.О., Ясницкий Л.Н. Массовая оценка и сценарное прогнозирование рыночной стоимости городской недвижимости

на основе технологии нейросетевого моделирования: монография. – Москва: РУСАЙНС, 2019. – 112 с.

2. Эдгар Оливейра. Оценка недвижимости с использованием искусственных нейронных сетей-систематический обзор литературы. Междисциплинарный научный журнал Core знаний. 04 год, Эд. 04, том 05, стр. 55-75 Апрель 2019. ISSN: 2448-0959

3. Черепанов Ф.М., Ясницкий Л.Н. Нейросимулятор 4.0. Свидетельство о регистрации программы для ЭВМ RUS 2014612546. Заявка № 2014610341 от 15.01.2014.

4. Ясницкий Л.Н., Ясницкий В.Л. Разработка и применение комплексных нейросетевых моделей массовой оценки и прогнозирования стоимости жилых объектов на примере рынков недвижимости Екатеринбурга и Перми // Имущественные отношения в Российской Федерации. 2017. № 3 (186). С. 68-84.

## NEURAL NETWORK SYSTEM FOR FORECASTING THE PRICE OF HOUSING PER SQUARE METER

*Tolochko Artem*

National Research University “Higher School of Economics”  
614070, Russia, Perm, st. Studentcheskaya, 38, tolochko\_artem@bk.ru

The article describes the development of a neural network to predict the price per square meter of housing, based on some of its characteristics. The system can be used by realtors and ad aggregators to set the most realistic price when submitting an ad.

**Keywords:** artificial intelligence, neural networks, forecasting, real estate, housing.

УДК 004.311

## НЕЙРОСЕТЕВАЯ СИСТЕМА ПРОГНОЗИРОВАНИЯ СРЕДНЕЙ ЗАРАБОТНОЙ ПЛАТЫ

*Каменских Любовь Андреевна*

Пермский государственный национальный исследовательский университет,  
ФИТ. 614990, Россия, г. Пермь, ул. Букирева, 15, lyubochka199@yandex.ru

В статье представлено описание разработки нейросетевой системы для прогнозирования результатов средней заработной платы. Система позволяет с большой точностью предсказать среднюю заработную плату в зависимости от года начислений, пола, сферы труда и образования. Нейросеть позволяет спрогнозировать динамику роста зарплат.

**Ключевые слова:** искусственный интеллект, нейросетевые технологии, прогнозирование, средняя зарплата, прогноз, зарплата.

**Введение.** В связи с непростой сложившейся геополитической, эпидемиологической и экономической ситуацией в России и мире, одной из самых актуальных проблем становится тема заработной платы, прожиточного минимума населения, ВВП, уровня занятости и безработицы.

Уровень заработной платы играет важную роль в выборе профессии, что влечет за собой получение соответствующего образования исходя из желаний, потребностей, ведь каждый человек хочет жить комфортно и благополучно.

При анализе официальных статистических данных выяснилось, что главными показателями уровня заработной платы являются: пол, сфера труда, образование.

Основная цель настоящей работы заключается в анализе, консолидации статистических данных об уровне заработной платы, а также создание и обучение нейросетевой модели на этих данных. Конечным результатом должна быть нейросетевая система, способная прогнозировать уровень заработной платы в зависимости от выбранных параметров.

Для создания нейросетевой системы были выбраны следующие параметры: X1 – сфера деятельности, X2 – год, X3 – пол, X4 – образование.

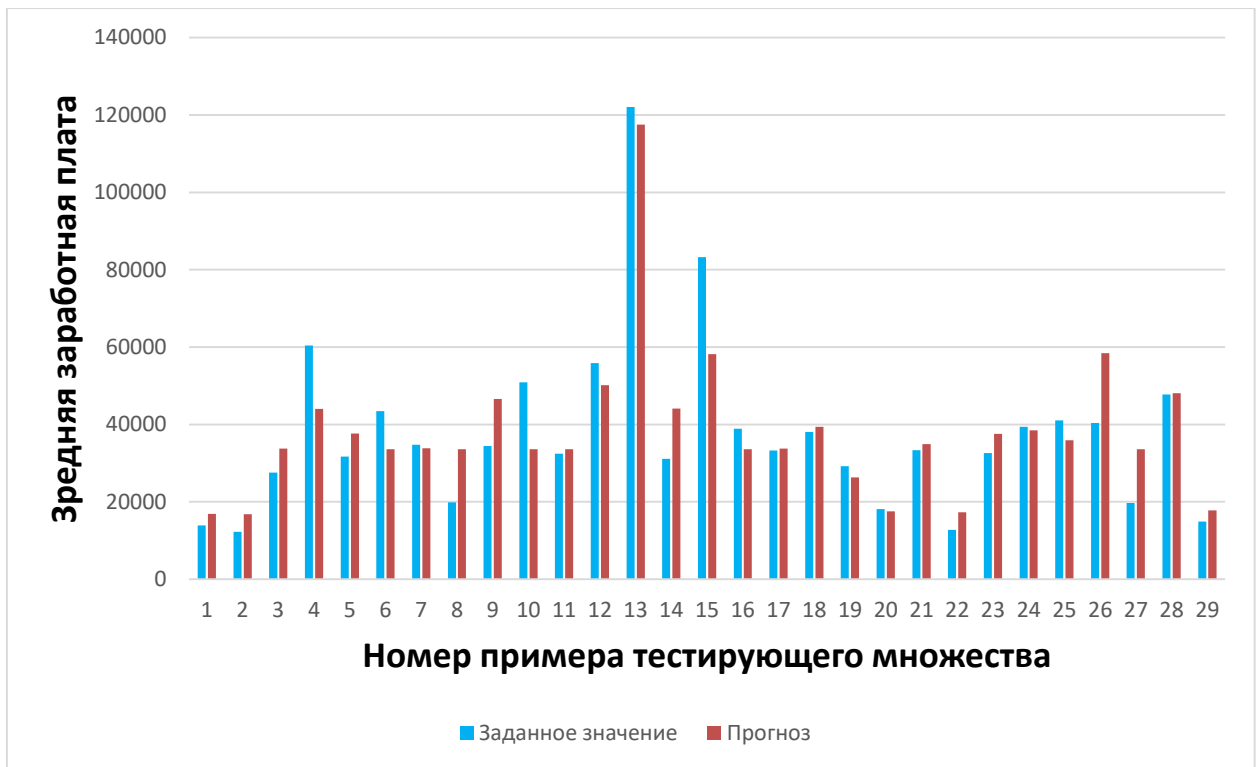
Весь объем данных был взят с официального источника государственной статистики [1]. Перед переходом к проектированию и обучению нейросетевой модели, была выполнена очистка исходного множества от противоречивых примеров, выбросов, дубликатов. Чистка проводилась параллельно со сбором данных (т. к. сбор осуществлялся вручную, то исключались сомнительные данные или данные, не имеющие полный набор данных по тому или иному параметру. После сбора данных они были подгружены в «Нейросимулятор 5» [2] и проверены его алгоритмом на выбросы с помощью методики [4]. Таким образом, объем итогового множества включает в себя 284 примера. Данное множество было разделено на обучающее и тестирующее в соотношении 90% и 10% соответственно.

Проектирование, обучение, тестирование нейронной сети выполнялись с помощью программы «Нейросимулятор 5» [2] по методике [3]. В результате оптимизации спроектированная нейронная сеть представляла собой персептрон, с функцией активации «Сигмоида», который имеет четыре входных нейрона, один выходной, а также два скрытых слоя с функцией активации «Тангенс Гиперболический» с двумя нейронами и одним соответственно.

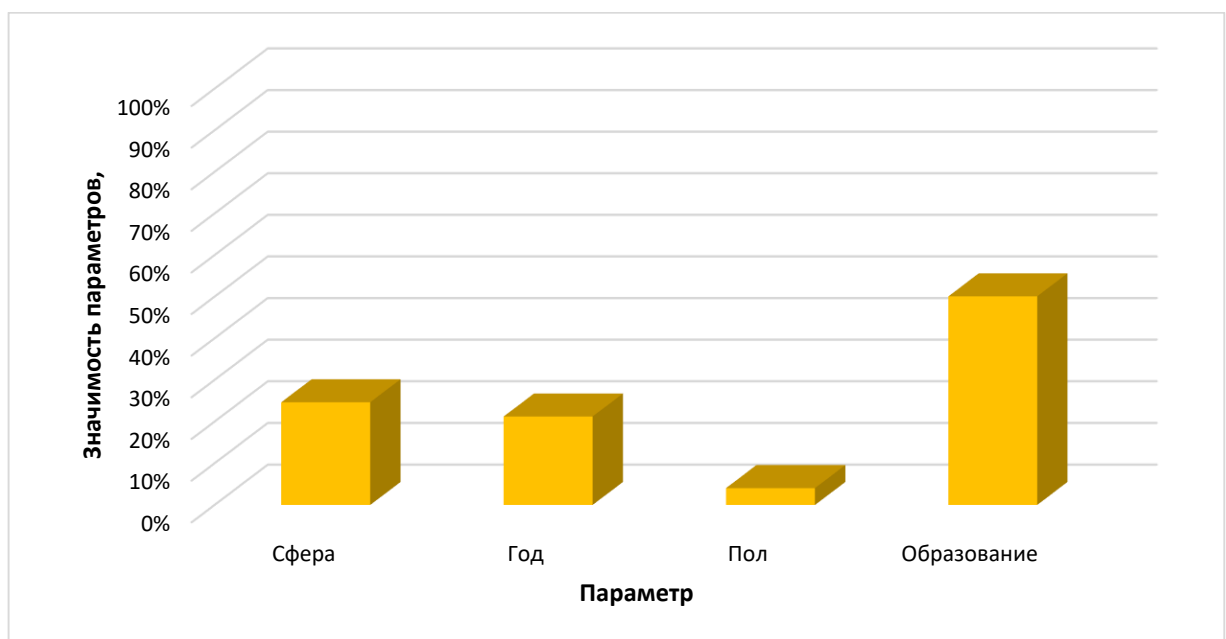
Для оценки корректной работы спроектированной нейронной сети использовалось тестирующее множество, состоящее из 29 примеров. Средняя относительная ошибка тестирования составила 8.6%, что можно считать приемлемым результатом. На рисунке 1 представлена гистограмма, демонстрирующая разницу между фактической и прогнозируемой нейросетью заработной платой.

Из результатов, изображенных на рисунке 1, можно сделать вывод об адекватной работе спроектированной нейронной сети.

Оценка значимости входных параметров была выполнена с помощью программы «Нейросимулятор 5», результат отображен на рисунке 2.



**Рисунок 1. Результат тестирования нейронной сети**

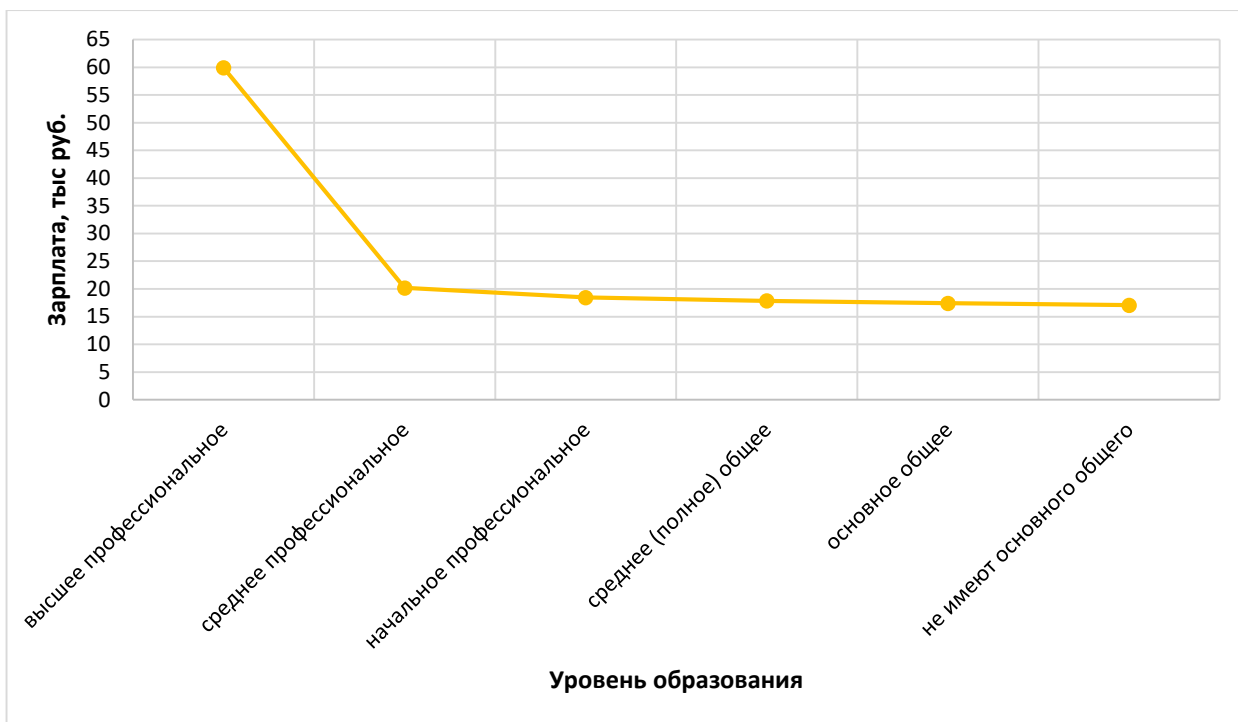


**Рисунок 2. Значимость входных параметров нейросетевой модели**

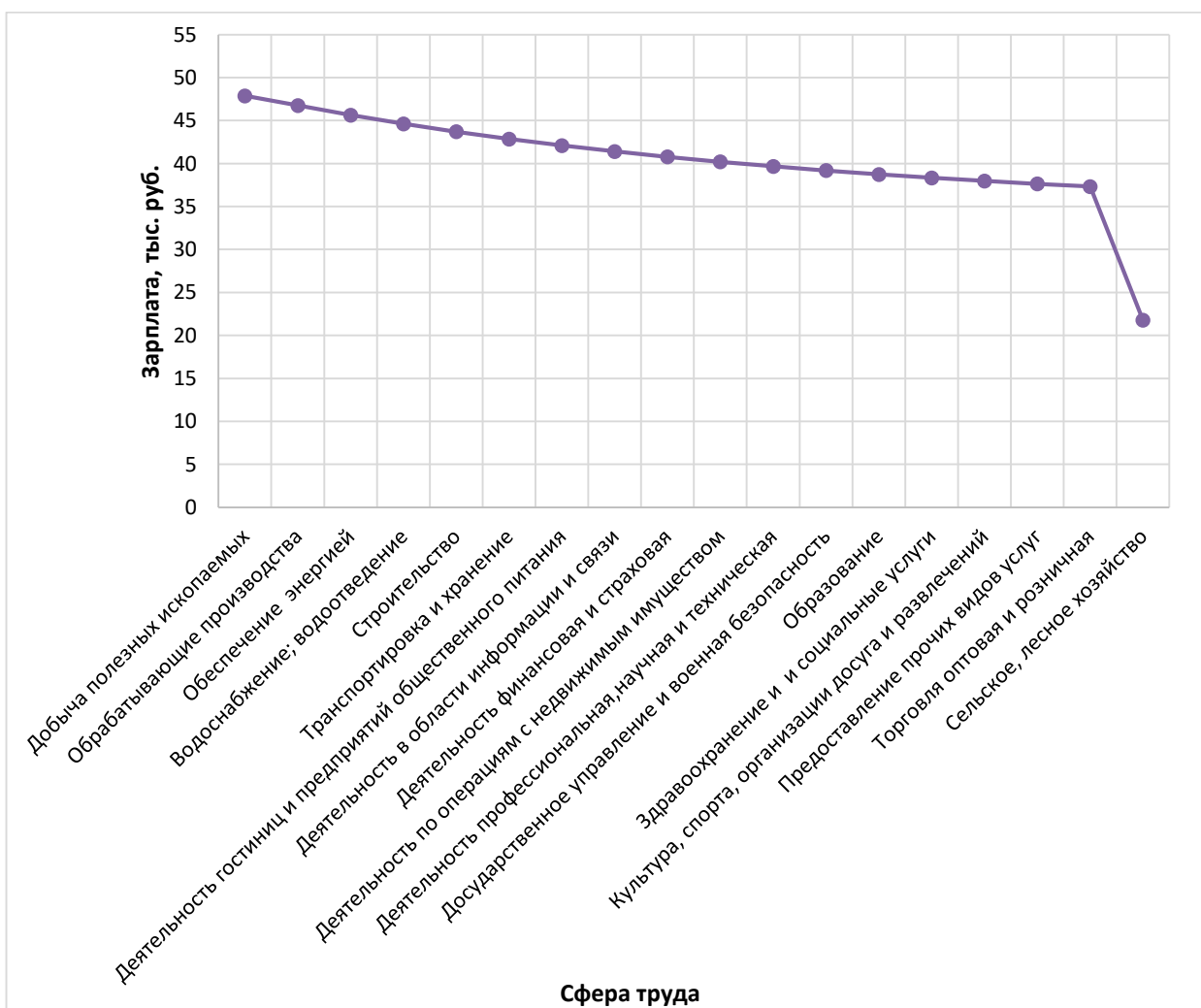
Как видно из рисунка 2, наиболее значимым показателем, влияющим на уровень заработной платы, является Образование.

Далее было проведено исследование полученных зависимостей между входными параметрами и уровнем заработной платы. Исследование производилось с помощью метода «замораживания» суть которого заключается в варьировании значения одного параметра и фиксировании значений всех других параметров.





**Рисунок 3. Зависимость зарплаты человека в зависимости от его образования**



**Рисунок 4. Зависимость зарплаты от сферы труда**

На рисунке 3 показан график зависимости заработной платы от образования. В том случае, когда сфера, год выплат, пол человека одинаковые, а образование разное, нейросеть прогнозирует значительное увеличение заработной платы человека. Можно сделать вывод, что чем выше образование человека, тем больше зарплату он получает, что не противоречит исходным данным, на основании которых строилось учебное множество.

На рисунке 4 продемонстрирована зависимость зарплаты при трех одинаковых показателях, но разной сфере труда. Неизменными показателями будет образование – среднее профессиональное, пол – женский и год – 2023. Можно увидеть, что больше всего зарабатывают люди, которые трудятся в сфере по добыче полезных ископаемых, а меньше всего – в сельском, лесном хозяйстве.

На рисунке 5 изображен график зависимости зарплаты от пола. Как видно из графика, зарплата мужчин выше, чем зарплата женщин почти во всех сферах, за исключением сельского и лесного хозяйства. При анализе исходных данных, за основу была взята таблица, где доля зарплаты женщин, меньше, чем доля мужчин. Поэтому можно сделать вывод, что нейросеть корректно прогнозирует новые результаты, опираясь на опыт предыдущей выборки.

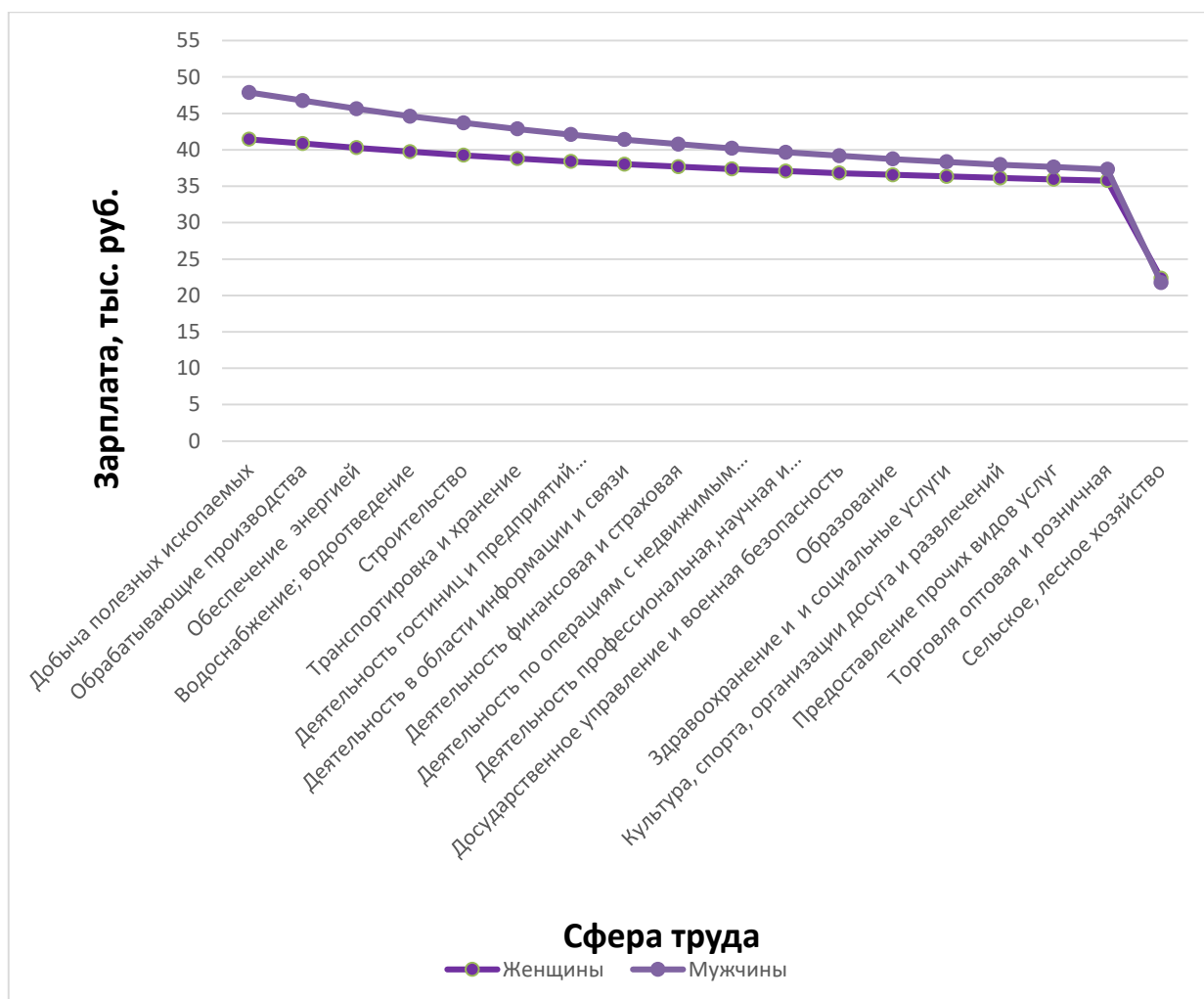
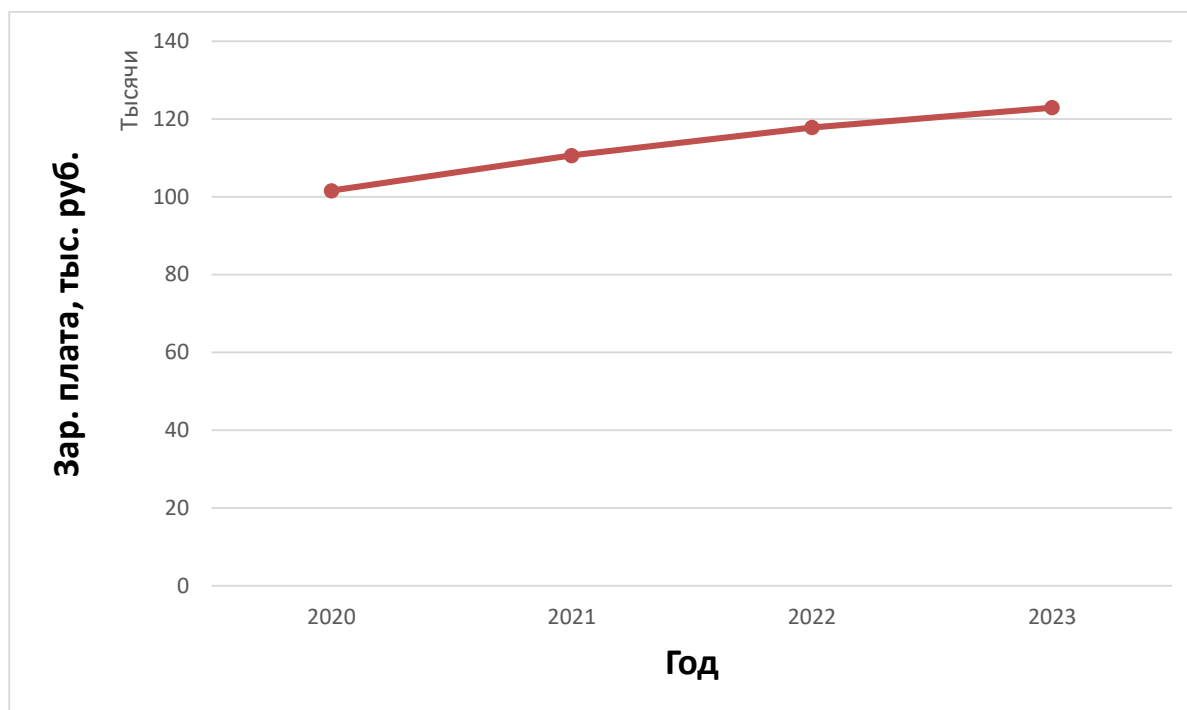


Рисунок 5. Зависимость зарплаты от пола

На рисунке 6 продемонстрирован рост зарплаты исходя из года выплат. Как видно из графика, с увеличением года увеличивается средняя зарплата. Исходные данные тоже имели такую тенденцию, что говорит о том, что наша нейросеть понимает смысл, чем больше год-тем выше зарплаты, это обусловлено и инфляцией, курсом доллара, экономическим состоянием мира в целом.



**Рисунок 6. Зависимость заработной платы от года**

Полученные результаты исследований не противоречат реальности, что подтверждает, что спроектированную нейронную сеть можно считать пригодной для прогнозирования расчета средней заработной платы.

**Заключение.** Построена система нейросетевого прогнозирования средней заработной платы исходя из 4 параметров, года выплат, пола, образования и сферы деятельности. Методом сценарного прогнозирования построены графики зависимостей прогнозируемой заработной платы от изменения входных параметров. Применение такого набора параметров в модели позволяет с высокой точностью прогнозировать среднюю заработную плату.

### **Библиографический список**

1. Rosstat.gov.ru – Федеральная служба государственной статистики. [Электронный ресурс]. Режим доступа: <https://rosstat.gov.ru/>
2. Черепанов Ф.М., Ясницкий Л.Н. Нейросимулятор 4.0. Свидетельство о регистрации программы для ЭВМ RUS 2014612546. Заявка № 2014610341 от 15.01.2014.
3. Ясницкий Л.Н. Интеллектуальные системы: учебник. М.: Лаборатория знаний, 2016. 221 с.
4. Yasnitsky L.N. Algorithm for searching and analyzing abnormal observations of statistical information based on the Arnold – Kolmogorov – Hecht-

## NEURAL NETWORK SYSTEM FOR FORECASTING THE AVERAGE WAGE

*Kamenskikh Lyubov A.*

Perm State University

Str. Bukireva, 15, Perm, Russia, 614990, lyubochka199@yandex.ru

The article describes the development of a neural network system for predicting the results of average wages. The system allows you to predict with high accuracy the average salary depending on the year of accrual, gender, sphere of work and education. The neural network allows you to see the dynamics of salary growth.

**Keywords:** artificial intelligence, neural network technologies, forecasting, average salary, forecast, salary.

УДК 004.032.26

## НЕЙРОСЕТЕВАЯ СИСТЕМА ПРОГНОЗИРОВАНИЯ РЕЙТИНГА КИНОФИЛЬМА В ПРОКАТЕ

*Неверова Елизавета Алексеевна*

Пермский государственный национальный исследовательский университет,  
ПМИ. 614990, Россия, г. Пермь, ул. Букирева, 15, ruinmath@bk.ru

В статье представлено описание разработки нейросетевой системы для прогнозирования рейтинга кинофильмов в прокате. Система позволяет с большой точностью предсказать рейтинг будущего фильма на основании данных о нем. С помощью разработанной интеллектуальной системы проведено исследование предметной области, выявлены закономерности, имеющие практической значение.

**Ключевые слова:** искусственный интеллект, нейросетевые технологии, прогнозирование, фильмы, рейтинг.

**Введение.** Данная модель может помочь оптимизировать финансовые затраты при планировании новых кинофильмов. Она позволяет делать прогнозы о рейтинге фильма и выбирать наиболее эффективный способ распределения бюджета на различные аспекты фильма, тем самым увеличивая его шанс понравится публике и критикам.

Эффективность использования искусственного интеллекта при оценке успешности продуктов киноиндустрии была показана в работах [1-2]. В данных работах получены значительные результаты. В настоящей работе для моделирования было решено использовать схожий, но несколько отличающийся список параметров, используемых для построения прогноза, далее представлено подробное его описание.

Основная цель настоящей работы заключается в сборе множества публичных данных о фильмах, а также создание и обучение нейросетевой модели на этих данных. Конечным результатом должна быть нейросетевая система, способная прогнозировать рейтинги фильмов в больше, чем 70% случаев.

Для создания нейросетевой системы были выбраны следующие параметры:

X1 – Год выпуска фильма.

X2 – Продолжительность фильма (мин).

X3 – Возрастное ограничение фильма: 1 – от 0 лет, 2 – от 6 лет, 3 – от 12 лет, 4 – от 16 лет, 5 – от 18 лет.

X4 – Бюджет фильма (в миллионах долларов США).

X5 – Наличие у фильма приквела: 0 – нет, 1 – да.

X6 – Фильм создан при участии страны «США»: 0 – нет, 1 – да.

X7 – Фильм создан при участии страны «Россия»: 0 – нет, 1 – да.

X8 – Фильм создан при участии страны «Германия»: 0 – нет, 1 – да.

X9 – Фильм создан при участии страны «Канада»: 0 – нет, 1 – да.

X10 – Фильм создан при участии страны «Великобритания»: 0 – нет, 1 – да.

X11 – Фильм относится к жанру «Фантастика»: 0 – нет, 1 – да.

X12 – Фильм относится к жанру «Драма»: 0 – нет, 1 – да.

X13 – Фильм относится к жанру «Мелодрама»: 0 – нет, 1 – да.

X14 – Фильм относится к жанру «Криминал»: 0 – нет, 1 – да.

X15 – Фильм относится к жанру «Боевик»: 0 – нет, 1 – да.

X16 – Фильм относится к жанру «Комедия»: 0 – нет, 1 – да.

X17 – Фильм относится к жанру «Триллер»: 0 – нет, 1 – да.

X18 – Фильм относится к жанру «Ужасы»: 0 – нет, 1 – да.

X19 – Фильм относится к жанру «Мультфильм»: 0 – нет, 1 – да.

X20 – Фильм относится к жанру «Фэнтези»: 0 – нет, 1 – да.

X21 – Фильм относится к жанру «Приключения»: 0 – нет, 1 – да.

X22 – Наличие у режиссера номинаций на премии «Оскар» и/или «Золотой глобус» и/или "Золотой Орел": 0 – нет, 1 – да.

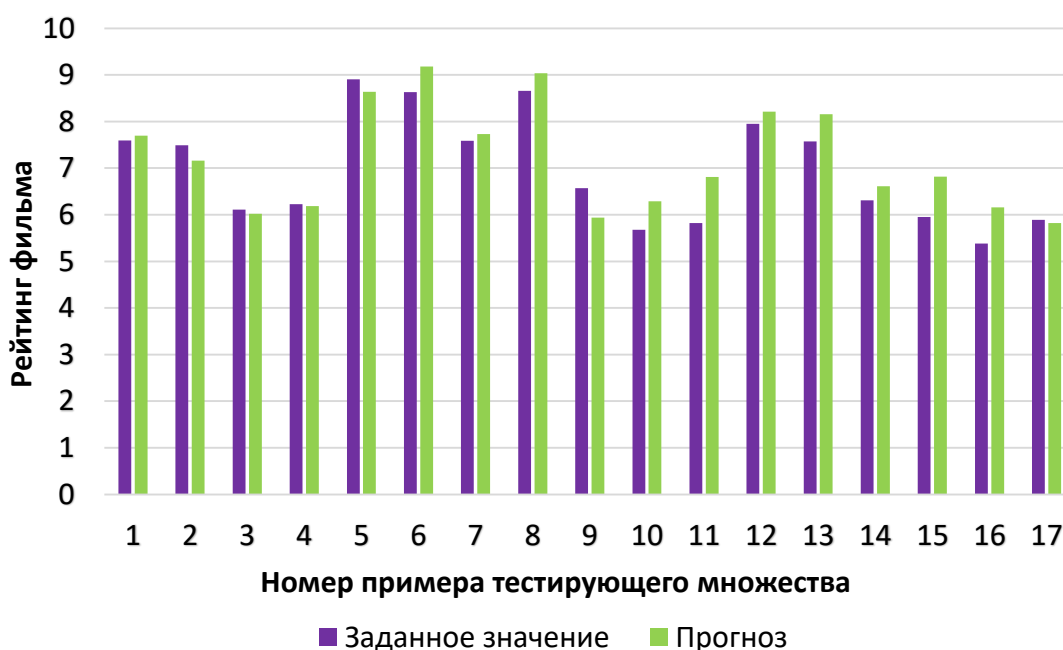
X23 – Наличие у актеров номинаций на премии «Оскар» и/или «Золотой глобус» и/или "Золотой Орел": 0 – нет, 1 – да.

Выходной параметр – рейтинг фильма в прокате.

Всего были собраны данные о 114 фильмах с официальных источников информации. Данное множество было разделено на обучающее и тестирующее в соотношении 85% и 15% соответственно. Большая часть данных была собрана с интернет ресурса [3].

Проектирование, обучение, тестирование нейронной сети выполнялись с помощью программы «Нейросимулятор 5» [4] по методике [5]. В результате оптимизации спроектированная нейронная сеть представляла собой персептрон, который имеет двадцать три входных нейрона, один выходной и один скрытый слой с двумя нейронами.

Для оценки корректной работы спроектированной нейронной сети использовалось тестирующее множество, состоящее из 17 примеров. Средняя относительная ошибка тестирования составила 11,7%, что можно считать приемлемым результатом. На рисунке 1 представлена гистограмма, демонстрирующая разницу между фактическим и прогнозируемым нейросетью рейтингами фильмов



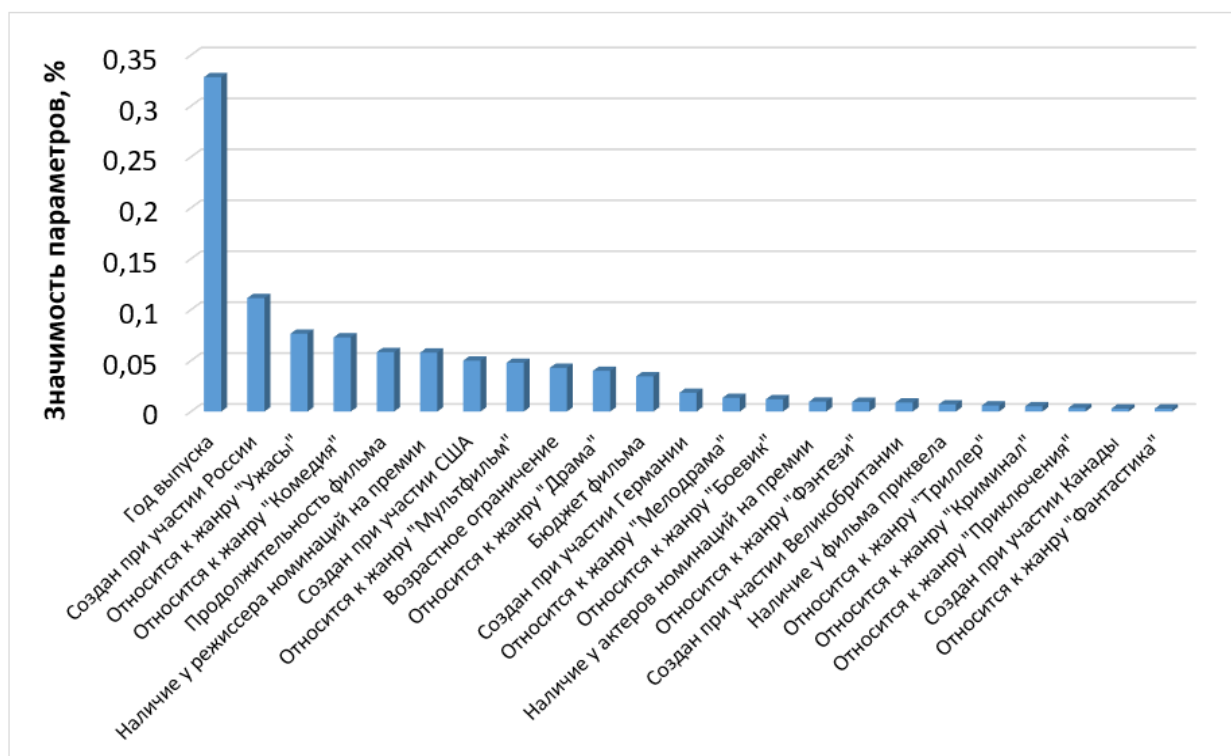
**Рисунок 1. Результат тестирования нейронной сети**

Из результатов, изображенных на рисунке 1, можно сделать вывод об адекватной работе спроектированной нейронной сети.

Оценка значимости входных параметров была выполнена с помощью программы «Нейросимулятор 5», результат отображен на рисунке 2.

Как видно из рисунка 2, очень важным параметром успешности фильма является год его выпуска. Так же важным параметром является снят ли фильм в России, анализируя узнаём, что это связано с плохо зарекомендовавшим себя Российским кинематографом.

Далее было проведено исследование полученных зависимостей между входными параметрами и рейтингом фильма. Исследование производилось с помощью метода «замораживания» [5], суть которого заключается в варьировании значения одного параметра и фиксации значений всех других параметров. Для этого были выбраны 3 фильма: «Лига справедливости», «Ночь в музее» и «Назад в будущее» информация о входных параметрах которых отражена в таблице 1.



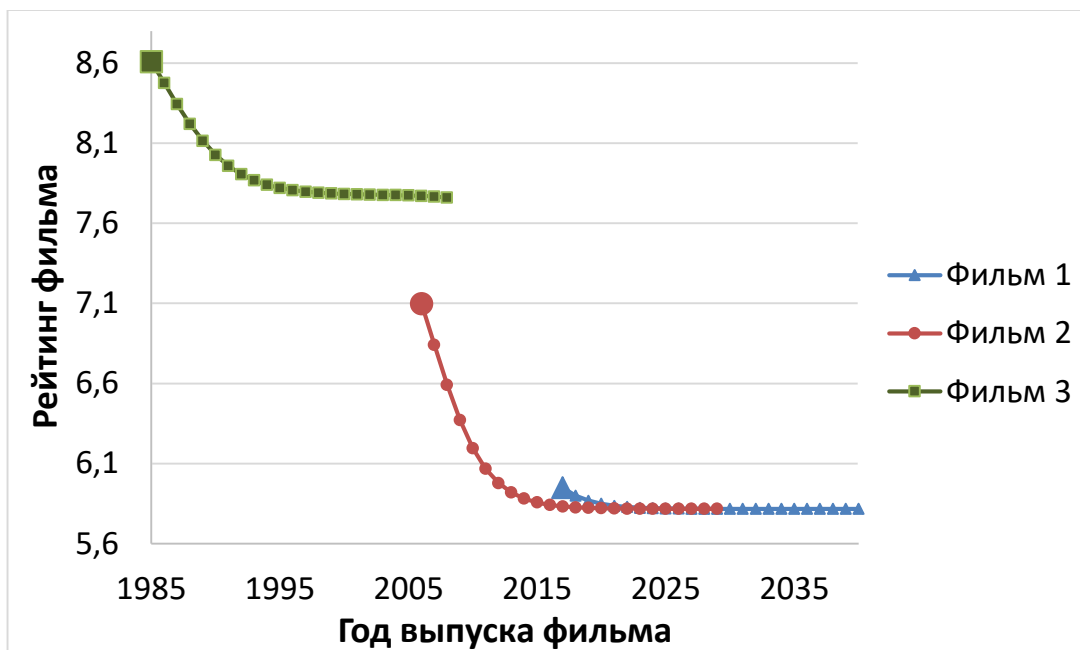
**Рисунок 2. Значимость входных параметров нейросетевой модели**

**Таблица 1**

**Характеристики фильмов, выбранных для исследования**

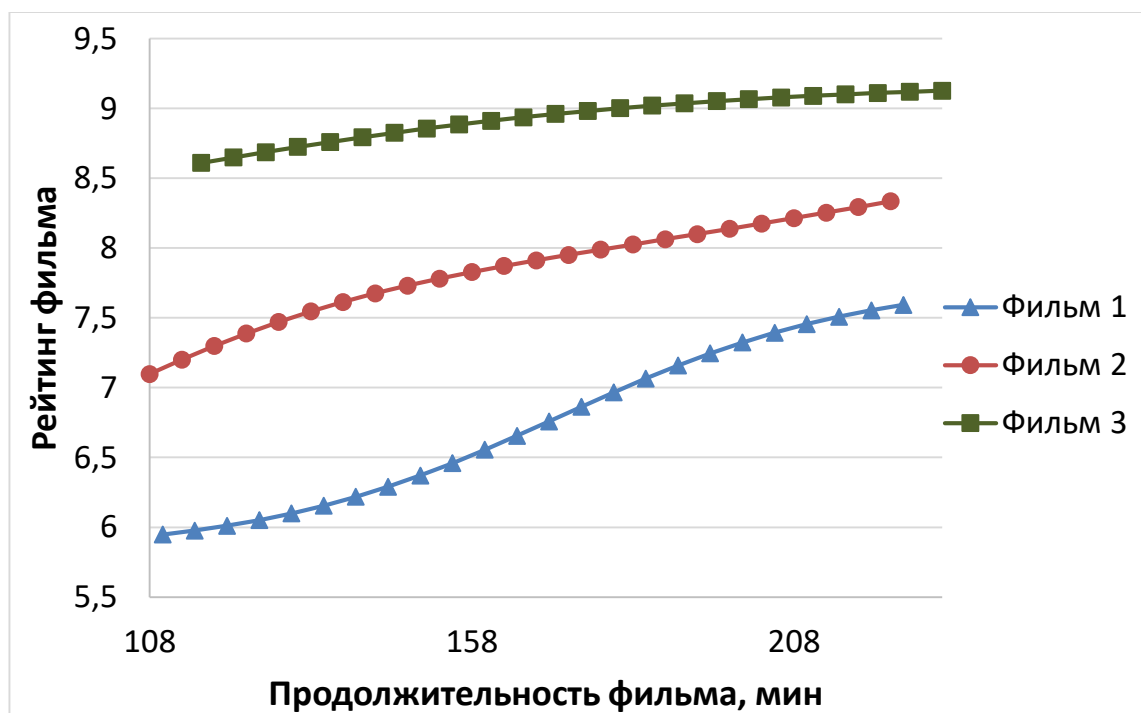
ФИЛЬМ	1 – Лига справедливости	2 – Ночь в музее	3 – Назад в будущее
X1	2017	2006	1985
X2	110	108	116
X3	4	3	3
X4	300	110	19
X5	1	0	0
X6	1	1	1
X7	0	0	0
X8	0	0	0
X9	1	0	0
X10	1	1	0
X11	1	0	1
X12	0	0	0
X13	0	0	0
X14	0	0	0
X15	1	0	0
X16	0	1	1
X17	0	0	0
X18	0	0	0
X19	0	0	0
X20	1	1	0
X21	0	1	1
X22	0	0	0
X23	1	1	0

На рисунке 3 показан график зависимости рейтингов фильмов от года выпуска. Как следует из графика, увеличение года выпуска картины приводит к уменьшению рейтинга.



**Рисунок 3. Зависимость рейтингов фильмов от года выпуска (крупными маркерами указаны изначальные рейтинги фильмов)**

На рисунке 4 показан график изменения рейтинга фильма в зависимости от увеличения продолжительности картины. Как видно из графика, продолжительность фильма увеличивает его рейтинг.



**Рисунок 4. Зависимость рейтингов от продолжительностей фильмов**



Полученные результаты исследований не противоречат реальности, что подтверждает, что спроектированную нейронную сеть можно считать пригодной для прогнозирования рейтингов кинофильмов в прокате.

**Заключение.** Построена система нейросетевого прогнозирования рейтингов фильмов. Спроектированная нейросетевая модель учитывает год выпуска фильма, его продолжительность, возрастное ограничение, бюджет, наличие у фильма приквела, при участии каких стран был создан фильм, к каким жанрам он относится, а также наличие премий у режиссёра и актеров. Проведены исследования рынка киноиндустрии. Методом сценарного прогнозирования построены графики зависимостей рейтингов фильмов от их продолжительности и года выпуска. Применение такого набора параметров в модели позволяет с высокой точностью прогнозировать рейтинг кинофильма.

### Библиографический список

1. Ясницкий Л.Н., Плотников Д.И. Экономико-математическая нейросетевая модель для оптимизации финансовых затрат в кинобизнесе // Фундаментальные исследования. 2016. № 11-2. С. 339-342.
2. Ясницкий Л.Н., Белобородова Н. О., Медведева Е.Ю. Методика нейросетевого прогнозирования кассовых сборов кинофильмов // Финансовая аналитика: проблемы и решения. 2017. Т. 10. № 4 (334). С. 449-463.
3. Kinopoisk.ru – интернет-сервис о кино. [Электронный ресурс]. Режим доступа: <https://www.kinopoisk.ru/>
4. Черепанов Ф.М., Ясницкий Л.Н. Нейросимулятор 4.0. Свидетельство о регистрации программы для ЭВМ RUS 2014612546. Заявка № 2014610341 от 15.01.2014.
5. Ясницкий Л.Н. Введение в искусственный интеллект. М.: Издательский центр «Академия», 2005. – 176 с.

## NEURAL NETWORK SYSTEM FOR FORECASTING THE RATING OF A FILM IN THE RENTAL

*Neverova Elizaveta A.*

Perm State University

Str. Bukireva, 15, Perm, Russia, 614990, ruinmath@bk.ru

The article presents a description of the development of a neural network system for predicting the rating of films at the box office. The system allows you to predict the rating of a future film with great accuracy based on data about it. With the help of the developed intellectual system, a study of the subject area was carried out, patterns of practical importance were identified.

**Keywords:** artificial intelligence, neural network, prediction, films, rating.

## НЕЙРОСЕТЕВАЯ СИСТЕМА ПРОГНОЗИРОВАНИЯ ОЦЕНКИ КОМПЬЮТЕРНЫХ ИГР

*Кондратенков Валерий Юрьевич, Плахина Татьяна Сергеевна,  
Орлов Александр Анатольевич*

Пермский государственный национальный исследовательский университет,  
КМБ. 614990, Россия, г. Пермь, ул. Букирева, 15, plahinats@psu.ru

В статье представлено описание разработки нейросетевой системы для прогнозирования оценки компьютерных игр. Система позволяет предсказать, какие средние оценки критиков и пользователей получит компьютерная игра на основании данных о ее издателе, двух основных жанрах, стоимости и годе выпуска.

**Ключевые слова:** искусственный интеллект, нейросетевые технологии, прогнозирование, компьютерные игры, оценка.

**Введение.** Ежегодно в игровой индустрии производится множество новых разнообразных игр. Одни похожи на то, что уже было создано ранее, другие, напротив, – результат некоторого «эксперимента» и совершенно не похожи на то, что уже выпускали ранее. Каждый желающий может найти то, что будет ему по вкусу.

Некоторые видеоигры оказываются успешнее, чем другие. Зачастую даже игры, принадлежащие одной серии, могут оставлять совершенно разное впечатление. Каждый игрок (далее пользователь) может выставить свою оценку для конкретной видеоигры. Именно по множеству таких оценок методом среднего арифметического высчитывается средняя оценка пользователей.

Кроме того, существует оценка критиков. Критик – это человек, который хорошо разбирается в данной предметной области и, как следствие, может качественно проанализировать и оценить материал [1]. Оценка критиков, которую мы используем в нашей системе, так же вычисляется как среднее арифметическое между оценками, которые были выставлены видеоигре.

Существует множество источников, на которых можно увидеть оценки критиков и пользователей по тем или иным играм [2-5]. В качестве оценок для большинства видеоигр были использованы данные с сайта Metacritic, так как данный источник является довольно авторитетным и на нем есть информация о большем количестве игр [2].

Основная цель настоящей работы заключается в сборе множества публичных данных о видеоиграх и их оценках (как со стороны критиков, так и со стороны пользователей), а также создание и обучение нейросетевой модели на этих данных. Конечным результатом должна быть нейросетевая система, способная прогнозировать оценки для игр, которые будут выпускаться в дальнейшем.

Для создания нейросетевой системы были выбраны следующие параметры: X1 – издатель видеоигры (используется номер издателя из таблицы 1), X2 – основной жанр игры (используется номер жанра из таблицы 2), X3 – дополнительный жанр игры (используется номер жанра из таблицы 2), X4 – стоимость видеоигры, X5 – год выпуска видеоигры.

**Таблица 1**

**Соотношение между основными издателями и номерами  
для входных данных X1**

<b>Издатель</b>	<b>Номер</b>
CD Project Red	1
Ubisoft	2
Activision	3
Rockstar	4
Valve	5
Riot games	6
Bethesda	7
Daedalic Entertainment	8
Square Enix	9
Microsoft	10
2K	11
CAPCOM	12
EA	13
From Software	14
Sony Interactive Entertainment	15
Larian studios	16
505 Games	17
Deep Silver	18
Прочее	19

Обучающее множество было собрано вручную. Перед переходом к проектированию и обучению нейросетевой модели, была выполнена очистка исходного множества от противоречивых примеров, выбросов, дубликатов. Например, некорректными примерами считались те, у которых была большая разница между оценкой критиков и оценкой пользователей (вероятнее всего причиной высокой разницы были технические проблемы версии игры). Выбросы обнаруживались по методике [14]. Таким образом, объем итогового множества включает в себя 397 примеров. Данное множество было разделено на обучающее и тестирующее в соотношении 90% и 10% соответственно. Так же были данные, содержащиеся и в тестовом, и в обучающем множестве. Все данные были собраны с интернет ресурсов [2-11].

Для вывода каждой оценки было решено сделать отдельную сеть, поскольку в таком случае результаты были более точными. Проектирование, обучение, тестирование нейронных сети выполнялись с помощью программы «Нейросимулятор 5» [12] по методике [13]. Спроектированная нейронная сеть для предсказания оценки критиков представляла собой персептрон, который имеет пять входных нейронов, один выходной и четыре скрытых слоя с пятью нейронами на первом, четырьмя на втором, тремя на третьем и

одним на четвертом. Нейронная сеть для предсказания оценки пользователей выглядит так же.

Таблица 2

**Соотношение между основными жанрами и номерами для входных данных X2 и X3**

Жанр	Номер
Шутер	1
Стратегия	2
Roguelike	3
Visual Novel	4
Песочница	5
Хоррор	6
МОВА	7
Гонки	8
RPG	9
Инди	10
ММО	11
Спортивная	12
Приключения	13
Прочее	14
CCG	15
Головоломка	16
Battle royal	17
Файтинг	18

Сперва опишем нейросеть для прогнозирования оценки критика. Для оценки корректной работы спроектированной нейронной сети использовалось тестирующее множество, состоящее из 23 примеров. Средняя относительная ошибка тестирования составила 7,8%, что можно считать приемлемым результатом. На рисунке 1 представлена гистограмма, демонстрирующая разницу между фактической и прогнозируемой нейросетью оценкой критиков.

Из результатов, изображенных на рисунке 1, можно сделать вывод об адекватной работе спроектированной нейронной сети.

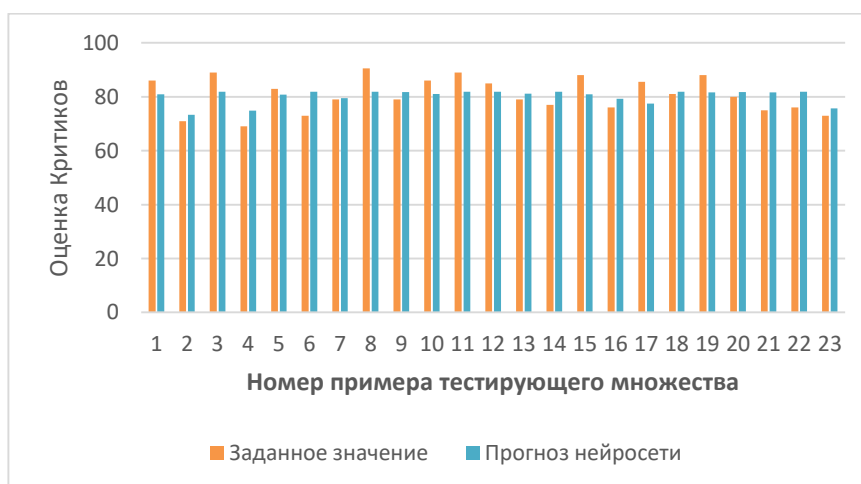
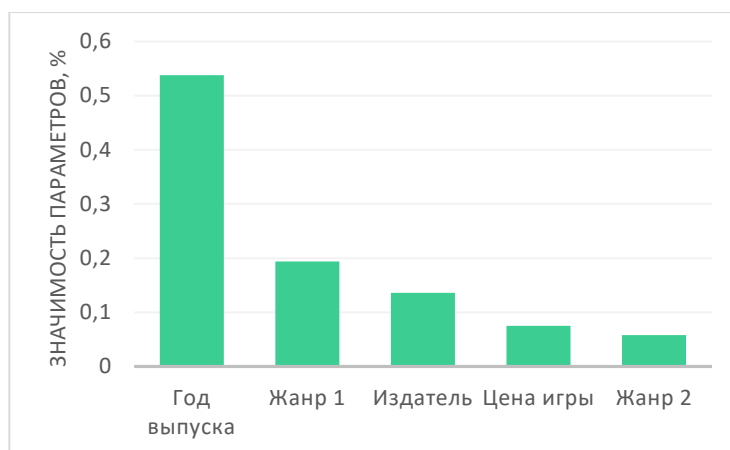


Рисунок 1. Результат тестирования нейронной сети для оценки критиков

Оценка значимости входных параметров была выполнена с помощью программы «Нейросимулятор 5», результат отображен на рисунке 2.

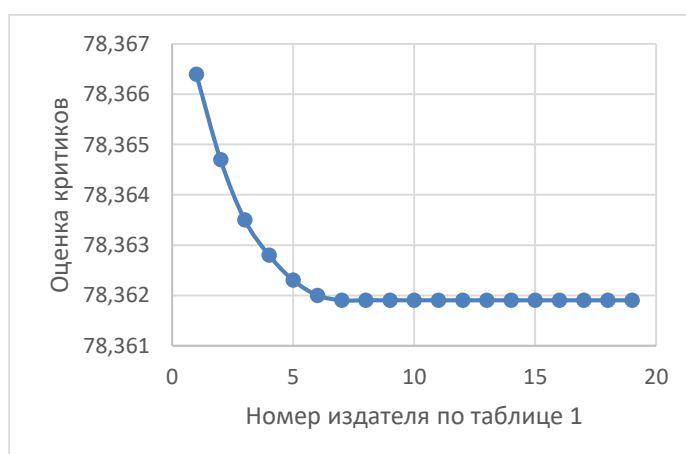


**Рисунок 2. Значимость входных параметров нейросетевой модели для оценки критиков**

Как видно из рисунка 2, наиболее значимыми для оценки критика являются такие параметры как год выпуска и основной жанр игры. Эти результаты вполне ожидаемы, так как оценка во многом базируется на жанре видеоигры и стоимость игры для критика не имеет большого значения.

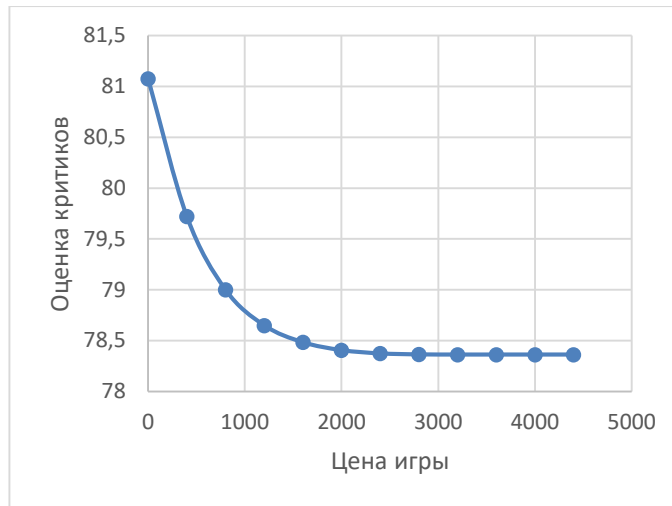
Далее было проведено исследование полученных зависимостей между входными параметрами и оценкой критиков. Исследование производилось с помощью метода «замораживания» [13], суть которого заключается в варьировании значения одного параметра и фиксации значений всех других параметров. Для этой цели была отобрана игра, для которой нейросеть выдала наименьшую ошибку.

На рисунке 3 показан график зависимости прогнозируемой оценки от издателя игры. По рисунку видно, что прогноз изменяется незначительно



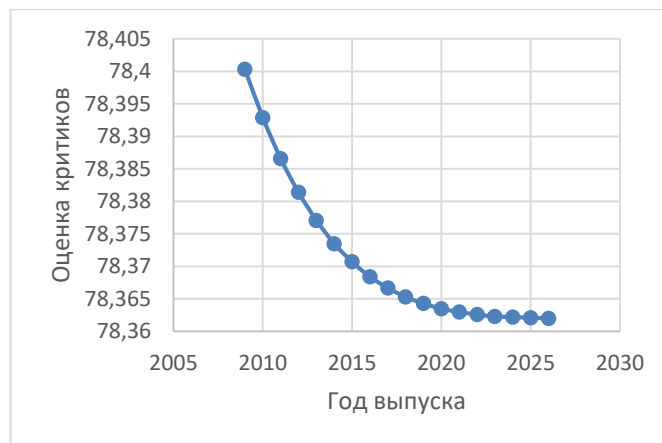
**Рисунок 3. Зависимость прогнозируемой оценки от издателя.**

На рисунке 4 показан график зависимости прогнозируемой оценки от цены игры. По рисунку видно, что прогноз изменяется немного больше, чем на рисунке 2, но только на промежутке от 0 до 2000 рублей.



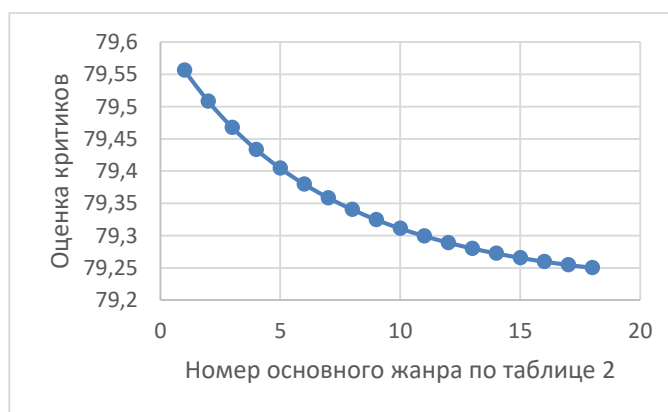
**Рисунок 4. Зависимость прогнозируемой оценки от цены.**

На рисунке 5 показан график зависимости прогнозируемой оценки от года выпуска игры. В этом случае, так же, как и в первом случае, оценка меняется незначительно.



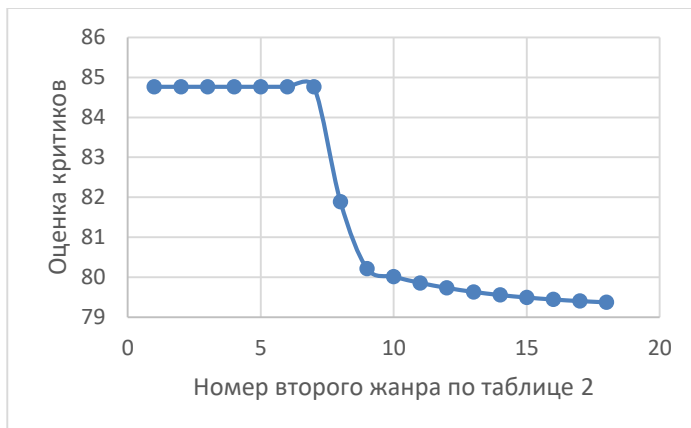
**Рисунок 5. Зависимость прогнозируемой оценки от года выпуска.**

На рисунке 6 показан график зависимости прогнозируемой оценки от основного жанра игры. В этом случае, так же оценка меняется незначительно.



**Рисунок 6. Зависимость прогнозируемой оценки от основного жанра.**

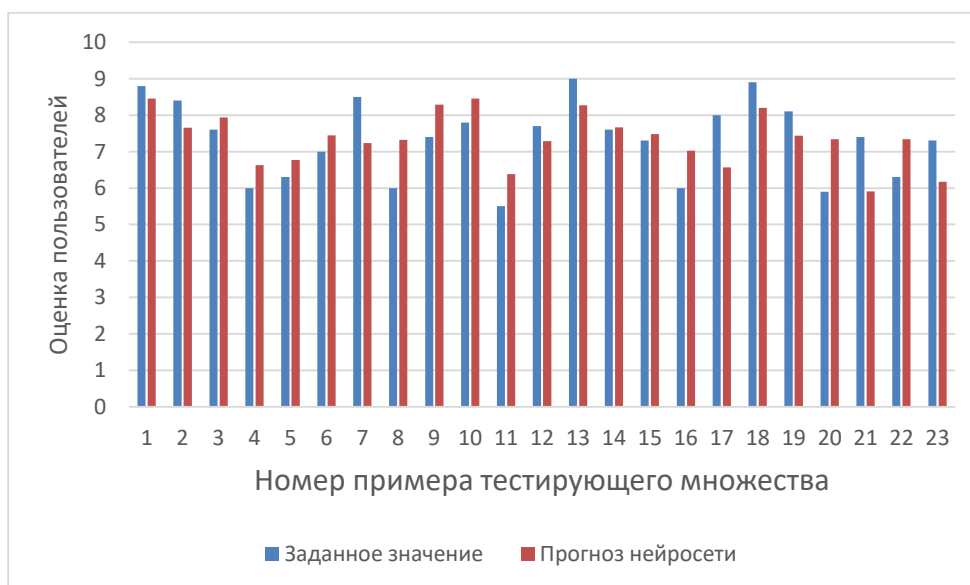
На рисунке 7 показан график зависимости прогнозируемой оценки от второго жанра игры. В этом случае, оценка значительно меняется при значении жанров 7, 8 и 9.



**Рисунок 7. Зависимость прогнозируемой оценки от второго жанра**

На основе этих данных можно сделать вывод, что прогноз нейросети практически не меняется при изменении какого-то одного параметра.

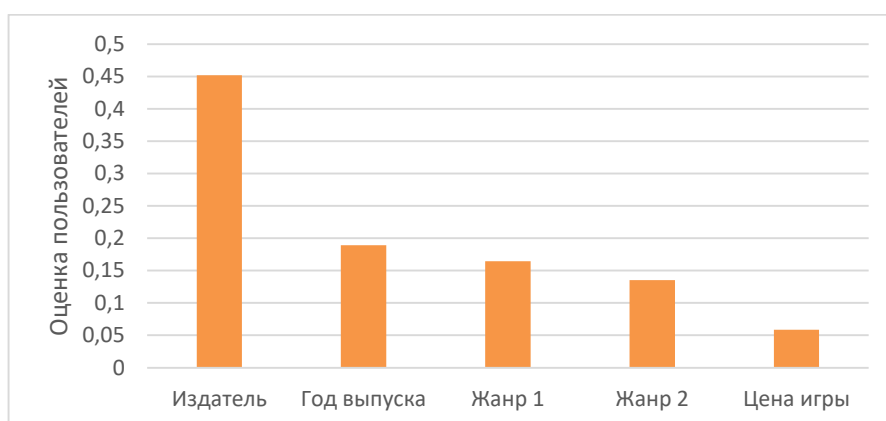
Далее проведем анализ нейросетевой модели для прогнозирования оценки пользователей. Для оценки корректной работы использует то же тестирующее множество из 23 примеров. Средняя относительная ошибка тестирования составила 11,7%, что немного больше, чем для оценки критиков, но также считается приемлемым результатом. На рисунке 8 представлена гистограмма, демонстрирующая разницу между фактической и прогнозируемой нейросетью оценкой пользователей.



**Рисунок 8. Результат тестирования нейронной сети для оценки пользователей**

Из результатов, изображенных на рисунке 8, можно сделать вывод о том, что нейронная сеть так же работает адекватно.

Тем же способом провели оценку входных. Результат отображен на рисунке 9.



**Рисунок 9. Значимость входных параметров нейросетевой модели для оценки пользователей**

Как видно из рисунка 9, наиболее значимыми для оценки пользователей являются такие параметры как издатель и год выпуска игры. Эти результаты также вполне соответствуют реальности.

Далее провели исследование полученных зависимостей между входными параметрами и оценкой пользователей. Исследование, как и с первой нейросетью, производили с помощью метода «замораживания» [13]. Для этой цели также была отобрана игра, для которой нейросеть выдала наименьшую ошибку.

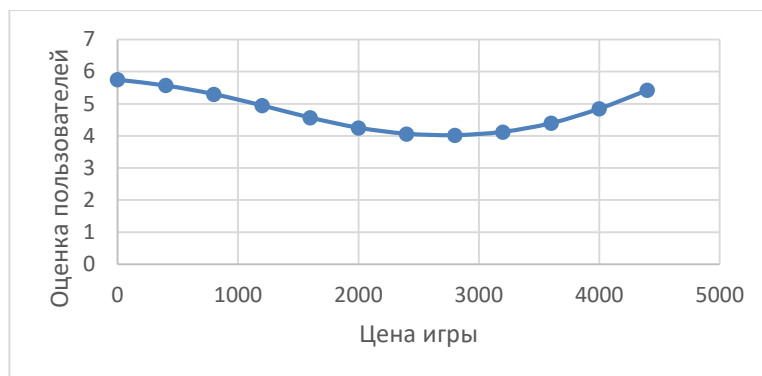
На рисунке 10 показан график зависимости прогнозируемой оценки от издателя игры. По рисунку видно, что прогноз очень сильно зависит от данного параметра.



**Рисунок 10. Зависимость прогнозируемой оценки от издателя**

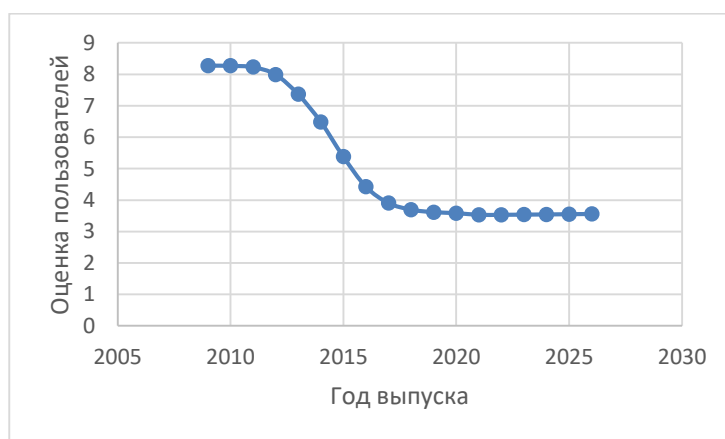
На рисунке 11 показан график зависимости прогнозируемой оценки от издателя игры. По рисунку видно, что прогноз изменяется сильнее, чем в первом случае, а самую низкую оценку будут иметь игры, которые стоят от 2400 до 3000 рублей.





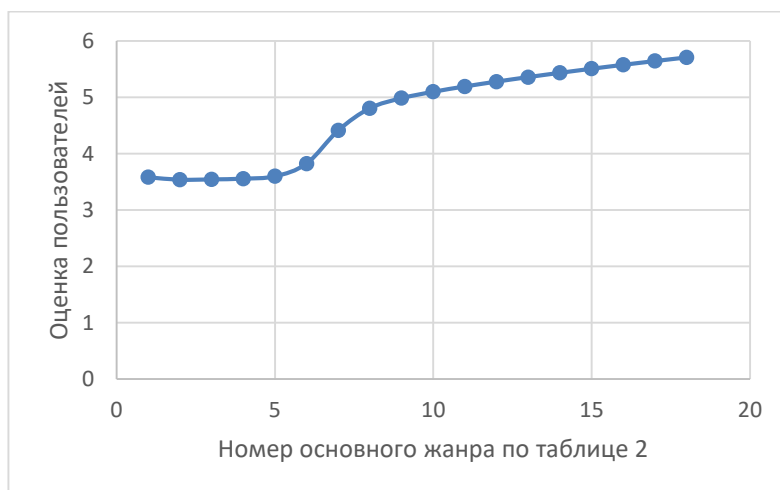
**Рисунок 11. Зависимость прогнозируемой оценки от цены**

На рисунке 12 показан график зависимости прогнозируемой оценки от года выпуска игры. По рисунку видно, что прогноз так изменяется сильнее, чем в первом случае. Если игра с исходными параметрами будет выходить в дальнейшем, то она будет иметь низкую оценку пользователей.

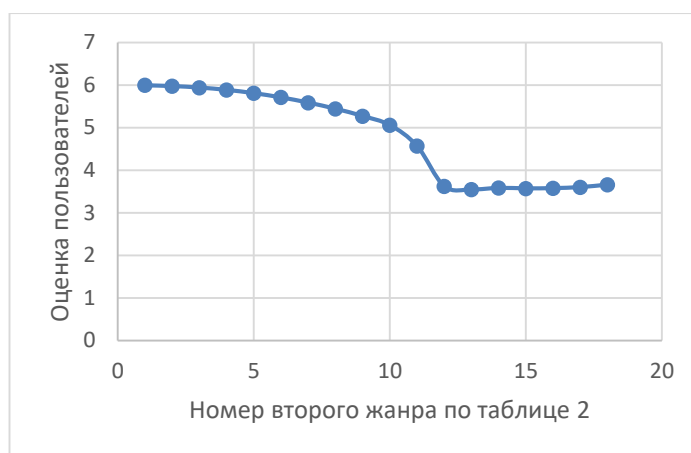


**Рисунок 12. Зависимость прогнозируемой оценки от года выпуска игры**

На рисунках 13 и 14 показан график зависимости прогнозируемой оценки от жанров. Видно, что оценка так же сильно меняется при изменении каждого из данных параметров.



**Рисунок 13. Зависимость прогнозируемой оценки от основного жанра.**



**Рисунок 14. Зависимость прогнозируемой оценки от второго жанра.**

На основании представленных выше данных можно сделать вывод, что изменение любого параметра существенно влияет на оценку, выдаваемую нейросетью.

Таким образом, мы с уверенностью можем утверждать, что обе спроектированные нейронные сети пригодны для прогнозирования оценок, которые получит компьютерная игра.

**Заключение.** Построены две системы нейросетевого прогнозирования: первая для прогнозирования оценки видеоигры со стороны критиков, вторая – оценка со стороны пользователей. Спроектированные нейросетевые модели учитывают пять параметров: издатель, основной жанр, второй жанр, стоимость видеоигры и год выпуска. Кроме того, построены графики зависимостей прогнозируемых оценок от изменения входных параметров. Применение такого набора параметров в модели позволяет с высокой точностью прогнозировать обе оценки компьютерных игр.

### Библиографический список

1. thefreedictionary.com – определение слова «критик» в словаре [Электронный ресурс]. Режим доступа: <https://www.thefreedictionary.com/Critics>
2. Metacritic – основной источник данных об оценке уже вышедших видеоигр [Электронный ресурс]. Режим доступа: <https://www.metacritic.com/>
3. IGN – дополнительный источник данных об оценке уже вышедших видеоигр [Электронный ресурс]. Режим доступа: <https://ru.ign.com/>
4. Video Games Reviews & News – GameSpot – дополнительный источник данных об оценке уже вышедших видеоигр [Электронный ресурс]. Режим доступа: <https://www.gamespot.com/>
5. GameInformer.com – дополнительный источник данных об оценке уже вышедших видеоигр [Электронный ресурс]. Режим доступа: <https://www.gameinformer.com/>
6. Steam – интернет-магазин видеоигр [Электронный ресурс]. Режим доступа: <https://store.steampowered.com/?l=russian>

7. Официальный сайт PlayStation – источник данных о стоимости видеоигр для PlayStation [Электронный ресурс]. Режим доступа: <https://www.playstation.com/ru-ru/>
8. UBISOFT Официальный магазин RU – интернет-магазин видеоигр от компании Ubisoft [Электронный ресурс]. Режим доступа: <https://store.ubi.com/ru/home>
9. Epic Games Store | Official Site – интернет-магазин видеоигр [Электронный ресурс]. Режим доступа: <https://www.epicgames.com/store/ru/>
10. Origin – интернет-магазин видеоигр от компании EA [Электронный ресурс]. Режим доступа: <https://www.origin.com/rus/ru-ru/store>
11. Microsoft Store | Xbox – источник данных о стоимости видеоигр для Xbox [Электронный ресурс]. Режим доступа: <https://www.xbox.com/ru-RU/microsoft-store>
12. Черепанов Ф.М., Ясницкий Л.Н. Нейросимулятор 4.0. Свидетельство о регистрации программы для ЭВМ RUS 2014612546. Заявка № 2014610341 от 15.01.2014.
13. Ясницкий Л.Н. Интеллектуальные системы: учебник. М.: Лаборатория знаний, 2016. 221 с.
14. Yasnitsky L.N. Algorithm for searching and analyzing abnormal observations of statistical information based on the Arnold – Kolmogorov – Hecht-Nielsen theorem // International Journal of Advanced Trends in Computer Science and Engineering. 2020. 9(2). Pp. 1814-1819. DOI: 10.30534/ijatcse/2020/139922020

## **NEURAL NETWORK SYSTEM FOR PREDICTING THE EVALUATION OF COMPUTER GAMES**

*Kondratenkov Valeriy Yu., Plakhina Tatyana S., Orlov Alexander A.*  
Perm State University  
Str. Bukireva, 15, Perm, Russia, 614990, plahinats@psu.ru

The article describes the development of a neural network system for predicting the evaluation of computer games. The system allows you to predict what average ratings of critics and users will receive a computer game based on data about its publisher, two main genres, cost and year of release.

**Keywords:** artificial intelligence, neural network, prediction, computer games, evaluation

## НЕЙРОСЕТЕВАЯ СИСТЕМА АВТОМАТИЗАЦИИ УПРАВЛЕНИЯ СУШИЛЬНОЙ КАМЕРОЙ ДЛЯ ПРОСУШИВАНИЯ СЕМЯН ШИШЕК

*Овчинникова Аделина Александровна*

Национальный исследовательский университет «Высшая школа экономики»  
614060, Россия, г. Пермь, ул. Бульвар Гагарина, 37а, deley@mail.ru

В статье представлено описание разработки нейросетевой системы для автоматического управления оборудованием в сушильной камере при просушивании семян шишек на лесосеменном предприятии. Данная система позволяет сотрудникам предприятия точно поддерживать температурные режимы для раскрытия шишки и сохранности качественных показателей всхожести семян. На основе времени просушивания и получаемых данных с датчиков температур и уровня влажности нейросеть способна определять в какой момент необходимо включить или выключить устройства в камере. С помощью разработанной интеллектуальной системы проведено исследование предметной области, выявлены закономерные условия просушивания семян в камере, имеющие практическое значение.

**Ключевые слова:** искусственный интеллект, нейросетевые технологии, шишки, лесосеменная отрасль, сушка семян шишек.

**Введение.** Восстановление лесных угодий напрямую зависит от качества высаженных семян, предоставляемых лесосеменными отраслями. Процесс подготовки и переработки семян в таких отраслях играет ключевую роль в сохранении жизненно важных свойств будущих растений. Для того, чтобы подготовить семена, например, ели или сосны, сотрудники лесосеменного предприятия должны для начала извлечь их из шишек через длительное просушивание в сушильной камере. При этом перед ними стоит главная задача – сохранение микроклиматических условий, чтобы семена не пересушились и в последствии не погибли.

К сожалению, в России процессы просушивания семян остаются на уровне ручного управления устройствами в сушильной камере. Сотрудникам предприятия в определенные сезоны года приходится оставаться на ночные смены, пока непрерывно работает сушильная камера. Для каждого вида семян ели и сосны должны соблюдаться свои показатели температуры и уровня влажности в камере. Автоматизированное управление устройствами сушильной камеры на основе нейросетевой сети во время сушки семян могло бы помочь предприятию в контроле за данным процессом без постоянного вмешательства сотрудников.

При анализе литературных источников выяснилось, что работ на тему автоматизации процессов просушивания семян на основе нейросети в лесо-

семенной отрасли – нет. Однако есть исследования, направленные на прогнозирование качества семян определенных культур во время просушивания. Так, [1] авторы применяют рекуррентную нейронную сеть для прогнозирования качества семян кукурузы при просушивании. В качестве входных параметров они учитывают температуру в камере и показатель влажности, а на выходе получают процент содержания влаги в семенах кукурузы. Процент влаги в любых семенах играет важную роль для их всхожести. Если семя пересушено, то оно уже считается погибшим и из него не вырастит растение. Если в семени наоборот содержится более высокий процент влаги, то оно является непригодным для длительного хранения.

Для сохранения качества семян сотрудники лесосеменного предприятия используют специальные шкафы, в которых находятся, как было сказано ранее, устройства для поддержания микроклиматических условий для каждого семян ели и сосны. К такому оборудованию относятся вытяжка, калорифер, вентиляция и печь. Вентиляция и печь работают непрерывно с самого запуска процесса просушивания, однако режим работы калорифера и вытяжки напрямую зависит от показателей температуры и уровня влажности в камере.

Микроклиматические условия для каждого вида семян отличаются, поэтому целью данной работы является создание и обучение нейросетевой модели, способной управлять устройствами сушильной камеры для поддержания оптимальных условий просушивания семян на примере шишек ели и сосны. Для создания нейросетевой системы были выбраны следующие параметры:

- X1 – вид загруженной шишки (один из вариантов: ель или сосна),
- X2 – показатель с датчика температуры,
- X3 – показатель с датчика уровня влажности в камере,

Выходными параметрами являются состояния устройств в сушильной камере:

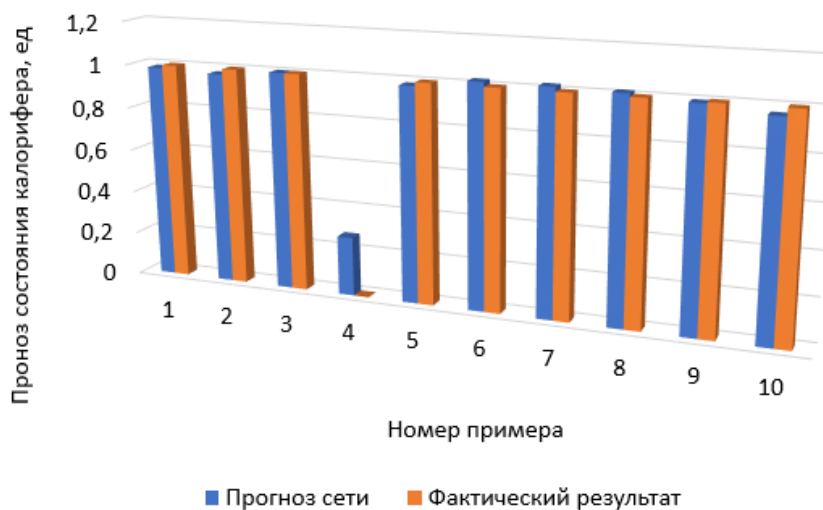
- Y1 – состояние калорифера,
- Y2 – состояние вытяжки,

Как было сказано ранее, в процессе просушивания семян устройства печи и вентиляции работают постоянно, поэтому данное оборудование было исключено из множества выходных параметров.

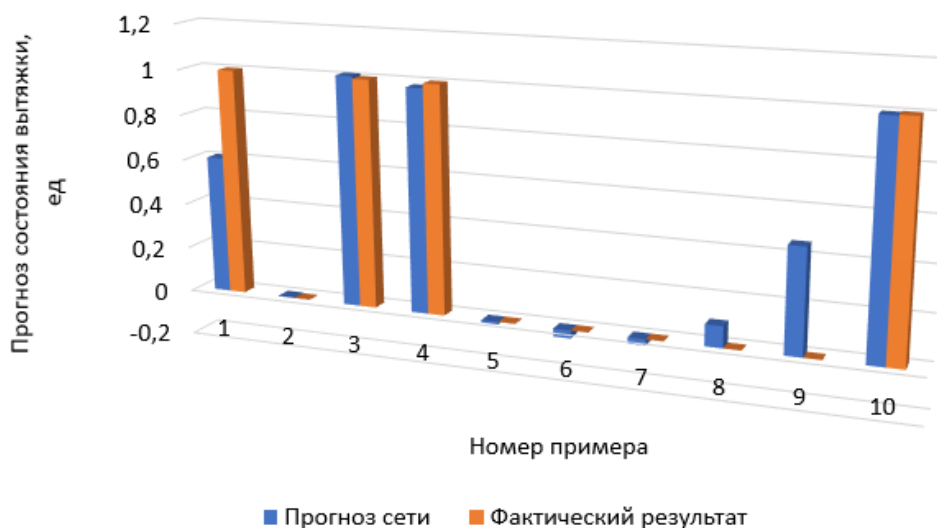
Обучающее множество было собрано на основе экспертного опыта сотрудника, занимающимся просушиванием семян шишек на лесосеменном предприятии «Компания ЛСР». Полученные данные были проанализированы на предмет выбросов с помощью однослойной нейросети. Примеры, сильно противоречащие общему множеству данных, были из него исключены. Таким образом, в итоговом множестве содержится 52 примера, из них 13 данных являются валидирующим множеством, 10 примеров – тестовыми, а остальные – обучающими.

Процессы проектирования, обучения и тестирования нейронной сети выполнялись с помощью программы «Нейросимулятор 5». Полученная сеть содержит три входных и два выходных нейрона, два скрытых слоя: на первом слое десять нейронов, на втором – шесть.

Для оценки корректной работы спроектированной нейронной сети использовалось тестирующее множество, состоящее из 10 примеров. Средняя относительная ошибка тестирования для Y1 (состояние калорифера) составила 4.2%, а для Y2 (состояние вытяжки) – 10.5%, что являются приемлемыми результатами. На рисунках 1 и 2 представлены гистограммы, демонстрирующие разницу между фактическими и прогнозируемыми нейросетью состояниями калорифера и вытяжки в сушильной камере.



**Рисунок 1. Доказательство адекватности состояния калорифера**



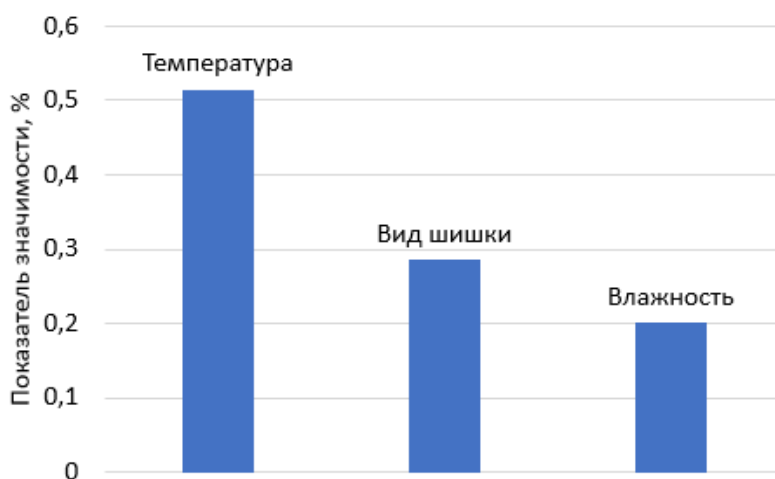
**Рисунок 2. Доказательство адекватности состояния вытяжки**

Из результатов, изображенных на рисунках 1 и 2, можно сделать вывод об адекватной работе спроектированной нейронной сети.

Оценка значимости входных параметров была выполнена с помощью программы «Нейросимулятор 5», результаты отображены на рисунке 3.

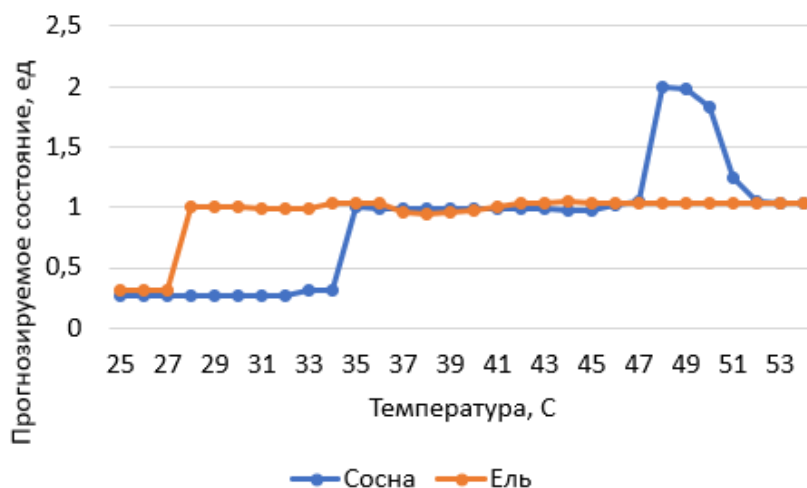
Как видно из рисунка 3, наиболее значимым параметром является температура в сушильной камере. Средний по значимости – вид шишки, так как для просушивания семян режим работы сушильной камеры напрямую зависит

от того, какая партия шишек находится на просушивании. После получения зависимостей между входными параметрами и состояниями устройств в камере, было проведено исследование с помощью метода «замораживания» [2]. Суть метода заключается в варьировании значения одного параметра и фиксации значений всех других параметров. Для этой цели был выбран пример, в котором на просушивании была отправлена партия ели, влажность в камере постоянная – 50%, а температура возрастает от 25 до 55 градусов Цельсия.



**Рисунок 3. Значимость входных параметров нейросетевой модели**

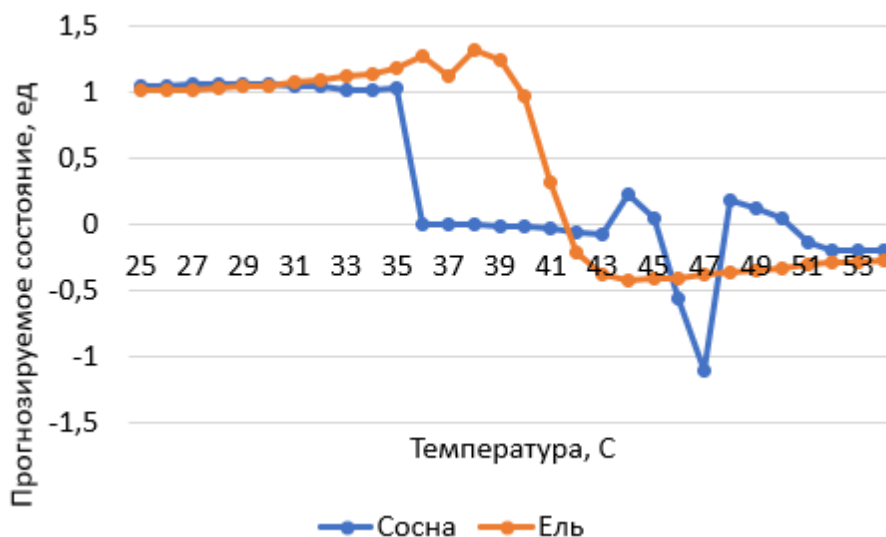
На рисунках 4 и 5 показаны графики зависимости прогнозируемого состояния калорифера и вытяжки от показателя температуры во время просушивания шишек ели и сосны.



**Рисунок 4. Зависимость прогнозируемого состояния калорифера от температуры**

Как можно заметить, в случае с елью, когда показатель температуры станет больше 27 градусов, сеть прогнозирует в дальнейшем непрерывную работу калорифера. Для семян сосны, судя по графику, его работа начинается при достижении температуры в 35 градусов Цельсия.

При тех же параметрах, в случае с вытяжкой, сеть прогнозирует другой режим работы. Начиная с 25 градусов Цельсия вытяжка должна будет работать в штатном режиме, но после того, как температура в камере с семенами ели поднимется до 41 градусов, то вытяжка прекращает свою работу. В случае с семенами сосны, то режим работы вытяжки должен будет смениться на выключенное состояние после 36 градусов Цельсия.



**Рисунок 5. Зависимость прогнозируемого состояния вытяжки от температуры**

Полученные результаты исследований не противоречат реальности, что подтверждает, что спроектированную нейронную сеть можно считать пригодной для автоматизации процесса просушивания семян шишек в сушильной камере.

**Заключение.** Построена нейросетевая система автоматизации управления сушильной камерой для просушивания семян шишек. Спроектированная нейросетевая модель учитывает 3 параметра: вид шишки, значение датчиков температуры и влажности в камере. Методом сценарного прогнозирования построены графики зависимостей состояний калорифера и вытяжки от изменения температуры. Применение такого набора параметров в модели позволяет точно прогнозировать состояние устройств в сушильной камере.

### Библиографический список

1. Daniel L Elliott, Russell E. Valentine, “Recurrent Neural Networks for Moisture Content Prediction in Seed Corn Dryer Buildings”, in 2011 IEEE 23rd International Conference on Tools with Artificial Intelligence, Boca Raton, FL, USA, November 7-9, 2011.
2. Ясницкий Л.Н. Введение в искусственный интеллект. М.: Издательский центр «Академия», 2005. – 176 с.



# NEURAL NETWORK CONTROL AUTOMATION SYSTEM OF THE DRYING CHAMBER FOR DRYING THE SEEDS OF CONES

*Ovchinnikova Adelina A.*

National Research University “Higher School of Economics”,  
Str. Gagarin Boulevard, 37a, Perm, Russia, 614060, deley@mail.ru

The article describes the development of a neural network system for automatic control of equipment in the drying chamber when drying the seeds of cones at a forest-seeding enterprise. This system allows the company's employees to accurately maintain temperature conditions for the disclosure of cones and the preservation of quality indicators of seed germination. Based on the drying time and the data received from temperature and humidity sensors, the neural network is able to determine at what point it is necessary to turn on or off the devices in the camera. With the help of the developed intellectual system, a study of the subject area was carried out, the natural conditions of drying seeds in the chamber, which are of practical importance, were revealed.

**Keywords:** artificial intelligence, neural network technologies, cones, forest seed industry, drying of cone seeds.

УДК 517.93; 519.83; 004.42

## ИНТЕЛЛЕКТУАЛЬНЫЙ ПРОГРАММНЫЙ КОМПЛЕКС МОДЕЛИРОВАНИЯ ПРОГНОЗИРОВАНИЯ И МИНИМАКСНОГО ОЦЕНИВАНИЯ ФАЗОВЫХ СОСТОЯНИЙ ДИСКРЕТНОЙ УПРАВЛЯЕМОЙ ДИНАМИЧЕСКОЙ СИСТЕМЫ

*Шориков Андрей Федорович*

Институт экономики УрО РАН,  
620014, Россия, г. Екатеринбург, ул. Московская, 29, afshorikov@mail.ru

В работе представлено описание структуры и функций интеллектуального компьютерного программного комплекса, позволяющего моделировать решение задач прогнозирования и минимаксного оценивания фазовых состояний линейной дискретной управляемой динамической системы. В статье приведено описание рассматриваемой дискретной управляемой динамической модели, функции основных модулей программного комплекса и описание разработанного пользовательского интерфейса. Основные модули программного комплекса позволяют моделировать решение задач формирования прогнозного множества рассматриваемой дискретной дина-

мической системы и его минимаксных оценок, а также реализовать удобный пользовательский интерфейс.

**Ключевые слова:** интеллектуальный программный комплекс, дискретная динамическая модель, прогнозирование, минимаксные оценки.

**Введение.** В статье представлено описание интеллектуального компьютерного программного комплекса, позволяющего моделировать решение задач прогнозирования фазовых состояний линейной дискретной управляемой динамической системы в заданный момент времени и формирования его минимаксных оценок. В основе этого программного комплекса – соответствующая дискретная управляемая динамическая математическая модель и разработанные алгоритмы построения прогнозного множества дискретной динамической системы и формирования его минимаксных оценок. Полученные в статье результаты примыкают к работам [1-3], основываются на результатах работ [4-6] и могут быть использованы для решения различных практических задач.

**Математическая модель исследуемого процесса.** Пусть на заданном целочисленном промежутке времени (далее просто – промежутке времени)  $\overline{0, T} = \{0, 1, \dots, T\}$  ( $T \in \mathbf{N}$ ;  $\mathbf{N}$  – множество всех натуральных чисел) рассматривается многошаговая управляемая система, динамика которой описывается векторно-матричным линейным дискретным рекуррентным уравнением вида

$$x(t+1) = A(t)x(t) + B(t)u(t), \quad x(0) = x_0, \quad t \in \overline{0, T-1}, \quad (1)$$

где  $x(t) = (x_1(t), x_2(t), \dots, x_n(t))'$  – фазовый вектор системы,  $x(t) \in \mathbf{R}^n$  ( $n \in \mathbf{N}$ ; здесь и далее, для  $k \in \mathbf{N}$ ,  $\mathbf{R}^k$  –  $k$ -мерное векторное пространство векторов-столбцов;  $x_0$  – заданное начальное значение фазового вектора), значения которого должны удовлетворять следующему заданному фазовому ограничению

$$\forall t \in \overline{0, T} : x(t) \in \mathbf{X}^*(t) \subset \mathbf{R}^n, \quad (2)$$

где  $\mathbf{X}^*(t)$  – непустой выпуклый многогранник-компакт (с конечным числом вершин) в пространстве  $\mathbf{R}^n$ , описываемый набором его вершин или конечной системой линейных алгебраических равенств и неравенств;  $u(t) = (u_1(t), u_2(t), \dots, u_p(t))'$  – вектор управляющего воздействия (управления),  $u(t) \in \mathbf{R}^p$  ( $p \in \mathbf{N}$ ), значения которого должны удовлетворять следующему заданному геометрическому ограничению

$$\forall t \in \overline{0, T-1} : u(t) \in \mathbf{U}^*(t) \subset \mathbf{R}^p, \quad (3)$$

где  $\mathbf{U}^*(t)$  – непустой выпуклый многогранник-компакт (с конечным числом вершин) в пространстве  $\mathbf{R}^p$ , описываемый набором его вершин или конечной системой линейных алгебраических равенств и неравенств.

Введем ряд определений, которые нужны для математической формализации задачи прогнозирования и минимаксного оценивания состояний рассматриваемой динамической системы (1) – (3).

Здесь и далее, для  $k \in \mathbf{N}$  и целочисленного промежутка времени  $\overline{\tau, \mathcal{G}}$  ( $\tau < \mathcal{G}$ ), символом  $\mathbf{S}_k(\overline{\tau, \mathcal{G}})$  будем обозначать метрическое пространство вектор-функций целочисленного аргумента (далее, просто – функций)  $\psi : \overline{\tau, \mathcal{G}} \rightarrow \mathbf{R}^k$ , в котором метрика  $\rho_k$  задается соотношением

$$\rho_k(\psi_1(\cdot), \psi_2(\cdot)) = \max_{t \in \overline{\tau, \mathcal{G}}} \|\psi_1(t) - \psi_2(t)\|_k \quad ((\psi_1(\cdot), \psi_2(\cdot)) \in \mathbf{S}_k(\overline{\tau, \mathcal{G}}) \times \mathbf{S}_k(\overline{\tau, \mathcal{G}})),$$

а символом  $\text{comp} \mathbf{S}_k(\overline{\tau, \mathcal{G}})$  – множество всех непустых и компактных, в смысле этой метрики, подмножеств пространства  $\mathbf{S}_k(\overline{\tau, \mathcal{G}})$  (здесь и далее, символом  $\|\cdot\|_k$  обозначается евклидова норма для элементов пространства  $\mathbf{R}^k$ ).

Используя ограничение (3), определим множество  $\mathbf{U}(\overline{\tau, \mathcal{G}}) \in \text{comp} \mathbf{S}_p(\overline{\tau, \mathcal{G}-1})$  ( $\tau < \mathcal{G}$ ) допустимых на промежутке времени  $\overline{\tau, T}$  программных управлений системы следующим соотношением

$$\mathbf{U}(\overline{\tau, \mathcal{G}}) = \left\{ u(\cdot) : u(\cdot) = \{u(t)\}_{t \in \overline{\tau, \mathcal{G}-1}} \in \mathbf{S}_p(\overline{\tau, \mathcal{G}-1}), \forall t \in \overline{\tau, \mathcal{G}-1}, u(t) \in \mathbf{U}^*(t) \right\}. \quad (4)$$

Назовем набор  $w(\tau) = \{\tau, x(\tau)\} \in \overline{0, T} \times \mathbf{R}^n$  ( $w(0) = w_0 = \{0, x_0\}$ )  $\tau$ -позицией дискретной управляемой динамической системы (1) – (3), или просто –  $\tau$ -позицией системы. Для каждого  $\tau \in \overline{0, T}$  определим также множество  $\mathbf{W}(\tau) = \{\tau\} \times \mathbf{X}^*(\tau)$  ( $\mathbf{W}(0) = \mathbf{W}_0 = \{w(0) = w_0 : w_0 = \{0, x_0\} \in \{0\} \times \mathbf{X}^*(0)\}$ ) всех допустимых  $\tau$ -позиций рассматриваемой системы.

Далее, пусть  $x(\cdot) = \varphi(\cdot; \overline{\tau, T}, x_\tau, u(\cdot))$  допустимая на промежутке времени  $\overline{\tau, T} \subseteq \overline{0, T}$  фазовая траектория системы (1), соответствующая допустимой паре  $(w(\tau), u(\cdot)) \in \mathbf{W}(\tau) \times \mathbf{U}(\overline{\tau, T})$  ( $w(\tau) = w_\tau = \{\tau, x_\tau\}$ ,  $w(0) = w_0 = \{0, x_0\}$ ), если  $\forall t \in \overline{\tau, T} : x(t) = \varphi(t; \overline{\tau, T}, x_\tau, u(\cdot)) \in \mathbf{X}^*(t)$ ,  $x(\tau) = x_\tau$ , где  $\varphi : \mathbf{R}^n \times \mathbf{S}_p(\overline{\tau, T-1}) \rightarrow \mathbf{S}_n(\overline{\tau, T})$  – оператор правой части системы (1), действующий на промежутке времени  $\overline{\tau, T}$ .

Обозначим символом  $\mathbf{X}(\cdot; \overline{\tau, T}, w_\tau, \mathbf{U}(\overline{\tau, T}))$  – трубку всех допустимых фазовых траекторий  $x(\cdot) = \varphi(\cdot; \overline{\tau, T}, x_\tau, u(\cdot))$  системы (1) (см., например, [1-4]), соответствующую паре  $(w_\tau, u_\tau(\cdot)) \in \mathbf{W}(\tau) \times \mathbf{U}(\overline{\tau, T})$ , а через  $\mathbf{X}(t; \overline{\tau, T}, w_\tau, \mathbf{U}(\overline{\tau, T}))$  обозначим ее сечение в период времени  $t \in \overline{\tau, T}$  ( $\mathbf{X}(t; \overline{\tau, T}, w_\tau, \mathbf{U}(\overline{\tau, T})) = \{x(t)\} = \{x_\tau\}$ ), которое называется *прогноznым множеством или областью достижимости* [1-4] рассматриваемой динамической системы в этот период времени. Отметим, что это есть множество всех допустимых фазовых векторов  $x(t)$ , соответствующих допустимой реализации  $\tau$ -позиции  $w(\tau) = w_\tau$ , т.е. это

множество тех, и только тех фазовых векторов  $x(t)$  системы, что для них существуют допустимые реализации программного управления  $u(\cdot) \in \mathbf{U}(\overline{\tau, T})$ , для которых справедливо:  $x(t) = \varphi(t; \overline{\tau, T}, x_\tau, u(\cdot)) \in \mathbf{X}(t; \overline{\tau, T}, w_\tau, \mathbf{U}(\overline{\tau, T}))$ .

Можно показать (аналогично рассуждениям, представленным в работе [4]), что для рассматриваемой динамической системы (1) – (3) каждое прогнозируемое множество  $\mathbf{X}(t; \overline{\tau, T}, w_\tau, \mathbf{U}(\overline{\tau, T}))$ ,  $t \in \overline{\tau + 1, T}$ , является выпуклым многогранником-компактом (с конечным числом вершин) пространства  $\mathbf{R}^n$ .

Для оценки качества процесса прогнозирования фазовых состояний рассматриваемой динамической системы (1) – (3) в период времени  $t \in \overline{\tau + 1, T}$  предлагается использовать минимаксные показатели (оценки) прогнозируемого множества  $\mathbf{X}(t; \overline{\tau, T}, w_\tau, \mathbf{U}(\overline{\tau, T}))$ : вектор  $x_t^{(e)} \in \mathbf{X}(t; \overline{\tau, T}, w_\tau, \mathbf{U}(\overline{\tau, T}))$  – его *чебышевский центр* и число  $r_t^{(e)} = r(\mathbf{X}(t; \overline{\tau, T}, w_\tau, \mathbf{U}(\overline{\tau, T})))$  – его *чебышевский радиус* (см., например, [4])

Тогда на основании сформированной модели (1) – (4) можно сформулировать следующую многошаговую задачу минимаксного оценивания прогнозируемого состояния рассматриваемой динамической системы (1) – (3).

**Задача.** Для заданных целочисленного промежутка времени  $\overline{0, T}$ , периода времени  $\tau$  ( $\tau \in \overline{0, T - 1}$ ) и  $\tau$ -позиции  $w(\tau) = \{\tau, x(\tau)\} = \{\tau, x_\tau\} = w_\tau \in \mathbf{W}(\tau)$  ( $w(0) = \{0, x(0)\} = \{0, x_0\} = w_0 \in \mathbf{W}_0$ ) дискретной управляемой динамической системы (1) – (3), для любого периода времени  $t \in \overline{\tau + 1, T}$  требуется сформировать прогнозируемое множество  $\mathbf{X}(t; \overline{\tau, T}, w_\tau, \mathbf{U}(\overline{\tau, T}))$  и его минимаксные оценки –  $x_t^{(e)} \in \mathbf{X}(t; \overline{\tau, T}, w_\tau, \mathbf{U}(\overline{\tau, T}))$  и  $r_t^{(e)} = r(\mathbf{X}(t; \overline{\tau, T}, w_\tau, \mathbf{U}(\overline{\tau, T})))$ , которые есть соответственно значение величины его чебышевского центра и чебышевского радиуса, путем реализации конечного числа одношаговых операций, допускающих их алгоритмизацию.

Основываясь на результатах работ [4-6], можно показать, что решение сформулированной задачи существует.

**Описание интеллектуального программного комплекса.** В рамках сформированной математической модели (1) – (4) в работах [4-6] формулируются нижеследующие задачи.

**Задача 1.** *Задача построения прогнозируемого множества* – нелинейная многошаговая задача формирования прогнозируемого множества всех допустимых фазовых состояний дискретной управляемой динамической системы (1) – (3) в заданный период времени.

**Задача 2.** *Задача формирования минимаксных оценок прогнозируемого множества* – нелинейная оптимизационная задача.

**Задача 3.** *Задача отображения результатов построения прогнозируемого множества и его минимаксных оценок.*

Математическая формализация и численные алгоритмы решения задач формирования прогнозируемого множества всех допустимых фазовых состояний дискретной управляемой динамической системы (1) – (3) в заданный период

времени и его минимаксных оценок основываются на результатах, опубликованных в работах [5,6] и представленных в монографии [4].

На основе формализованного описания и численных алгоритмов решения задач 1-3, сформулированных в рамках математической модели (1) – (4), в среде Delphi 7 создан моделирующий интеллектуальный компьютерный программный комплекс «Программный комплекс «PROGNOZ-2022» для реализации построения и минимаксного оценивания областей достижимости (прогнозных множеств) линейных дискретных управляемых динамических систем» (разработчики – Шориков А.Ф., Тюлюкин В.А.).

Данный программный комплекс позволяет реализовать формирование решений задач прогнозирования и минимаксного оценивания фазовых состояний рассматриваемой дискретной управляемой динамической системы (1) – (3). Для решения конкретной задачи пользователю предлагается произвести ввод исходных данных, определяющих все параметры системы (1). В программном комплексе предусмотрена возможность ввода параметров, описывающих ограничения (2), (3) на допустимые значения соответственно фазового вектора и вектора управляющих воздействий системы, а также количества периодов времени для реализации решения конкретной задачи прогнозирования и минимаксного оценивания.

Решение задачи формирования прогнозного множества (области достижимости) рассматриваемой динамической системы (1) – (3) основывается на *общем рекуррентном алгебраическом методе построения областей достижимости* (Шориков А.Ф. [4]) линейных дискретных управляемых динамических систем и сводится к реализации конечной последовательности решений только одношаговых задач линейного математического программирования и алгебраических операций над векторами в  $\mathbf{R}^n$  и  $\mathbf{R}^p$ . Алгоритм решения задачи формирования минимаксных оценок прогнозного множества основывается на *методе редукции к задаче выпуклого математического программирования* (Шориков А.Ф. [4]) и сводится к решению задачи выпуклого математического программирования и реализации алгебраических операций над векторами в пространстве  $\mathbf{R}^n$ .

Выходными результатами решения задач 1 и 2 – прогнозирования и минимаксного оценивания фазовых состояний дискретной управляемой динамической системы (1) – (3), является графическое и табличное описание сформированных прогнозного множества и его минимаксных оценок.

Перечислим функции основных модулей программного комплекса.

**Модуль А (Интерфейс)** – реализует интеллектуальный пользовательский интерфейс для выбора конкретной задачи, настройки входных данных, решения задачи и вывода данных, в удобной для пользователя форме.

**Модуль 1 (Прогнозирование)** – реализует настройку входных данных для конкретной задачи 1, решение задачи и вывод данных, в удобной для пользователя форме.

**Модуль 2 (Минимаксные оценки)** – реализует настройку входных данных для конкретной задачи 2, решение задачи и вывод данных, в удобной для пользователя форме.

Отметим, данный программный комплекс можно дополнять другими программными модулями и реализовать его сопряжение с различными корпоративными информационными системами. В настоящее время интеллектуальный программный комплекс «Программный комплекс «PROGNOZ-2022» для реализации построения и минимаксного оценивания областей достижимости (прогнозных множеств) линейных дискретных управляемых динамических систем» внедряется в практику работы ряда институтов РАН и учебных заведений.

**Заключение.** В работе представлено описание используемой в интеллектуальном программном комплексе дискретной управляемой динамической модели для формализации задач построения и минимаксного оценивания прогнозного множества, разработанных алгоритмов, функции его основных модулей и описание пользовательского интерфейса. Основные модули программного комплекса позволяют реализовать компьютерное моделирование решения задач прогнозирования и минимаксного оценивания фазовых состояний рассматриваемой дискретной управляемой динамической системы. Данный программный комплекс может использоваться в качестве инструментального средства для решения различных практических задач техники, экономики, медицины и др.

*Работа выполнена в соответствии с Планом НИР Института экономики УрО РАН.*

### **Библиографический список**

1. Красовский Н.Н. Теория управления движением. М.: Наука, 1968.
2. Красовский Н.Н., Субботин А.И. Позиционные дифференциальные игры. – М.: Наука, 1974.
3. Куржанский А.Б. Управление и наблюдение в условиях неопределенности. – М.: Наука, 1977.
4. Шориков А.Ф. Минимаксное оценивание и управление в дискретных динамических системах. Екатеринбург: Изд-во Урал. гос. ун-та, 1997.
5. Тюлюкин В.А., Шориков А.Ф. Об одном алгоритме построения области достижимости линейной управляемой системы // Негладкие задачи оптимизации и управление. Свердловск. УроАН СССР. 1988. С. 55-61.
6. Тюлюкин В.А., Шориков А.Ф. Алгоритм решения задачи терминального управления для линейной дискретной системы // Автоматика и телемеханика. 1993. № 4. С. 115-127.

# INTELLIGENT SOFTWARE PACKAGE FOR MODELING PREDICTING AND MINIMAX ESTIMATION OF PHASE STATES OF A DISCRETE-TIME CONTROLLED DYNAMICAL SYSTEM

*Shorikov Andrey F.*

Institute of Economics of the Ural Branch of the RAS  
Str. Moskovskaya, 29, Ekaterinburg, Russia, 620014, afshorikov@mail.ru

The article describes the structure and functions of an intelligent computer software package that allows modeling the solution of predicting problems and minimax estimation of phase states of a linear discrete-time controlled dynamical system. The article describes the discrete-time controlled dynamical model under consideration, the functions of the main modules of the software package and the description of the developed user interface. The main modules of the software package allow you to simulate the solution of the problems of forming the predictive set of the discrete dynamic system under consideration and its minimax estimates, as well as to implement a user-friendly interface.

**Keywords:** intelligent software complex, discrete-time dynamical model, predicting problem, minimax estimations.

УДК 004.032.26

## НЕЙРОСЕТЕВАЯ СИСТЕМА ПРОГНОЗИРОВАНИЯ ОДОБРЕНИЯ КРЕДИТА

*Брагин Константин Алексеевич*

Пермский государственный национальный исследовательский университет,  
ФИТ, 614990, Россия, г. Пермь, ул. Букирева, 15,  
bragkostiy99@gmail.com

В статье представлено описание разработки нейросетевой системы для прогнозирования одобрения кредита. Система позволяет с большой точностью предсказать, будет ли одобрен кредит в зависимости от финансового состояния и социальных характеристик заемщика.

**Ключевые слова:** искусственный интеллект, нейросетевые технологии, прогнозирование, кредит, финансовое состояние человека, показатель долговой нагрузки.

**Введение.** На сегодняшний день мировая индустрия кредитования находится в жестких условиях конкурентной борьбы за каждого клиента и вынуждена крайне стремительно реагировать на внешние изменения условий банковского кредитования.

Инвестиционно-банковский бизнес является одной из наиболее быстро развивающихся отраслей, на которую приходится 5% мирового внутреннего валового продукта и около 6% всех налоговых поступлений.

В современных условиях кредит является отличной возможностью улучшить качество своей жизни: обновить автомобиль, совершить путешествие по миру, сделать давно задуманный ремонт, приобрести multifunctionальную бытовую технику и многое другое. Говоря банковским языком, кредит – это ссуда в денежной или товарной форме, предоставляемая на условиях срочности, платности и возвратности [1].

Интересно и то, что просматривается зависимость между социальными характеристиками основной группы заемщиков (такими как возраст, социальное положение, пол, материальное положение, регион проживания) и целями потребительских кредитов [2].

Поэтому для поддержания конкурентоспособности потребительского кредитования, необходимо определиться с содержанием и последовательностью конкретных действий для достижения поставленных целей, что и определяет в широком смысле процесс планирования и прогнозирования деятельности банков. И если процесс планирования деятельности в большей мере ложится на плечи маркетингового отдела банка, то с задачей прогнозирования ему вполне могут помочь современные информационные технологии, такие как нейросетевое моделирование. Как показал опыт Пермской научной школы искусственного интеллекта [3], правильно обученная нейросетевая система может добиться большой точности в прогнозировании, что в свою очередь может в значительной мере увеличить прибыль как малоизвестным банкам, так и крупным банковским холдингам по всему миру, и, соответственно, повысить уровень предоставляемых банковских услуг.

При анализе исследуемой области выяснилось, что работ на тему исследования факторов, влияющих на одобрение потребительского кредита, достаточно много [4-7]. Так, О. С. Сычугова в своей работе [4] исследовали зависимости между социальными характеристиками и целями потребительского кредита, в одном из самых больших банков Альфа-банк. В своем исследовании они анализировали личный опыт относительно использования кредитов и математическим путем обосновали абсолютную нерациональность одобрения кредита.

Исследование авторы статьи проводили посредством текстовой аналитики и факторного анализа собственного опыта, в результате чего авторы сделали следующие выводы:

1. Три главных фактора, влияющих на *одобрение кредита*: наличие личного транспорта, высокий доход, брак.
2. Три главных фактора, влияющих на *отвержение кредита*: наличие непогашенных задолженностей, плохая кредитная история, низкий доход.

Основная цель настоящей работы заключается в сборе большого множества информации о одобрении или отказе в потребительском кредите, а также создание и обучение нейросетевой модели на этих данных. Конечным результатом должна быть нейросетевая система, способная прогнозировать одобрение кредита в больше чем 70% случаев.



Для создания нейросетевой системы были выбраны следующие параметры: X1 – тип кредита, X2 – ежемесячный платеж, X3 – сумма кредита, X4 – тип клиента, X5 – группа доходности, X6 – пол, X7 – наличие личного транспорта, X8 – недвижимость, X9 – количество детей, X10 – семья, X11 – тип дохода, X12 – образование, X13 – семейное положение, X14 – место жительства, X15 – возраст, X16 – клиент банка, X17 – время владения транспортом, X18 – профессия, X19 – количество членов семьи, X20 – решение банка.

Обучающее множество было взято из социальной сети специалистов по обработке данных и машинному обучению Kaggle [8] и включала выборку одобрения кредитов множества европейских банков. Перед переходом к проектированию и обучению нейросетевой модели, были выполнены следующие шаги:

1. очистка исходного множества от противоречивых примеров,
2. очистка исходного множества от выбросов по методике [11],
3. очистка исходного множества от дубликатов.

Например, некорректными примерами считались такие, в которых отсутствовала информация по одному из рассматриваемых параметров или же при наличии всех факторов, при которых одобрение кредита невозможно, кредит был одобрен.

Таким образом, объем итогового множества включает в себя 21 528 примеров одобрений и отказов в потребительском кредите. Данное множество было разделено на обучающее и тестовое множества в соотношении 70% и 30%, соответственно. Таким образом, в обучающее множество попало 15 070 примеров, а в тестовое – 6 458 примеров.

Проектирование, обучение и тестирование нейросетевой модели выполнялись с помощью программы «Нейросимулятор 5» [9] по методике [10]. В результате оптимизации спроектированная нейронная сеть представляла собой персептрон, который имеет *19 входных нейронов, один скрытый слой с 12-ю нейронами и 1 выходной нейрон.*

Средняя относительная ошибка тестирования составила 14,14%, что можно считать приемлемым результатом. На рисунке 1 представлена гистограмма, демонстрирующую разницу между фактическим и прогнозируемым нейросетью результатом.

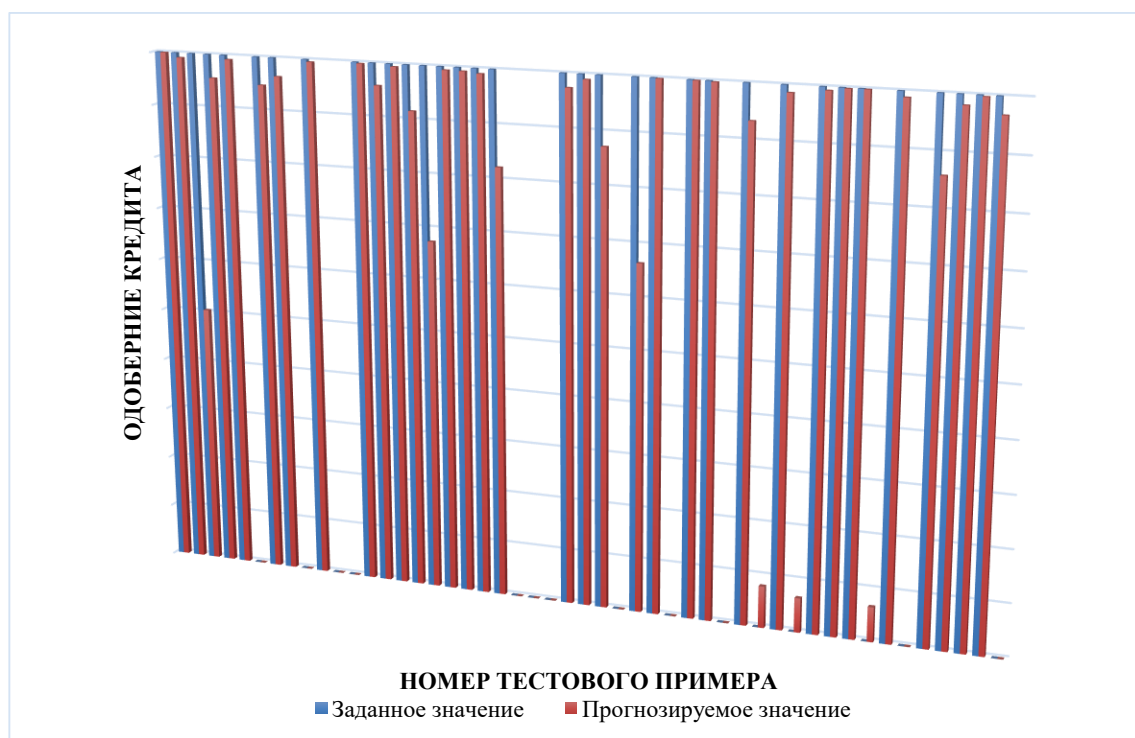
Из результатов, изображенных на рисунке 1, можно сделать вывод об адекватной работе спроектированной нейронной сети.

Оценка значимости входных параметров была выполнена с помощью программы «Нейросимулятор 5», результат отображен на рисунке 2.

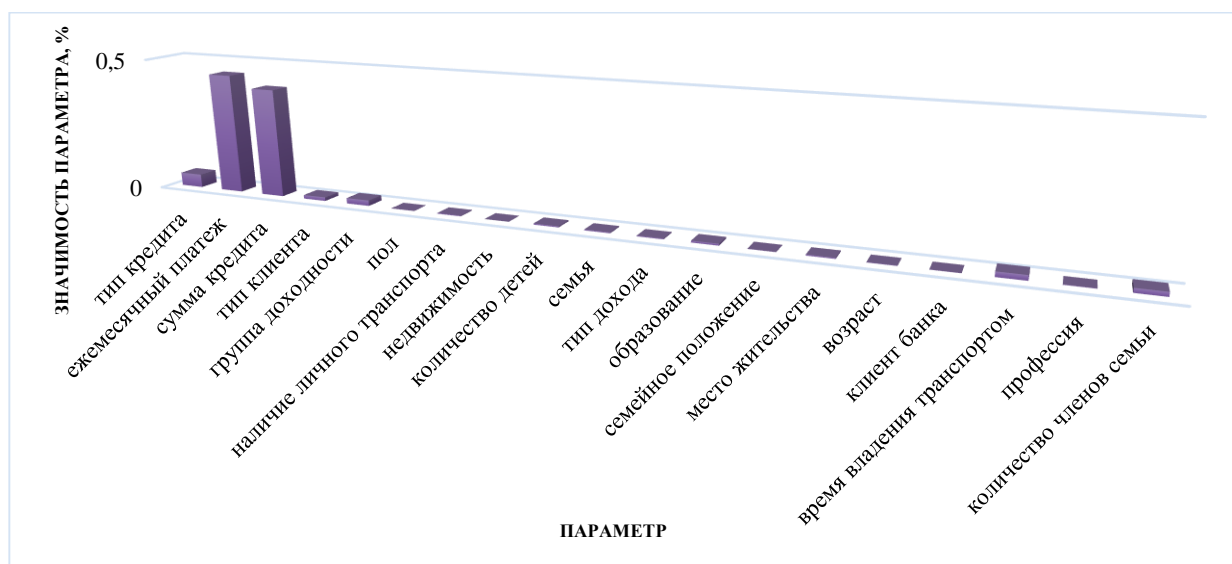
На основании данных, продемонстрированных на рисунке 2, наиболее значимыми параметрами нейросетевой модели являются X1 (тип кредита), X2 (ежемесячный платеж), X3 (сумма кредита). Наиболее значимым параметром является ежемесячный платеж, который сможет платить заемщик, что вполне ожидаемо, ведь банку не выгодно выдавать кредит без возврата денежных средств вовремя.

Далее было проведено исследование полученных зависимостей между входными параметрами и удовлетворенностью постояльцев отелей. Исследование проводилось с помощью метода «замораживания» [10], суть которого

заключается в варьировании значения одного параметра и фиксировании значений всех других параметров. Для этой цели был отобран «нейтральный пример», про который нейросеть не может с уверенностью сказать, будет ли одобрен кредит заемщику.

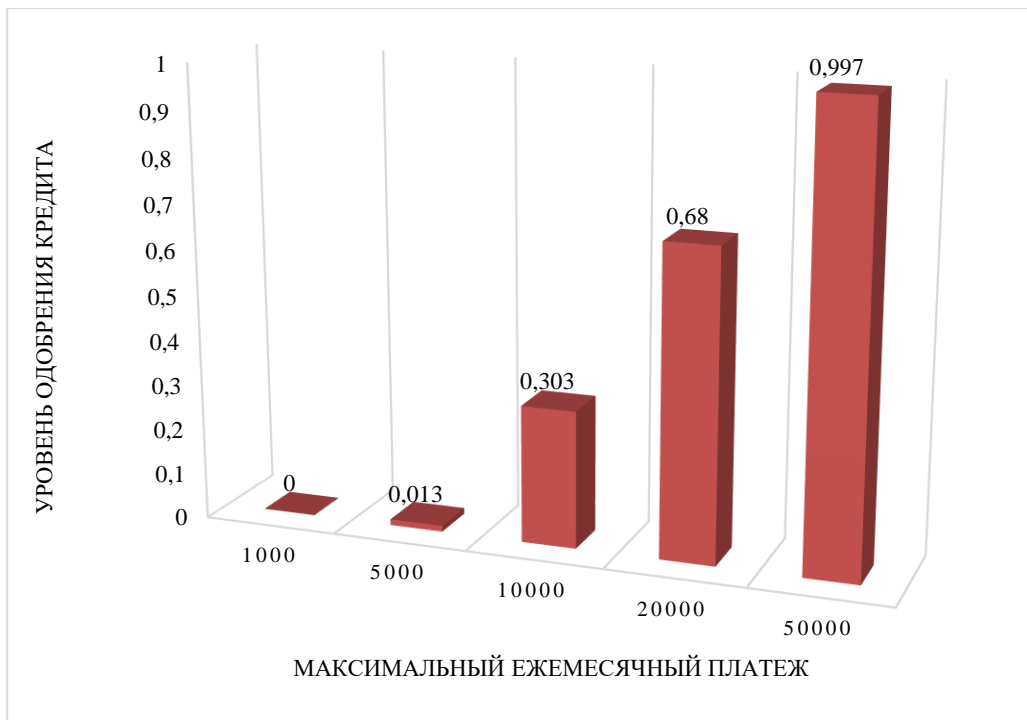


**Рисунок 1. Результат тестирования обученной нейросетевой модели**



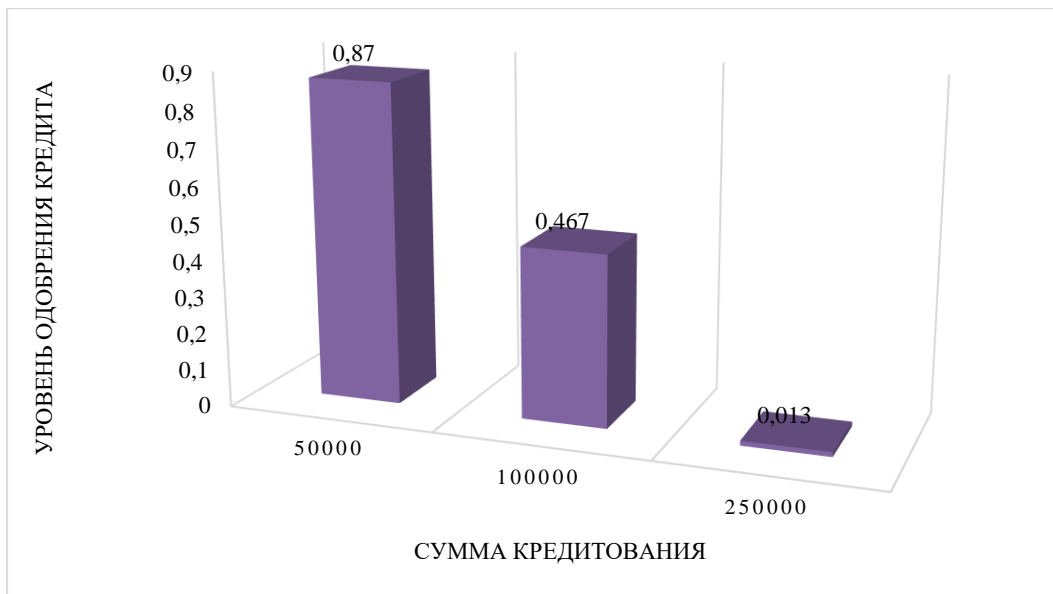
**Рисунок 2. Значимость параметров нейросетевой модели**

На рисунке 3 показан график зависимости прогнозируемого значения одобрения кредита от их возможности выплачивать кредит. Заметна тенденция, подтверждающая высокую значимость данного параметра в спроектированной нейросетевой модели: чем выше ежемесячный платеж, который сможет выплачивать клиент, тем выше уровень одобрения кредита.



**Рисунок 3. Зависимость прогнозируемого значения от максимального ежемесячного платежа**

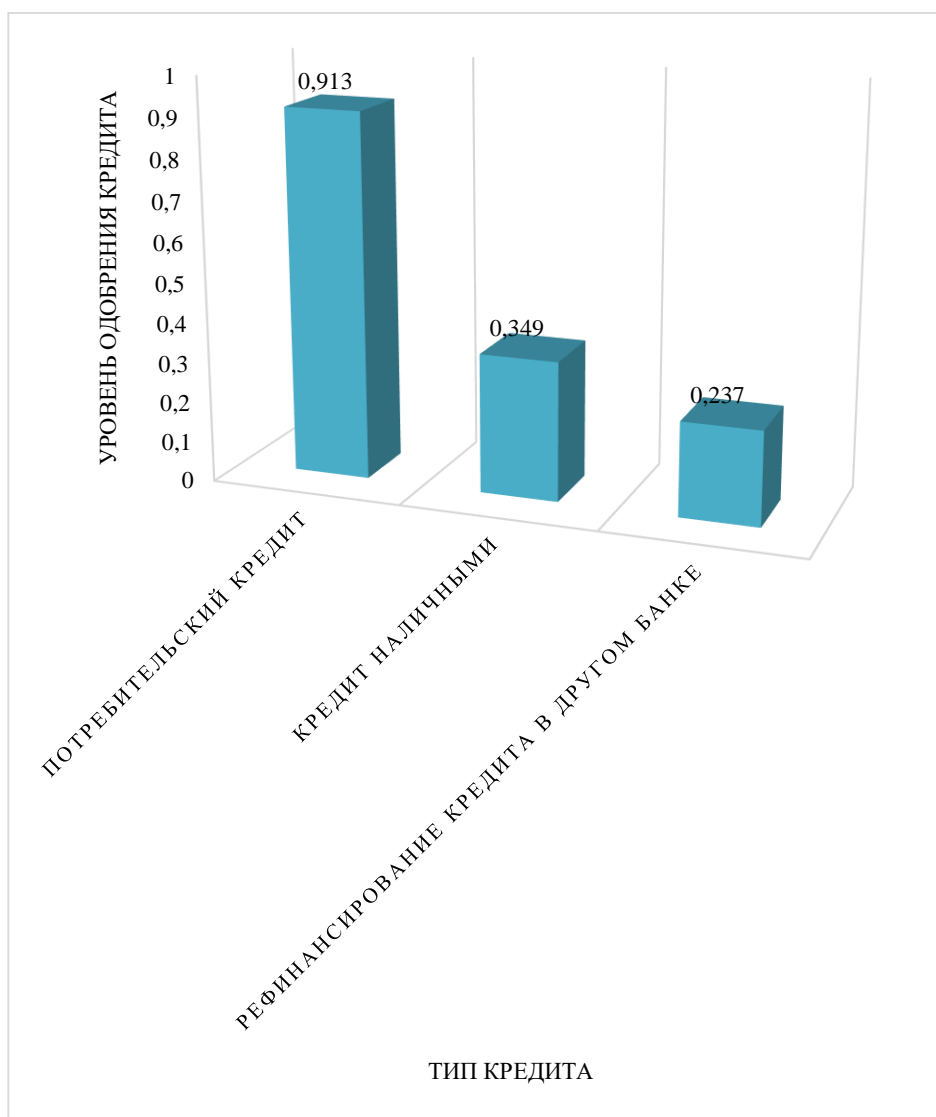
На рисунке 4 продемонстрирована зависимость прогнозируемого уровня одобрения кредита от суммы кредита. По результатам исследования выяснилось, что банки одобряли кредит заемщикам меньшей суммой кредитования гораздо чаще, в то же время обратная ситуация наблюдается с большими суммами, при которых одобрение кредита маловероятно.



**Рисунок 4. Зависимость прогнозируемого значения от суммы кредитования**

На рисунке 5 продемонстрирована зависимость прогнозируемого уровня одобрения кредита от типа кредита. Выдаваемые кредиты бывают потребительские, кредит наличными и рефинансирование кредита в другом банке.

Такие результаты могут натолкнуть на следующую мысль: банкам не выгодно выдавать кредит наличными, вероятно это происходит по причине более трудоемкого оборота наличных средств. Также банки достаточно редко одобряют рефинансирования кредитов других банков, возможно, это происходит по причине недоверия к людям, которые неспособны выплатить кредит в другом банке. В тоже время, потребительские кредиты достаточно легко получить, т.к. сумма таких займов обычно не большая.



**Рисунок 5. Зависимость прогнозируемого значения от типа кредита**

Полученные результаты исследований не противоречат реальности, что подтверждает, что спроектированную нейронную сеть можно считать пригодной для прогнозирования одобрения кредитов заемщиков в банках по всему миру.

**Заключение.** Построена система нейросетевого прогнозирования одобрения потребительских кредитов. Спроектированная нейросетевая модель учитывает 19 параметров, таких как: тип кредита, ежемесячный платеж, сумма кредита, тип клиента, группа доходности, пол, наличие личного транспорта, недвижимость, количество детей, семья, тип дохода, образова-

ние, семейное положение, место жительства, возраст, клиент банка, время владения транспортом, профессия, количество членов семьи. Методом сценарного прогнозирования построены графики зависимостей прогнозируемого одобрения кредитов от изменения входных параметров. Применение такого набора параметров в модели позволяет с высокой точностью прогнозировать одобрение кредитов заемщиков. В процессе исследования выявлен параметр, имеющий наибольшую значимость – максимальный ежемесячный платеж.

### Библиографический список

1. Тамбиева, Х. М. Индивидуальный подход как средство привлечения клиентов в процессе ипотечного кредитования / Х. М. Тамбиева // Мир науки, культуры, образования. – 2015. – № 3(52). – С. 265-267.
2. Киреева О.В., Педанова Е.Ю., Дёмин А.Н. Отношение к кредитам и социальный статус личности // Южно-российский журнал социальных наук. 2017. №3. URL: <https://cyberleninka.ru/article/n/otnoshenie-k-kreditam-i-sotsialnyu-status-lichnosti> (дата обращения: 23.12.2021).
3. Ясницкий Л.Н., Бржевская А.С., Черепанов Ф.М. О возможностях применения методов искусственного интеллекта в сфере туризма // Сервис plus. – 2010 – №4. – С.111-115.
4. Сычугова, О. С. Анализ практики применения и сравнительная оценка экономической эффективности потребительского и ипотечного кредитования в современных условиях / О. С. Сычугова // Молодежь и наука. – 2014. – № 2. – С. 40.
5. Показатель долговой нагрузки как финансовый элемент кредитования / А. А. Навасардян, О. И. Хамзина, Е. В. Банникова, В. В. Евстафьева // Финансовая экономика. – 2019. – № 12. – С. 69-71.
6. Долгих, И. А. Анализ и оценка кредитоспособности заемщика / И. А. Долгих // Роль и место информационных технологий в современной науке : сборник статей Международной научно-практической конференции, Самара, 17 января 2019 года. – Самара: Общество с ограниченной ответственностью "ОМЕГА САЙНС", 2019. – С. 63-68.
7. Разоренная, А. А. К вопросу о платежеспособности и кредитоспособности заемщика / А. А. Разоренная, М. Р. Тахумов // Школа молодых новаторов : Сборник научных статей международной молодежной научной конференции. В 2-х томах, Курск, 19 июня 2020 года. – Курск: Юго-Западный государственный университет, 2020. – С. 217-221.
8. Kaggle – социальная сеть специалистов по обработке данных и машинному обучению. [Электронный ресурс]. Режим доступа: <https://www.kaggle.com>.
9. Черепанов Ф.М., Ясницкий Л.Н. Нейросимулятор 4.0. Свидетельство о регистрации программы для ЭВМ RUS 2014612546. Заявка № 2014610341 от 15.01.2014.
10. Ясницкий Л.Н. Интеллектуальные системы: учебник. М.: Лаборатория знаний, 2016. 221 с.

11. Yasnitsky L.N. Algorithm for searching and analyzing abnormal observations of statistical information based on the Arnold – Kolmogorov – Hecht-Nielsen theorem // International Journal of Advanced Trends in Computer Science and Engineering. 2020. 9(2). Pp. 1814-1819. DOI: 10.30534/ijatcse/2020/139922020

## EURAL NETWORK CREDIT APPROVAL PREDICTION SYSTEM

*Bragin Konstantin A.*

Perm State University,

614990, Russia, Perm, st. Bukireva, 15, bragkostiy99@gmail.com

The article describes the development of a neural network system for predicting loan approval. The system allows you to predict with high accuracy whether a loan will be approved depending on the financial condition and social characteristics of the borrower.

**Keywords:** artificial intelligence, neural network technologies, forecasting, credit, financial condition of a person, debt load indicator.

УДК 681.5.01

## АДАПТИВНОЕ ГРУППОВОЕ УПРАВЛЕНИЕ ПЕРЕОПРЕДЕЛЕННЫМ ОБЪЕКТОМ НА ПРИМЕРЕ КАМЕРЫ СГОРАНИЯ ГАЗОТУРБИННОГО ДВИГАТЕЛЯ

*Южаков Александр Анатольевич*

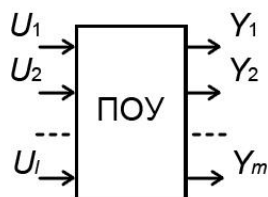
*Сторожев Сергей Александрович*

Пермский национальный исследовательский политехнический университет,  
614000, Россия, г. Пермь, ул. Комсомольский проспект, 29,  
sepra5@mail.ru

Рассматривается способ адаптивного группового управления переопределенным объектом. Разработан адаптивный нечеткий групповой регулятор системы автоматического управления подачей топлива в камеру сгорания газотурбинного двигателя, который позволяет учитывать значения сразу нескольких каналов управления при формировании управляющего воздействия с учетом изменяющихся условий эксплуатации газотурбинного двигателя.

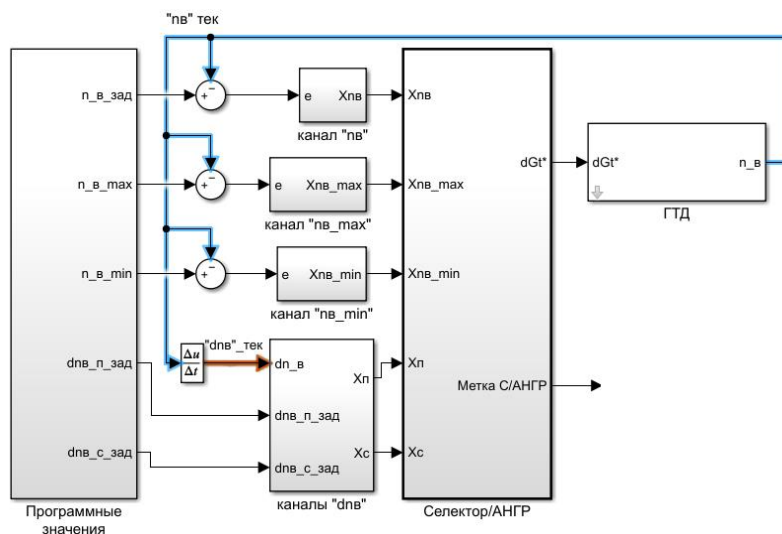
**Ключевые слова:** переопределенный объект управления, нечеткий групповой регулятор; метод взвешенного среднего; фаззификатор; дефаззификатор; газотурбинный двигатель; камера сгорания, условия эксплуатации.

**Введение.** Объекты управления, у которых число управляющих воздействий  $l$  меньше чем число управляемых координат  $m$  называются переопределенными (см. рис. 1) [1].



**Рисунок 1. Переопределенный объект управления**

Системы автоматического управления (САУ) переопределенными объектами должны решать задачу выработки необходимых управляющих воздействий для соблюдения требований по точности и надежности. Примером является САУ подачи топлива в камеру сгорания газотурбинного двигателя (ГТД) [2], представленная на рисунке 2. Она содержит одно управляющее воздействие – количество топлива в камере сгорания «Gт» и две управляемых координаты – частота оборотов вентилятора «*пв*» и производная оборотов вентилятора «*дпв*».



**Рисунок 2. Структурная схема САУ подачи топлива в камеру сгорания ГТД с селектором**

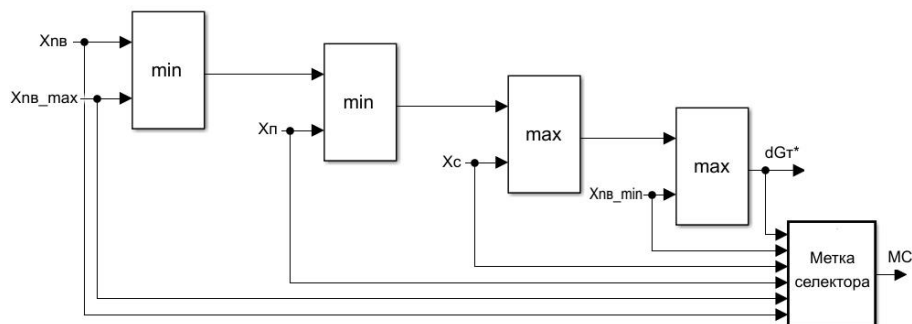
Здесь: «канал «пв»», «канал «пв\_max»», «канал «пв\_min»», «каналы «дпв»» – каналы управления САУ ГТД; «Селектор/АНГР» – устройство формирования управляющего воздействия на дозатор ГТД

Для формирования управляющего воздействия используются 5 каналов управления:

- 1 – канал частоты вращения вентилятора «пв» (основной регулятор режима);
- 2 – канал максимальной частоты вращения вентилятора «пв\_max» (канал ограничения по максимуму);

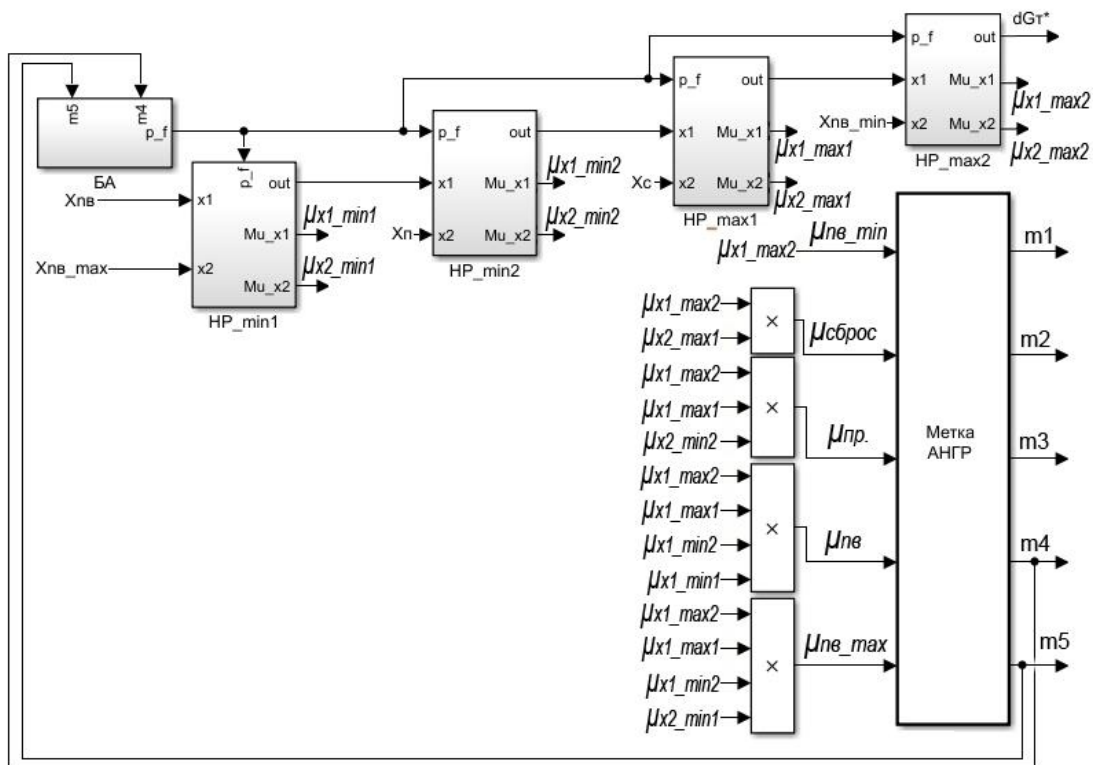
- 3 – канал минимальной частоты вращения вентилятора «пв\_min» (канал ограничения по минимуму);
- 4 – канал приемистости ускорения частоты вращения вентилятора «dпв\_п» (канал разгона);
- 5 – канал сброса ускорения частоты вращения вентилятора «dпв\_с» (канал сброса).

Селектор формирует управляющее воздействие (см. рис. 3).



**Рисунок 3. Структурная схема селектора**

Здесь: «Xпв», «Xпв\_max», «Xпв\_min», «Xп», «Xс» – управляющие воздействия каналов управления; «min», «max» – элементы на базе многозначной логики; «dГт\*» – управляющее воздействие на дозатор топлива в камеру сгорания ГТД



**Рисунок 4. Структурная схема адаптивного нечеткого группового регулятора**

Здесь «Xпв», «Xпв\_max», «Xпв\_min», «Xп», «Xс» – управляющие воздействия каналов управления; «HP\_min», «HP\_max» – нечеткие регуляторы



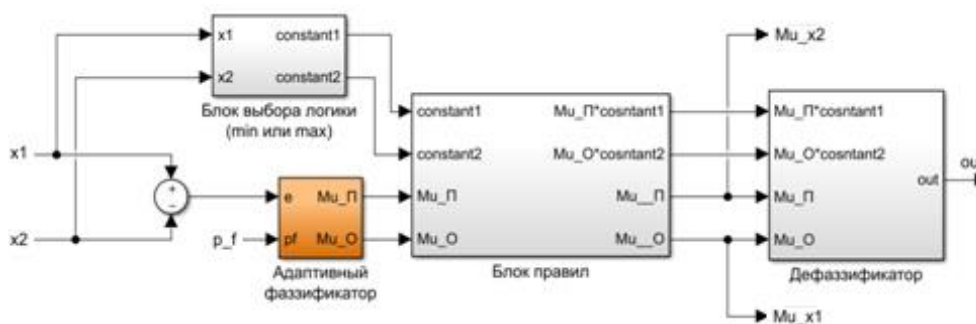
по минимуму и по максимуму с учетом адаптации; «dGт\*» – выход нечеткого адаптивного группового регулятора; «m4», «m5» – индикаторы работы каналов приемистости и сброса; «БА» – блок адаптации; «р\_f» – параметр адаптации нечетких регуляторов

Выбранный селектором канал управления вступает в работу, а обратные связи остальных каналов разрываются, что является недостатком использования многозначной логики. Еще одним недостатком селектора является отсутствие адаптации к изменяющимся условиям, т.е. отсутствует согласование между каналами управления.

Вышеперечисленные недостатки устраняются с помощью применения адаптивного нечеткого группового регулятора (АНГР) взамен селектора.

**Адаптивный нечеткий групповой регулятор.** Он состоит из нечетких регуляторов по минимуму и по максимуму и блока адаптации (см. рис. 4).

Структура нечетких регуляторов по минимуму и по максимуму приведена на рисунке 5.



**Рисунок 5. Структурная схема нечеткого регулятора по минимуму (по максимуму) с учетом адаптации**

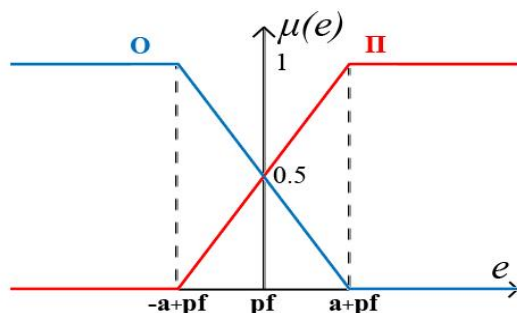
Где «x1», «x2» – входы нечеткого регулятора; «e» – отклонение входа «x2» от входа «x1»; «Mu\_P» – степень принадлежности к терм-множеству «П»; «рf» – значение параметра для адаптации фаззификатора; «Mu\_O» – степень принадлежности к терм-множеству «О»; «constant1», «constant2» – выходы правил; «Mu\_x1», «Mu\_x2» – значения степеней принадлежности фаззификатора для нечеткого регулятора по минимуму (для нечеткого регулятора по максимуму они меняются местами); «out» – выход нечеткого регулятора

Графическое представление фаззификатора представлен на рисунке 6.

Математическое описание терм-множеств «О» и «П» приведено в (1) [3].

$$\mu_P(e, a, pf) = \begin{cases} 0, & e \leq -a + pf; \\ \frac{e + a - pf}{2 \cdot a}, & -a + pf < e < a + pf; \\ 1, & e \geq a + pf; \end{cases}$$

$$\mu_O(e, a) = \begin{cases} 1, & e \leq -a + pf; \\ \frac{a + pf - e}{2 \cdot a}, & -a + pf < e < a + pf; \\ 0, & e \geq a + pf. \end{cases} \quad (1)$$



**Рисунок 6. Адаптивный фаззификатор нечеткого регулятора:**  
 «e» – отклонение входа «x2» от входа «x1»; «μ(e)» – степень принадлежности;  
 «O», «П» – терм-множества; «a» – параметр нечеткого регулятора;  
 «pf» – значение сформированное блоком адаптации

Блок правил для нечеткого регулятора по минимуму:

1) Если «e» (разность «x1» и «x2») положительная, то выход равен «constant1» (т.е. «x2») (Если «e» = «П», то «x2»)

2) Если «e» отрицательная, то выход равен «constant2» (т.е. «x1») (Если «e» = «O», то «x1»)

Блок правил для нечеткого регулятора по максимуму:

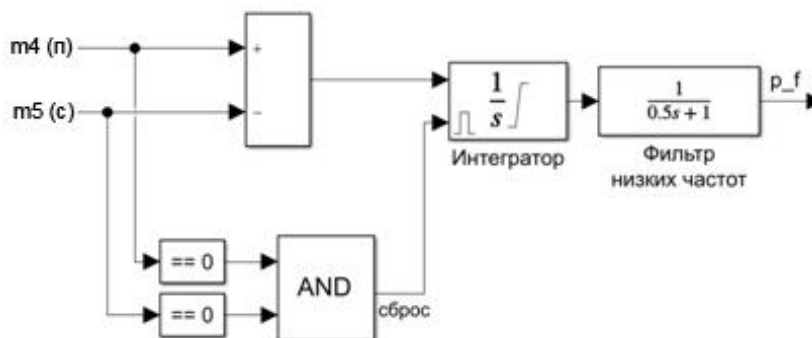
1) Если «e» положительная, то выход равен «constant1» (т.е. «x1») (Если «e» = «П», то «x1»)

2) Если «e» отрицательная, то выход равен «constant2» (т.е. «x2») (Если «e» = «O», то «x2»)

Значения из блока правил поступают на дефаззификатор. Дефаззификация происходит по методу среднего взвешенного (2) [3].

$$out = \frac{\mu_{П} \cdot constant1 + \mu_{O} \cdot constant2}{\mu_{П} + \mu_{O}} \quad (2)$$

На рисунке 7 приведен блок адаптации.



**Рисунок 7. Структурная схема блока адаптации,**  
 где «Mr», «Ms» – индикаторы работы контуров приемистости  
 и сброса; «rf» – значение сдвига терм-множеств

Во время работы контура приемистости (см. «Mr») терм-множества сдвигаются вправо (знак «+»), а во время работы контура сброса (см. «Ms») – влево (знак «-»). Интегратор имеет ограничения минимального и максимального зна-

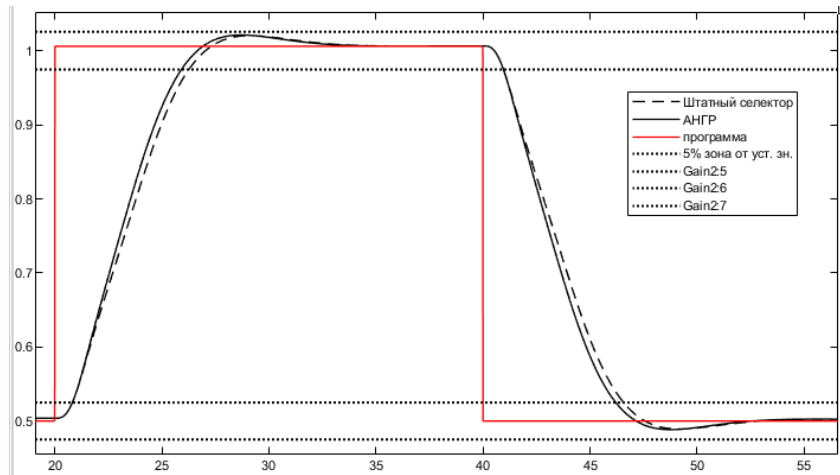
чений  $\pm l$  (крайнее левое и крайнее правое положения терм-множеств). Сигнал сброса формируется в момент отключения контуров приемистости и сброса. Фильтр обеспечивает плавное перемещение терм-множеств.

Результаты эксперимента. Был проведен эксперимент с САУ подачей топлива в камеру сгорания ГТД с селектором и АНГР. Для определения качества управления построена следующая целевая функция (3) [4]:

$$F = A \cdot (\sum \dot{t}_1 + \sum \dot{t}_2) + B \cdot (\sum \dot{\sigma}_1 \% + \sum \dot{\sigma}_2 \%) \rightarrow \min \quad (3)$$

где  $\dot{t}_1, \dot{\sigma}_1 \%$  – нормированные значения показателей качества при первом изменении программного значения оборотов вентилятора «n\_в»;  $\dot{t}_2, \dot{\sigma}_2 \%$  – нормированные значения показателей качества при втором изменении программного значения оборотов вентилятора «n\_в»,  $A$  и  $B$  – весовые коэффициенты.

На рисунке 8 приведены результаты эксперимента



**Рисунок 8. Сравнение оборотов вентилятора «n\_в» САУ ГТД с селектором и АНГР**

Показатели качества САУ ГТД с селектором:  $t_1^C = 26.26 - 20 = 6.26$  с;

$$t_2^C = 46.57 - 40 = 6.57 \text{ с}; \sigma_1^C \% = \frac{|h_{max} - h_{ycm}|}{dh} = \frac{|1.021 - 1|}{0.5} = 4.2\%;$$

$$\sigma_2^C \% = \frac{|h_{min} - h_{ycm}|}{dh} = \frac{|0.489 - 0.5|}{0.5} = 2.2\%;$$

Показатели качества САУ ГТД с АНГР:  $t_1^{АНГР} = 25.91 - 20 = 5.91$  с;

$$t_2^{АНГР} = 46.20 - 40 = 6.20 \text{ с}; \sigma_1^{АНГР} \% = \frac{|h_{max} - h_{ycm}|}{dh} = \frac{|1.02135 - 1|}{0.5} = 4.27\%;$$

$$\sigma_2^{АНГР} \% = \frac{|h_{min} - h_{ycm}|}{dh} = \frac{|0.4890 - 0.5|}{0.5} = 2.26\%;$$

АНГР уменьшает время переходного процесса на приемистости на 0.35 с и на сбросе на 0.37 с, при этом увеличивается перерегулирование на приемистости на 0.07% и на сбросе на 0.06% по сравнению с селектором.

Результаты сведены в таблицу и нормированы.

Регулятор	$t_1$	$\sigma_1$	$t_2$	$\sigma_2$
Селектор	6,26	4,2	6,57	2,2
АНГР	5,91	4,27	6,2	2,26
<b>max</b>	<b>6,26</b>	<b>4,27</b>	<b>6,57</b>	<b>2,26</b>
Селектор(Н)	1	0,983607	1	0,973451
АНГР(Н)	0,944089	1	0,943683	1

По значениям целевой функции ( $F_{\text{селектор}}=1.98$ ;  $F_{\text{АНГР}}=1.94$ ), АНГР улучшает качество управления на 2%.

### Заключение

1. Предложен новый метод адаптации при проектировании нечеткого группового регулятора подачи топлива в камеру сгорания ГТД.
2. Адаптивный нечеткий групповой регулятор повышает качество управления на 2% по сравнению с селектором.
3. Адаптивный нечеткий групповой регулятор дополняет теорию нечетких систем [2, 5].

### Библиографический список

1. Цыпкин, Я. З. Адаптация и обучение в автоматических системах / Я. З. Цыпкин. – М.: Наука, 1968. – 400 с.
2. Гостев В.И. Нечеткие регуляторы в системах автоматического управления. –К.: Радиоматор, 2008. – 972с.
3. Гинзбург, С. А. Математическая непрерывная логика и изображение функций / С. А. Гинзбург. – М.: Энергия, 1968. – 136 с.
4. Андриевская Н.В., Андриевский О.А., Леготкина Т.С., Хижняков Ю.Н., Сторожев А.А., Никулин В.С., Южаков А.А., Кузнецов М.Д. Нейронно-нечеткое управление выбросами вредных веществ авиационного газотурбинного двигателя // Мехатроника. Автоматизация. Управление. 2020. Т. 2. № 6. С. 348 – 355.
5. Хижняков Ю.Н. Нечеткое, нейронное и гибридное управление: учеб. пособие / Ю.Н. Хижняков. – Пермь: Изд-во Перм. нац. исслед. политехн. ун-та, 2013. – 303 с.

## ADAPTIVE GROUP CONTROL OF A REDETERMINATED OBJECT ON THE EXAMPLE OF A GAS TURBINE ENGINE COMBUSTION CHAMBER

*Yuzhakov Alexander A., Storozhev Sergey A.*

Perm National Research Polytechnic University

Str. Komsomolsky prospect, 29, Perm, Russia, 614000, cepra5@mail.ru

A method for adaptive group control of an overridden object is considered. An adaptive fuzzy group controller of the automatic control system for fuel

supply to the combustion chamber of a gas turbine engine has been developed, which allows taking into account the values of several control channels at once when forming a control action, taking into account the changing operating conditions of a gas turbine engine.

**Keywords:** redefined control object, fuzzy group controller; weighted average method; fuzzifier; defuzzifier; gas turbine engine; combustion chamber, operating conditions.

УДК 681.5.01

## ВЫБОР И ОПТИМИЗАЦИЯ ПАРАМЕТРОВ РЕГУЛЯТОРА ЭЛЕМЕНТА ПЕЛЬТЬЕ БЛОКА БУЦ, РАЗРАБОТАННОГО НА ОТЕЧЕСТВЕННОЙ ЭЛЕМЕНТНОЙ БАЗЕ

*Соколов Дмитрий Александрович,  
Нацанов Максим Анварович,  
Загвозкин Владимир Дмитриевич,  
Хижняков Юрий Николаевич*

Пермский национальный исследовательский политехнический университет,  
614000, Россия, г. Пермь, ул. Комсомольский проспект, 29,  
forstudy01@mail.ru

Современные проблемы автоматизации сложных нелинейных (недетерминированных) объектов связаны с отсутствием их математического описания. Как правило, недетерминированные объекты разного класса требуют разного рода управления. В нашем случае регулирование температуры лазерного диода осуществляется элементом Пельтье. Ввиду разработки блока БУЦ на отечественной элементной базе встает вопрос о выборе и оптимизации параметров регулятора элемента Пельтье.

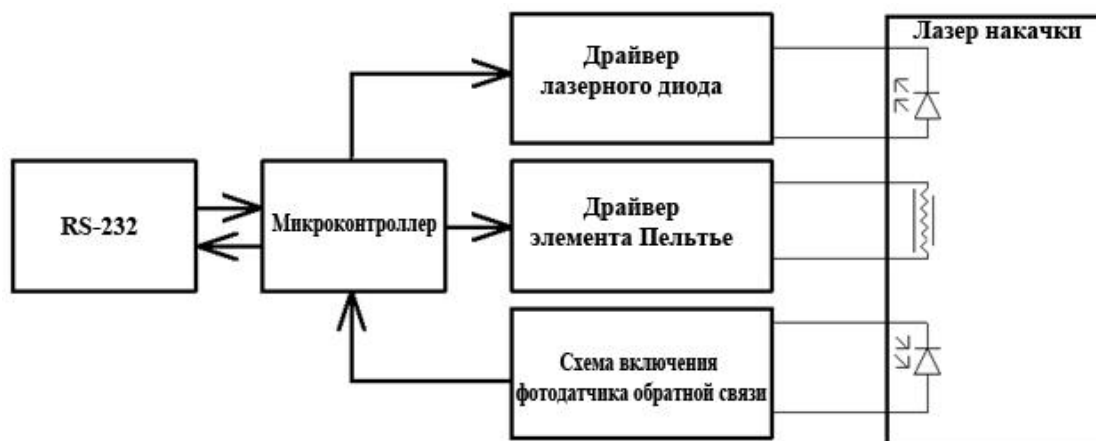
**Ключевые слова:** элемент Пельтье, регулятор, оптимизация.

**Введение.** Ниже будут рассматриваться: релейный регулятор, И-регулятор, ПИ-регулятор, ПИД-регулятор, нечеткий регулятор с П составляющей, нечеткий регулятор с ПД составляющими для термостабилизации температуры кристалла лазерного диода (ЛД) при помощи элемента Пельтье встроенного в лазерный диодный модуль посредством блока управления цифрового (БУЦ) усилителя спонтанной эмиссии (УСЭМ).

УСЭМ – квантовый волоконно-оптический генератор, собранный на основе эрбиевого усилителя с цепью оптической положительной обратной связи. [1] Применяется в волоконно-оптических линиях передачи для восстановления уровня оптического сигнала. Преимуществом эрбиевых усилителей является отсутствие преобразования в электрический сигнал, возможность одновременного усиления сигналов с разными длинами волн, что обуславли-

вадет возможность усиления спектрально-мультиплексированного сигнала, практически точное соответствие рабочего диапазона эрбиевых усилителей области минимальных оптических потерь световодов на основе кварцевого стекла, сравнительно низкий уровень шума и простота включения в волоконно-оптическую систему. [2]

На рисунке 1 приведена функциональная схема БУЦ. Регулятор лазерного диода блока БУЦ реализован в микроконтроллере в виде программного кода.



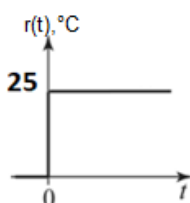
**Рисунок 1. Функциональная схема БУЦ**

Элемент Пельтье — это термоэлектрический преобразователь, принцип действия которого основан на эффекте Пельтье — возникновении разности температур при протекании электрического тока. [3]

В основе работы элементов Пельтье лежит контакт двух полупроводниковых материалов с разными уровнями энергии электронов в зоне проводимости. При протекании тока через контакт таких материалов электрон должен приобрести энергию, чтобы перейти в более высокоэнергетическую зону проводимости другого полупроводника. При поглощении этой энергии происходит охлаждение места контакта полупроводников. При протекании тока в обратном направлении происходит нагревание места контакта полупроводников, дополнительно к обычному тепловому эффекту. [3]

**Постановка эксперимента.** Был поставлен ряд экспериментов, нацеленных на выявление показателей качества описанных выше регуляторов, вид задающего воздействия – ступенчатое (рисунок 2). Эксперименты были осуществлены посредством фиксирования изменений значений регуляторов на четырех различных промежутках температур:

1. От 0 до 25 градусов:

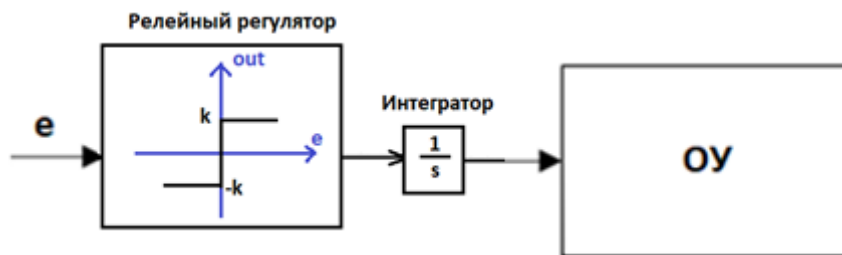


**Рисунок 2. Задающее воздействие, подаваемое на вход регуляторов (0-25)**

Были использованы следующие типы регуляторов:

Релейный регулятор.

Структурная схема приведена на рисунке 3.

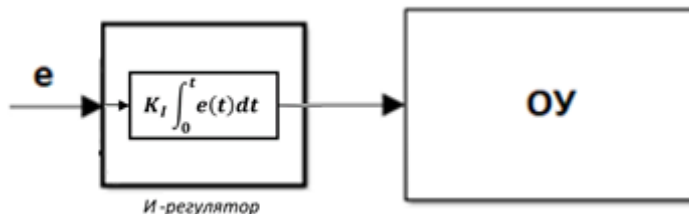


**Рисунок 3. Структурная схема САУ с релейным регулятором**

Итогом проведенных экспериментов стали данные зафиксированные в таблицах, а также построенные по ним графики и на основании целевых функций, были найдены наиболее оптимальные коэффициенты релейного регулятора:  $K_{\text{наг.}} - 0.0019$ ,  $K_{\text{охл.}} - 0.0019$ .

И-регулятор.

Структурная схема приведена на рисунке 4.

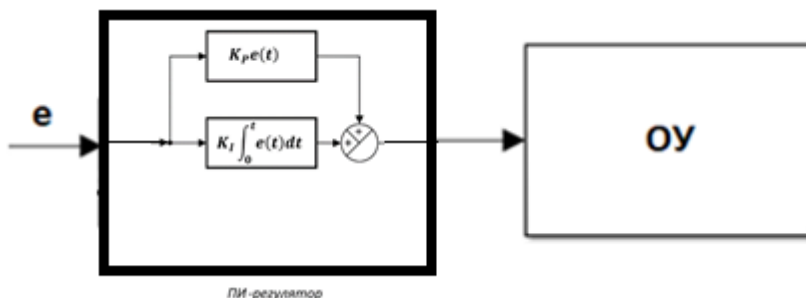


**Рисунок 4. Структурная схема И-регулятора**

Итогом проведенных экспериментов стали данные, зафиксированные в таблицах, а также построенные по ним графики и на основании целевых функций, был найден наиболее оптимальный коэффициент И-регулятора:  $K_I - 0.00035$ .

ПИ-регулятор.

Структурная схема приведена на рисунке 5.



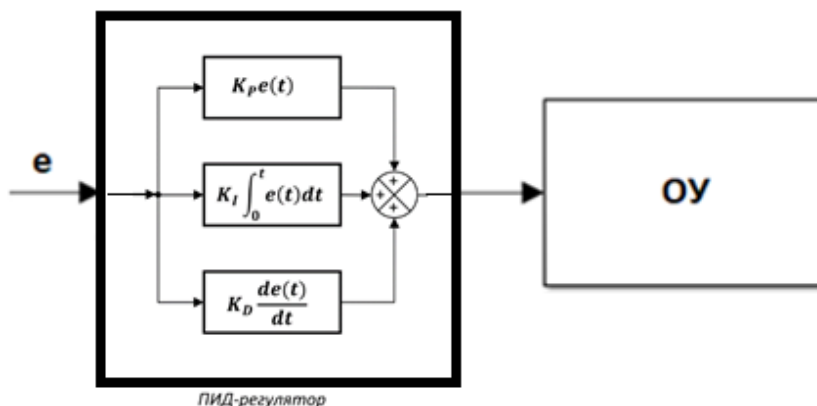
**Рисунок 5. Структурная схема ПИ-регулятора**

Итогом проведенных экспериментов стали данные, зафиксированные в таблицах, а также построенные по ним графики и на основании целевых

функций, были найдены наиболее оптимальные коэффициенты ПИ-регулятора:  $K_p = 0.00035$ ,  $K_i = 3.5$ .

ПИД-регулятор.

Структурная схема приведена на рисунке 6.

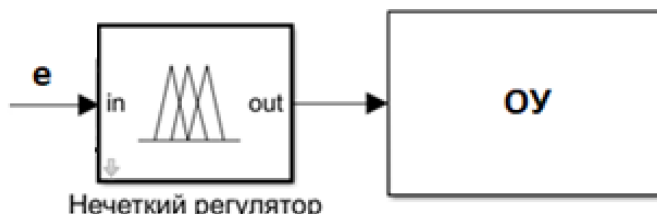


**Рисунок 6. Структурная схема ПИД-регулятора**

Итогом проведенных экспериментов стали данные, зафиксированные в таблицах, а также построенные по ним графики и на основании целевых функций, были найдены наиболее оптимальные коэффициенты ПИД-регулятора:  $K_p = 0.00035$ ,  $K_i = 3.5$ ,  $K_d = 0.001$

Нечеткий П-регулятор.

Структурная схема приведена на рисунке 7.



**Рисунок 7. Структурная схема Нечеткого П-регулятора**

Итогом проведенных экспериментов стали данные, зафиксированные в таблицах, а также построенные по ним графики и на основании целевых функций, были найдены наиболее оптимальные коэффициенты нечеткого П-регулятора:  $f$  (ширина рабочего диапазона по ошибке)=8;  $i=2$ ;  $U_v/U_m$  (коэффициенты увеличения/уменьшения воздействия)=0.001;  $U_v$  В (коэффициент быстрого увеличения воздействия)=0.0105;  $U_m$  В (коэффициент быстрого уменьшения воздействия)=0.005;

Нечеткий ПД-регулятор.

Структурная схема приведена на рисунке 8.

Фаззификатор данного регулятора представлен тремя термами множеств от ошибки (рисунок 9).

И тремя термами множеств от изменения ошибки (рисунок 10).



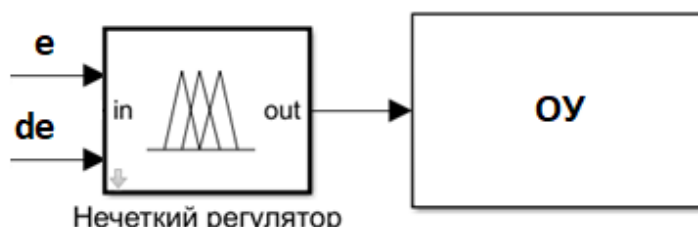


Рисунок 8. Структурная схема Нечеткого ПД-регулятора

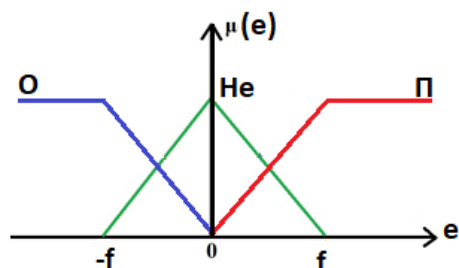


Рисунок 9. Термы ошибки Нечеткого ПД-регулятора, где «О» – отрицательная ошибка; «He» – нулевая ошибка; «П» – положительная ошибка

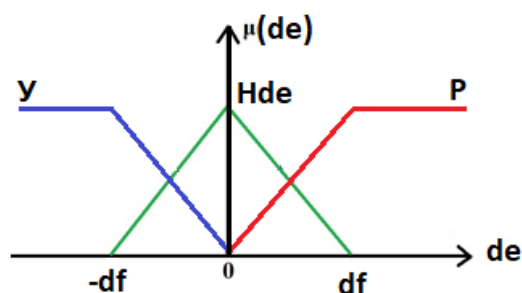


Рисунок 10. Термы производной по ошибке Нечеткого ПД-регулятора, где «У» – ошибка убывает; «Hde» – изменение ошибки нулевое; «Р» – ошибка растет

Блок правил описывается следующими выражениями:

- 1) Если  $e$  «О» и  $de$  «У», то нагреваем быстро;
- 2) Если  $e$  «О» и  $de$  «Hde», то нагреваем;
- 3) Если  $e$  «О» и  $de$  «Р», то не изменяем;
- 4) Если  $e$  «He» и  $de$  «У», то нагреваем;
- 5) Если  $e$  «He» и  $de$  «Hde», то не изменяем;
- 6) Если  $e$  «He» и  $de$  «Р», то охлаждаем;
- 7) Если  $e$  «П» и  $de$  «У», то не изменяем;
- 8) Если  $e$  «П» и  $de$  «Hde», то охлаждаем;
- 9) Если  $e$  «П» и  $de$  «Р», то охлаждаем быстро;

Для устранения нечеткости окончательного результата (дефаззификации) существует достаточно большое количество методов перехода к точным значениям. Наиболее часто используется метод средневзвешенного [5]:

$$out = \frac{\sum \mu_i * out_i}{\sum \mu_i}$$

где  $out$  – результат дефаззификации,  $\mu_i$  – функция принадлежности нечеткого множества  $out_i$ .

Итогом проведенных экспериментов стали данные, зафиксированные в таблицах, а также построенные по ним графики и на основании целевых функций, были найдены наиболее оптимальные коэффициенты нечеткого ПД-регулятора:  $f$  (ширина рабочего диапазона по ошибке)=7;  $df$  (ширина рабочего диапазона по изменению ошибки)=50 Nag/Onl (коэффициенты нагрева/охлаждения)=0.2; Nag B/Ohl B (коэффициенты быстрого нагрева/охлаждения)=5;

### Графики переходного процесса, полученные по итогам эксперимента

График переходного процесса от 0 до 25 градусов представлен на рисунке 11:

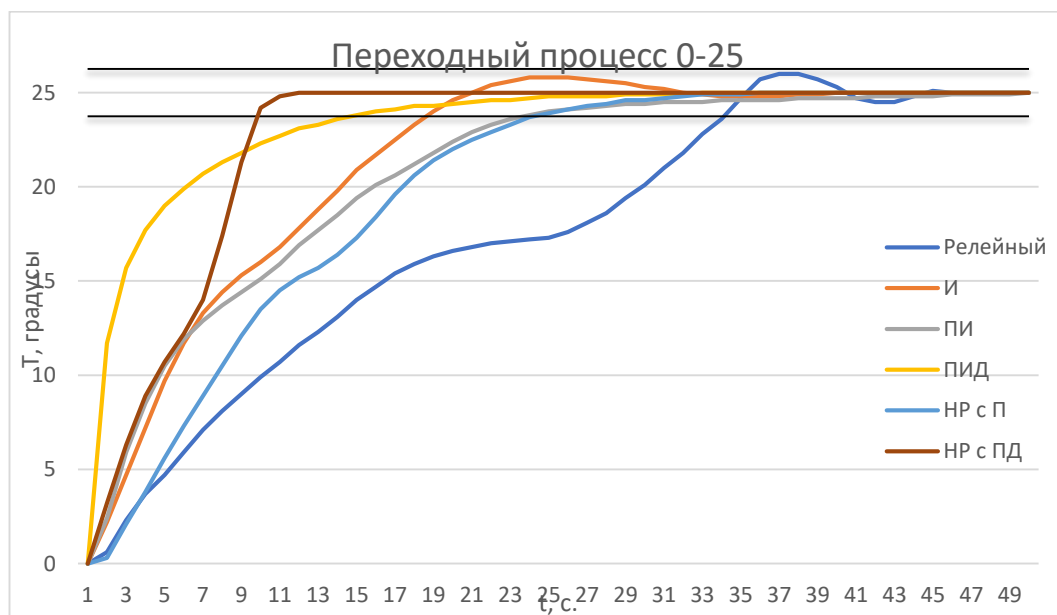


Рисунок 11. Графики переходного процесса, полученные по итогам эксперимента

Далее, приведена оценка результатов эксперимента по целевой функции

**Целевая функция** – обобщенный показатель системы, который характеризует степень достижения системой ее цели [5]. Составление целевой функции одна из важнейших задач при проектировании системы. Однако не существует общей теории построения целевых функций.

В нашем случае с различными видами регуляторов, будем использовать следующий вид целевой функции:

$$F = A * t_{\text{пп}} + B * \sigma,$$

где  $t_{\text{пп}}$  – время переходного процесса;  $\sigma$  – перерегулирование; А и В весовые коэффициенты и их сумма равна единице (для нас оба слагаемых равнозначны, поэтому коэффициенты выбраны одинаковыми, по 0.5);

Сравнение результатов эксперимента приведено на рисунке 12:

Изменение температуры от 0 до 25						
	Тпп, сек	Перерег., градусы	Тпп Относит.	Перерег. Относит.		
Реле рег.	34	1	1,000	1,000	F Реле рег.	1,000
И рег.	18	0,8	0,529	0,800	F И рег.	0,665
ПИ рег.	23	0	0,676	0,000	F ПИ рег.	0,338
ПИД рег.	14	0	0,412	0,000	F ПИД рег.	0,206
НР с П.	23	0	0,676	0,000	F НР с П.	0,338
НР с ПД рег.	9	0	0,265	0,000	F НР с ПД рег.	0,132
max	34	1				

**Рисунок 12. Таблица итогового сравнения целевой функции регуляторов (0-25)**

**Выводы.** В ходе выполнения работы было проведено исследование различных типов регуляторов элемента Пельтье в составе блока БУЦ усилителя спонтанной эмиссии (УСЭМ).

Так, для достижения цели работы было разработано 6 различных регуляторов элемента Пельтье, которые помимо описанных выше тестов проходили тестирование в термокамере на температурах в -10, -40 и -50 градусов по Цельсию, в том числе была проведена оценка результатов экспериментов по целевой функции.

В итоге был разработан нечёткий регулятор с ПД составляющими, который по целевой функции превзошел все рассмотренные выше регуляторы, в том числе часто используемый «классический» ПИД-регулятор на 40%.

### Библиографический список

1. Стариков С. С., Кель О. Л., Вольхин И. Л. Измерение шумов волоконно-оптических источников излучения // Вестник Пермского университета. Физика. 2019. № 1. С. 66–73. doi: 10.17072/1994-3598-2019-1-66- 73
2. EDFA. [Электронный ресурс]. Режим доступа: <https://ru.wikipedia.org/wiki/EDFA>
3. Элемент Пельтье. [Электронный ресурс]. Режим доступа: [https://ru.wikipedia.org/wiki/%D0%AD%D0%BB%D0%B5%D0%BC%D0%B5%D0%BD%D1%82\\_%D0%9F%D0%B5%D0%BB%D1%8C%D1%82%D1%8C%D0%B5](https://ru.wikipedia.org/wiki/%D0%AD%D0%BB%D0%B5%D0%BC%D0%B5%D0%BD%D1%82_%D0%9F%D0%B5%D0%BB%D1%8C%D1%82%D1%8C%D0%B5)
4. Гостев В.И. Нечеткие регуляторы в системах автоматического управления. – К.: Радиомотор, 2008. – 972с.
5. Хижняков Ю.Н. Нечеткое, нейронное и гибридное управление: учеб. пособие / Ю.Н. Хижняков. – Пермь: Изд-во Перм. нац. исслед. политехн. ун-та, 2013. – 303 с.

# SELECTION AND OPTIMIZATION OF PARAMETERS OF THE PELTIER ELEMENT CONTROLLER OF THE DIGITAL CONTROL UNIT BLOCK DEVELOPED ON THE DOMESTIC ELEMENT BASE

*Sokolov Dmitriy A., Natsanov Maksim A.  
Zagvozhkin Vladimir D., Hizhnyakov Yuri N.*

Perm National Research Polytechnic University

Str. Komsomolsky prospect, 29, Perm, Russia, 614000, forstudy01@mail.ru

Modern problems of automation of complex nonlinear (non-deterministic) objects are associated with the lack of their mathematical description. As a rule, nondeterministic objects of different classes require different kinds of management. In our case, the temperature control of the laser diode is carried out by the Peltier element. Due to the development of the digital control unit block on the domestic element base, the question arises about the selection and optimization of the parameters of the Peltier element regulator.

**Keywords:** Peltier element, regulator, optimization.

УДК 332.3

## НЕЙРОСЕТЕВАЯ СИСТЕМА ОПРЕДЕЛЕНИЯ ОПТИМАЛЬНОЙ АГРОКУЛЬТУРЫ ДЛЯ ВОЗДЕЛЫВАНИЯ НА УЧАСТКЕ

*Агаев Артём Русланович*

Пермский государственный национальный исследовательский университет,  
ПМИ. 614990, Россия, г. Пермь, ул. Букирева, 15, artem5agaev@gmail.com

В статье представлено описание разработки нейросетевой системы для определения наиболее подходящей агрокультуры для земельного участка. Система позволяет с большой точностью определить агрокультуру, возделывание которой на участке будет эффективнейшим, основываясь на данных об этом участке. С помощью разработанной интеллектуальной системы проведено исследование предметной области, выявлены закономерности, имеющие практической значение.

**Ключевые слова:** искусственный интеллект, нейросетевые технологии, сельское хозяйство.

**Введение.** Сельское хозяйство обеспечивает население большей частью продуктов питания и некоторых материалов, например, хлопком. Однако суммарная площадь участков, пригодных для возделывания, ограничена, что ставит задачу повышения эффективности сельскохозяйственных работ. При освоении

земельного участка прежде всего необходимо правильно выбрать культуру, которой в наибольшей степени подходят условия на этом участке.

При анализе литературных источников выяснилось, что, несмотря на необходимость использования нейросетевых технологий в сельском хозяйстве, отмеченную во многих статьях, например [1-2], практически нет работ, связанных с распределением агрокультур по участкам.

Основная цель настоящей работы заключается в сборе множества публичных данных о земельных участках и культурах, дающих наибольший относительный урожай на этих участках, а также создание и обучение нейросетевой модели на этих данных. Конечным результатом должна быть нейросетевая система, наиболее точно определяющая подходящую для данного участка культуру.

Набор данных для обучения нейронной сети был взят с платформы Kaggle [3]. Перед переходом к проектированию и обучению модели, была выполнена очистка исходного множества от выбросов и пропущенных значений, произведено преобразование категориальных признаков в числовой формат и их разделение.

Для проектирования системы были выбраны следующие параметры:

- X1 – уровень содержания азота в почве, кг/га
- X2 – уровень содержания фосфора в почве, кг/га
- X3 – уровень содержания калия в почве, кг/га
- X4 – температура в градусах Цельсия
- X5 – относительная влажность
- X6 – мера кислотности почвы
- X7 – количество осадков в мм

Выходные параметры D1 – D2 – мера эффективности выращивания каждой культуры на данном участке. Список культур (номер в списке соответствует номеру выходного параметра):

1. рис
2. кукуруза
3. нут
4. фасоль
5. голубиный горох
6. мотыльковая фасоль
7. маш
8. урд
9. чечевица
10. гранат
11. бананы
12. манго
13. виноград
14. арбуз
15. дыня
16. яблоки
17. апельсины
18. папайя

19. кокос
20. хлопок
21. джут
22. кофе

Проектирование, обучение, тестирование нейронной сети выполнялись с помощью программы «Нейросимулятор 5» [4] по методике [5]. В результате оптимизации спроектированная нейронная сеть представляла собой персептрон, который имеет семь входных нейронов, двадцать два выходных и один скрытый слой с шестьюдесятью нейронами.

Для оценки корректной работы спроектированной нейронной сети использовалось тестирующее множество, состоящее из 215 примеров. Средняя относительная ошибка тестирования не превышает 3%, что можно считать хорошим результатом.

Для наглядного представления разницы между ожидаемым и полученным результатами была построена гистограмма (рисунок 1) суммарных ошибок всех выходных параметров для каждого примера. Для наглядности на гистограмме отображена только часть тестов. Стоит отметить, что из всех 215 примеров, только 14 с ошибкой выше 30%.



**Рисунок 1. Сравнение ожидаемых и полученных значений**

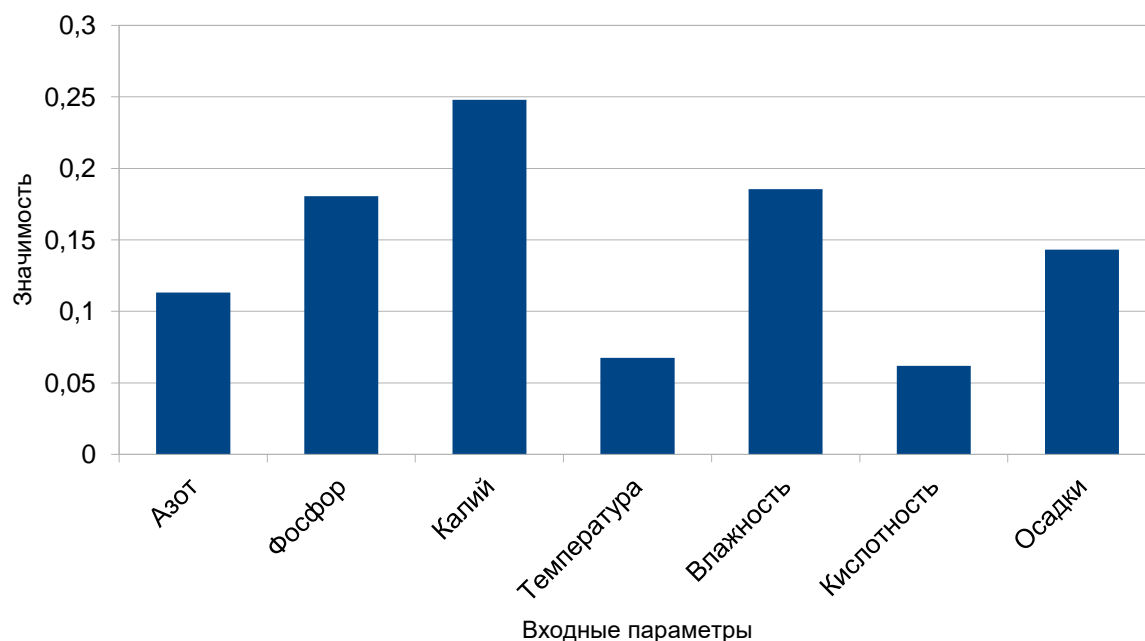
Из результатов, изображенных на рисунке 1, можно сделать вывод об адекватной работе спроектированной нейронной сети.

Оценка значимости входных параметров была выполнена с помощью программы «Нейросимулятор 5», результат отображен на рисунке 2.

Как видно из рисунка 2, наиболее значимыми являются уровень содержания калия, фосфора, и количество осадков.

Далее было проведено исследование полученных зависимостей между входными параметрами и рекомендуемой культурой. Исследование производилось с помощью метода «замораживания» [5], суть которого заключается в

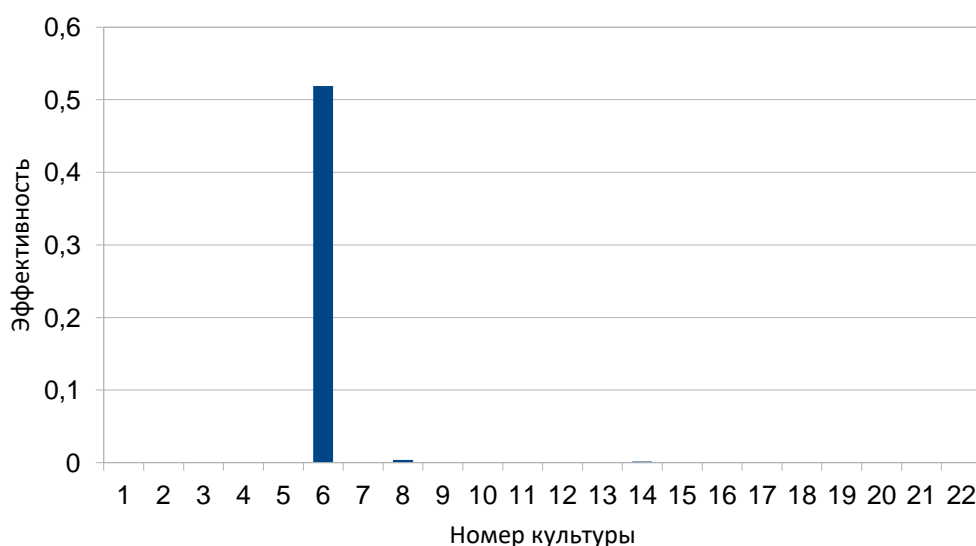
варьировании значения одного параметра и фиксировании значений всех других параметров.



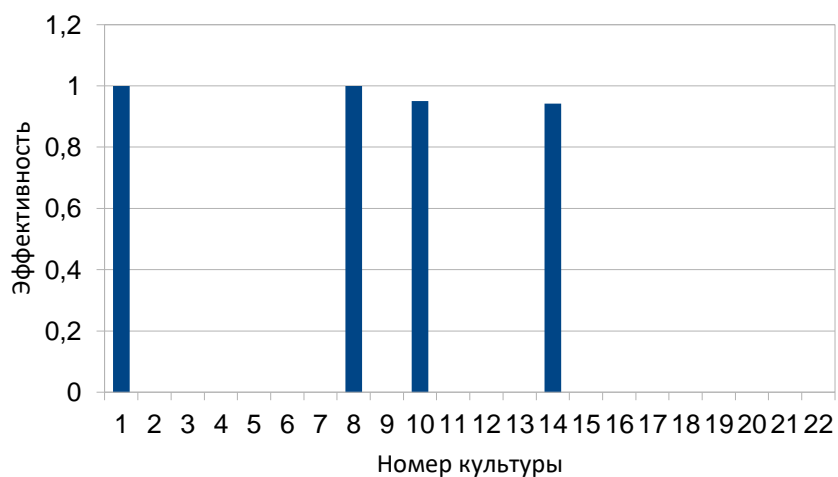
**Рисунок 2. Значимость входных параметров нейросетевой модели**

На рисунках 3-4 видно, как с увеличением количества осадков повышается эффективность выращивания риса и некоторых других неприхотливых растений.

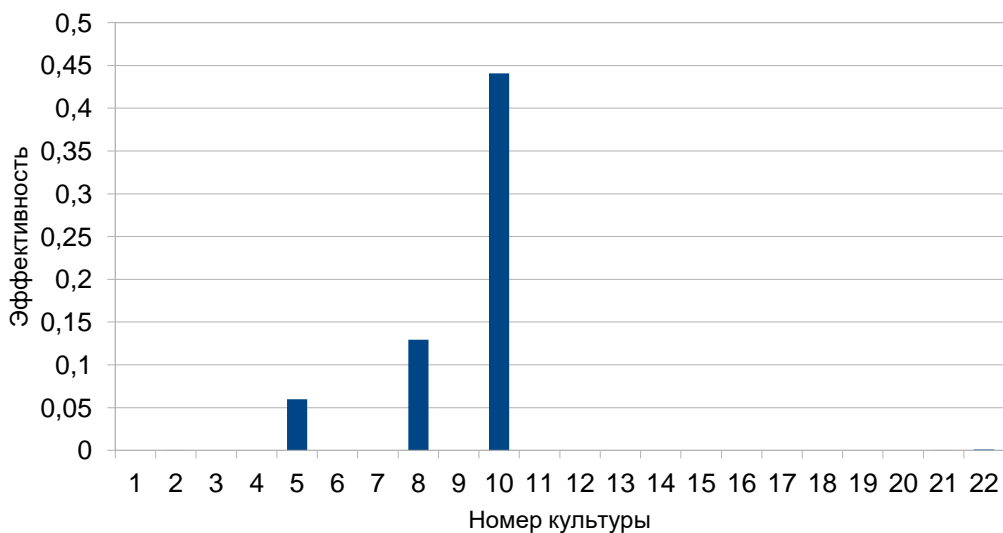
На рисунках 5-6 изображены диаграммы, демонстрирующие изменение эффективности при повышении температуры.



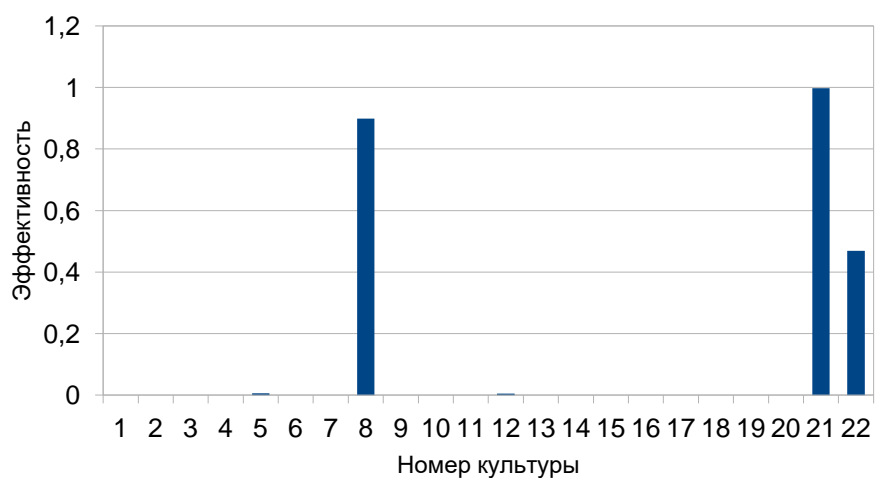
**Рисунок 3. Эффективность культур в бедной почве при низком количестве осадков в тёплом климате**



**Рисунок 4. Эффективность культур в бедной почве при большом количестве осадков в тёплом климате**



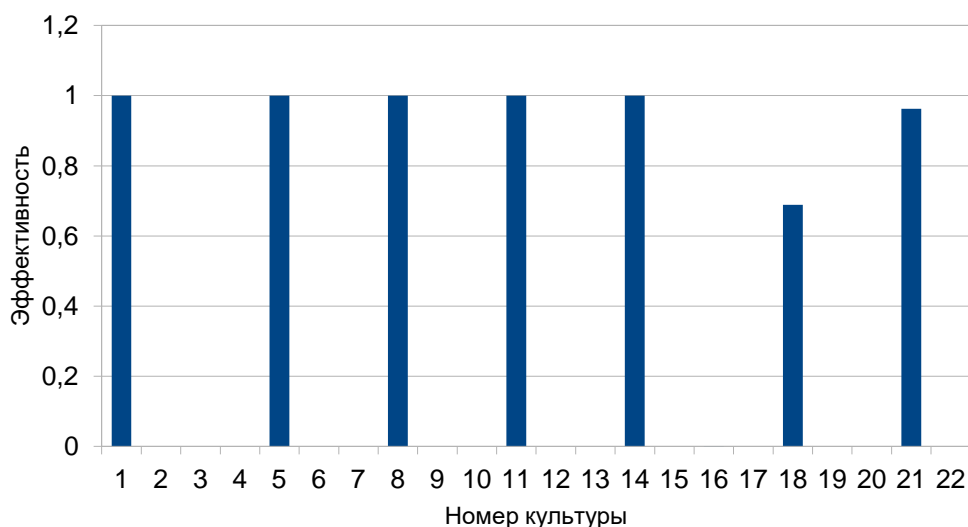
**Рисунок 5. Эффективность культур в бедной почве при температуре в 20 градусов Цельсия**



**Рисунок 6. Эффективность культур в бедной почве при температуре в 30 градусов Цельсия**



Попробуем теперь увеличить насыщенность почвы микроэлементами (рисунок 7). Видна тенденция к увеличению эффективности многих культур.



**Рисунок 7. Эффективность культур в плодородной почве**

Полученные результаты исследований не противоречат реальности, что подтверждает, что спроектированную нейронную сеть можно считать пригодной для выбора наиболее эффективной культуры для возделывания. Помимо этого, полученный инструмент можно использовать для выявления правильной стратегии изменения условий, если требуется вырастить определённую культуру – руководствуясь изменениями показаний нейросети подобрать правильное количество удобрений, усилить орошение на определённый уровень.

**Заключение.** Построена система нейросетевого прогнозирования эффективнейшей агрокультуры в условиях данного земельного участка. Спроектированная нейросетевая модель учитывает 7 параметров: уровень содержания азота в почве, уровень содержания фосфора в почве, уровень содержания калия в почве, температура, относительная влажность, кислотность почвы, количество осадков. Методом сценарного прогнозирования построены графики зависимости выбора самой эффективной культуры от некоторых параметров.

### **Библиографический список**

1. Погоньшев В.А., Ториков В.Е. Нейронные сети в цифровом сельском хозяйстве. Журнал «Вестник Брянской государственной сельскохозяйственной академии». 2021г., номер 5, стр. 68-71.

2. Скворцов Е.А., Набоков В.И., Некрасов К.В., Скворцова Е.Г., Кротов М.И. Применение технологий искусственного интеллекта в сельском хозяйстве. Журнал «Аграрный вестник Урала». 2019, № 8, стр. 91-98.

3. Kaggle – Crop Recommendation Dataset. [Электронный ресурс] [Режим доступа: <https://www.kaggle.com/datasets/atharvaingle/crop-recommendation-dataset>]

4. Черепанов Ф.М., Ясницкий Л.Н. Нейросимулятор 4.0. Свидетельство о регистрации программы для ЭВМ RUS 2014612546. Заявка № 2014610341 от 15.01.2014.

5. Ясницкий Л.Н. Введение в искусственный интеллект. М.: Издательский центр «Академия», 2005. – 176 с.

## NEURAL NETWORK SYSTEM FOR PREDICTING THE MOST EFFECTIVE CROP

*Agaeв Artem R.*

Perm State University  
artem5agaev@gmail.com

The article presents a description of the development of a neural network system to determine the most suitable crop for a given area. The system allows you to determine with great accuracy the agricultural crop, the cultivation of which on the site will be the most effective, based on the data on this site.

With the help of the developed intellectual system, a study of the subject area was carried out, patterns of practical importance were identified.

**Keywords:** artificial intelligence, neural network, agronomy.

УДК 004.896

## РЕЗУЛЬТАТЫ РАЗРАБОТКИ ПРОТОТИПА ПРОГНОСТИЧЕСКОЙ СИСТЕМЫ ПО МОДЕЛИРОВАНИЮ РАСПРЕДЕЛЕНИЯ ТЕМПЕРАТУРЫ ПО ГЛУБИНЕ СТВОЛА СКВАЖИНЫ В ПРОЦЕССЕ ПАРОЦИКЛИЧЕСКИХ ОБРАБОТОК

*Степанов Владимир Анатольевич<sup>1</sup>, Култышева Светлана Николаевна<sup>2</sup>*

<sup>1</sup> Пермский государственный национальный исследовательский университет,  
614990, Россия, г. Пермь, ул. Букирева, 15, svamail@list.ru

<sup>2</sup> Пермский национальный исследовательский политехнический университет,  
614990, Россия, г. Пермь, ул. Комсомольский проспект, 29,  
skultysheva@mail.ru

В статье представлено описание результатов разработки инженерного программного модуля для прогнозирования параметров теплоносителя по стволу скважины в процессе проведения пароциклических обработок на примере пермо-карбонной залежи Усинского месторождения. Спроектированы варианты архитектур нейронных сетей и выполнено их предварительное обучение, валидация и тестирование.

**Ключевые слова:** пароциклическая обработка, геофизические исследования, нейронная сеть, обучающее и валидирующее множество, тестовая скважина.

**Введение.** С каждым годом все большее значение приобретает добыча высоковязкой нефти. Основными причинами развития добычи тяжелой нефти в мире являлись рост цен на нефть и развитие технологий добычи высоковязкой нефти, обеспечивающих достижение высокого коэффициента извлечения (до 60% от начальных геологических запасов нефти) [1, 2].

В России разрабатываются два крупных месторождения сверхвязкой нефти, – Ярегское и Усинское (пермо-карбоновая залежь). Разработка залежей высоковязких нефтей осуществляется как традиционными способами (истощение, заводнение), так и с помощью термических методов.

Наиболее распространенным и эффективным методом повышения нефтеотдачи при добыче высоковязких нефтей считается закачка пара в пласт [2-5]. Она способствует повышению степени извлечения нефти из продуктивного пласта, которое обеспечивается вследствие облегчения условий фильтрации за счет снижения вязкости пластовой нефти и локального повышения пластового давления в зоне воздействия теплоносителем [3, 6]. Кроме того, теплоноситель обеспечивает объемное расширение породы и содержащегося в нем флюида, снижение поверхностного натяжения и приводит к изменению типа смачиваемости породы.

Закачка пара может осуществляться как в нагнетательные, так и в добывающие скважины. В случае закачки в добывающие скважины воздействие паром осуществляют в циклическом режиме. Технология проведения пароциклической обработки (ПЦО) представляет собой последовательность этапов: после закачки необходимого количества пара в добывающую скважину закрывают на пропитку, после чего скважину запускают в эксплуатацию с пониженной вязкостью [3, 7].

Во многих странах для добычи высоковязкой нефти используют закачку пара в пласт. Процесс тщательно исследован при фильтрации теплоносителя через образцы керна в лабораторных условиях. Также имеется значительный опыт проведения данной технологии и промысловых условиях. Применение моделирования позволяет математически описать исследуемую задачу, что способствует увеличению понимания сущности процесса ПЦО и влечет за собой повышение эффективности проектирования тепловых методов воздействия на пласт.

Разработано множество моделей вытеснения нефти паром. Численные и аналитические модели имеют разное назначение. Аналитическое моделирование больше применяют для выполнения инженерных расчетов, характеризующих конкретный этап теплового воздействия. Численные модели используют как инструмент для решения задач по проектированию и для возможности оптимизации закачки пара в пласт. В отличие от аналитических численные модели требуют значительного объема информации об изучаемом объекте и процессах в нем. Определение метода моделирования основывает-

ся на характере используемых исходных данных, их полноте и качестве. Поэтому в настоящее время актуальной задачей является разработка технологий моделирования, позволяющих обеспечить максимальную точность прогноза при любых соотношениях в системе «уровень знаний-объем данных». В частности, даже в таком случае, когда неизвестны уравнения, описывающие процессы в системе. В связи с обозначенным, все большее применение сегодня находят нейронные сети и прочие методы, работоспособные при недостаточном количестве и качестве исходных данных [8, 9]. Далее кратко описаны применения искусственного интеллекта в области нефтедобычи.

Запатентовано [10] использование нейронных сетей в области количественного прогнозирования нефтегазопродуктивных типов в геологическом разрезе и построения карт распределения на основе комплексной интерпретации сейсмических, геохимических и теромобарических атрибутов. Применение нейронных сетей позволяет изучать межскважинное пространство в условиях высокоуглеродистых отложений битуминозного типа. Результатом прогноза является повышение эффективности геолого-разведочных работ за счет повышения точности при проектировании размещения скважин различного назначения, а также площади проведения сейсморазведочных работ.

Исследователь [3] предлагает применять нейронные сети и математическое моделирование при решении задач оптимизации работы глубинно-насосного оборудования, спускаемого в скважину после проведения ПЦО, в частности штанговых установок. Созданный алгоритм позволяет автоматически регулировать скорость движения штока штанговой установки на основе определения динамики дебита скважины и времени эффективной работы спущенного оборудования.

Применение нейронных сетей для автоматизации процесса обработки и интерпретации данных ГИС [11, 12] вызвано значительным объемом геофизической информации и большими трудозатратами на интерпретацию данных, широким спектром условий проведения исследований и задач, решаемых посредством этих исследований. Отличительной чертой программных продуктов является работа в условиях неопределенностей. Одной из основных задач является учет максимального количества возможных ситуаций. Опытно-промышленное опробование показало, что эффективность использования технологии, в частности количество правильных ответов, близость к реальной ситуации составила 87% [11].

Текущий уровень развития технологий в области нефтедобычи показывает, что одним из основных источников информации по-прежнему остаются геофизические исследования скважин (ГИС). Они позволяют получить количественную информацию о коллекторе и процессах, происходящих в нем, за счет комплексирования петрофизического, алгоритмического, методического и метрологического обеспечения [13]. Конечной целью ГИС является получение наиболее исчерпывающих представлений об объекте исследования.

Изучение температуры пласта пермо-карбоновой залежи Усинского месторождения возможно лишь двумя способами – проведение прямых заме-

ров в скважинах посредством спуска геофизического прибора и контроль температуры по датчикам глубинно-насосного оборудования.

При проведении ПЦО замер температуры выполняют в середине цикла ПЦО и по окончании закачки пара. Таким образом, проведение ГИС при ПЦО – это единственный безальтернативный на сегодняшний день способ получения информации о степени прогрева пласта и эффективности закачки пара.

Традиционно применяемые детерминированные методы математического моделирования, необходимые для оптимального управления процессом ПЦО, не позволяют учитывать всего многообразия гео-физико-химических явлений процесса ПЦО. Они требуют введения определенных упрощающих гипотез, а потому их эффективность принципиально ограничена. Вследствие этого, например, по данным проведения геофизических исследований в процессе ПЦО пермо-карбоневой залежи Усинского месторождения за 2020 г отмечается «недостаточная результативность исследований». Так, из 105 проведенных исследований 38% неинформативны. Причиной низкого охвата исследованиями является, как правило, отсутствие возможности спуска прибора ГИС до интервалов перфорации. Поэтому такие исследования ограничиваются замерами параметров теплоносителя на глубинах до 500 м, что значительно выше интервалов перфорации и не характеризует степень прогрева пласта. Программных же инструментов, которые бы позволяли прогнозировать распределение теплофизических характеристик закачиваемого пара вглубь ствола скважины без применения прибора ГИС, пока не существует.

Перед исследователями поставлена задача оценки возможности создания инженерного программного продукта, который посредством прогнозирования параметров теплоносителя в процессе ПЦО позволит повысить степень изученности теплового воздействия и усилить контроль за разработкой залежей высоковязкой нефти. Последний в свою очередь обеспечит более эффективное вовлечение запасов нефти в разработку. Создание симулятора позволило бы определять характеристики теплоносителя в любой точке ствола скважины в течение цикла ПЦО без необходимости проведения геофизических исследований на скважинах.

Задача характеризуется отсутствием математической модели, описывающей всю сложность теплофизических процессов, происходящих по пути движения пара в скважине, а также множеством неопределенностей и ограниченным набором исходных данных.

Решение задачи найдено посредством использования возможностей нейронных сетей. Современный аппарат нейронных сетей позволяет строить математические модели, учитывающие неограниченное количество факторов различной природы, обучать модели на исторических данных (имеющемся опыте) и, следовательно, выявлять и использовать даже такие скрытые закономерности моделируемых явлений, которые на современном уровне развития науки еще не изучены и не известны [14].

Задача нейросетевого моделирования состоит в том, чтобы, используя исторические данные геофизических исследований по оценке распределения теплофизических характеристик закачиваемого пара на разной глубине сква-

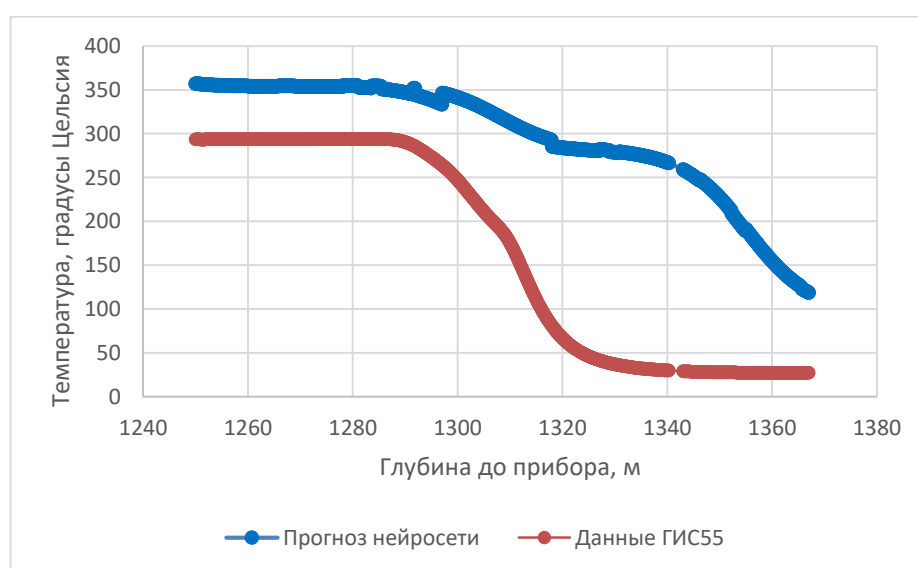
жины, обучить нейронную сеть прогнозировать теплофизические характеристики пара на разных глубинных отметках скважины, используя на входе только параметры скважины и характеристики закачиваемого в скважину пара на уровне ее устья.

Опытным путем установлено, что математическая модель прогнозирования параметров теплоносителя при ПЦО в общем случае должна иметь 17 входных параметров и 7 выходных параметров. Следуя рекомендациям [14], задача построения демонстрационного прототипа прогностической системы разбита на 7 подзадач, состоящих в проектировании семи нейронных сетей, объединенных общим интерфейсом.

Проверка гипотезы о возможности создания нейросетей, прогнозирующих параметры теплоносителя по глубине скважины при ПЦО производилась следующим образом. Из набора исходных данных выбиралась одна скважина, используемая в качестве тестовой. Данные по этой скважине исключались и из оставшихся данных формировалось обучающее и валидирующее множества в соотношении 90 : 10. Эти множества содержали соответственно по 17 входных параметров и один выходной параметр, например, температуру теплоносителя. Нейронные сети обучались различными методами. Среднеквадратичная относительная ошибка на обучающем и валидирующем множествах составила 2,7% и 3,6% соответственно.

После валидации проверка прогностических свойств разрабатываемого прототипа инженерного программного продукта для прогнозирования параметров теплоносителя проводилась на тестовом множестве, представляющем собой данные ГИС на тестовой скважине (которые ни в обучении, ни в валидации не участвовали).

Результаты сопоставления прогнозируемого с помощью прототипа распределения температуры по глубине ствола тестовой скважины с данными геофизических исследований приведены на рисунке 1.



**Рисунок 1. Результаты сопоставления прогнозируемого с помощью прототипа распределения температуры по глубине ствола исследуемой скважины с данными геофизических исследований**

Как видно из рисунка, результаты прогнозирования распределения температуры теплоносителя по глубине ствола тестовой скважины совпадают между собой на качественном уровне, при этом имеют некоторые отличия на количественном уровне.

**Заключение.** Как следует из результатов прогнозного тестирования разработанного прототипа, нейросетевые прогнозы распределения теплоносителя по глубине тестовой скважины на качественном уровне совпадают с результатами геофизических исследований, при этом имеют некоторые отличия на количественном уровне. Поэтому можно утверждать, что принципиальная возможность прогнозирования распределения теплофизических характеристик теплоносителя по глубине скважины при ПЦО существует и может быть реализована на практике.

### Библиографический список

1. Brooks R.T. Experiences in eliminating steam breakthrough & providing zonal isolation in SAGD wells / R. T. Brooks, H. Tavakol // SPE Western Regional Meeting (21-23 March, Bakersfield, California, USA). 2012. Conference paper SPE-153903-MS. P. DOI: 10.2118/153903-MS

2. Оптимизация технологических параметров при пароциклическом воздействии на нефтяные пласты / К. М. Федоров, А. П. Шевелев, А. Я. Гильманов, Т. Н. Ковальчук // Вестник Тюменского государственного университета. Физико-математическое моделирование. Нефть, газ, энергетика. – 2020. – Т. 6. – № 2(22). – С. 145-161. – DOI 10.21684/2411-7978-2020-6-2-145-161. – EDN ZUMYIV.

3. Рожкин, М. Е. Разработка методики контроля работы штанговых установок при добыче высоковязких нефтей / М. Е. Рожкин // Автоматизация, телемеханизация и связь в нефтяной промышленности. – 2011. – № 10. – С. 13-16. – EDN OIVCKJ.

4. Ansari S. The role of emulsions in steam-assisted-gravity-drainage (SAGD) oil-production process: a review / S. Ansari, R. Sabbagh, Y. Yusuf, D. S. Nobes // SPE Journal. 2019. Vol. 24. No. 6. Pp. 1-21. DOI: 10/2118/199347-PA

5. Антониади Д.Г., Гарушев А.Р., Ишханов В.Г. Настольная книга по термическим методам добычи нефти. – Краснодар: Советская Кубань, 2000. – 464 с.

6. Шахмеликьян, М. Г. Анализ применения технологии пароциклического метода интенсификации добычи вязких и высоковязких нефтей / М. Г. Шахмеликьян, Л. К. Нвизуг-Би // Наука. Техника. Технологии (политехнический вестник). – 2018. – № 4. – С. 217-242. – EDN YVLMZN.

7. Рузин, Л.М. Технологические принципы разработки залежей аномально вязких нефтей и битумов [Текст]: монография / Л.М.Рузин, И.Ф. Чупров: под ред. Н.Д. Цхадая. – Ухта: УГТУ, 2007. – 244 с.

8. Концепция эффективного проектирования разработки месторождений углеводородов. Программные решения / А. С. Гавришь, В. П. Косяков,

А. Ю. Боталов [и др.] // Нефтепромысловое дело. – 2015. – № 11. – С. 75-85. – EDN UXPNAD.

9. URL: <http://shahab.pe.wvu.edu/> (дата обращения 24.09.2022)

10. Патент № 2677981 С1 Российская Федерация, МПК G01V 11/00, G01V 9/00. Способ выявления нефтегазопродуктивных типов геологического разреза в межскважинном пространстве в высокоуглеродистых отложениях битуминозного типа : № 2017137485 : заявл. 26.10.2017 : опубл. 22.01.2019 / Е. А. Копилевич, Н. Д. Сурова, М. Б. Скворцов, Г. В. Кузнецов ; заявитель Федеральное государственное бюджетное учреждение "Всероссийский научно-исследовательский геологический нефтяной институт" (ФГБУ "ВНИГНИ"). – EDN MGBJXQ.

11. Вахитова, Г. Р. Разработка компьютерной обработки данных геофизических исследований при освоении и эксплуатации скважин, основанных на базах знаний : специальность 04.00.12 : автореферат диссертации на соискание ученой степени кандидата технических наук / Вахитова Гузель Ринатовна. – Уфа, 1998. – 24 с. – EDN ZKMFVL.

12. Свидетельство о государственной регистрации программы для ЭВМ № 2021614112 Российская Федерация. Программный пакет «ГеоПоиск 10» для комплексной интерпретации данных геофизического исследования скважин («ГеоПоиск 10») : № 2021613153 : заявл. 09.03.2021 : опубл. 19.03.2021 / И. Д. Колкер, А. В. Коломиец, В. Г. Тульчинский, П. Г. Тульчинский. – EDN DISMOD.

13. Мухамедиев, Р. И. Средства автоматизации обработки данных геофизического исследования скважин на месторождениях урана пластово-инфильтрационного типа / Р. И. Мухамедиев, Я. И. Кучин // Cloud of Science. – 2015. – Т. 2. – № 3. – С. 451-464. – EDN UYUDLV.

14. Ясницкий Л.Н. Интеллектуальные системы : учебник. М.: Лаборатория знаний, 2016. 221 с. <https://cloud.mail.ru/public/5H3z/4LxCpishxw>

## **RESULTS OF THE DEVELOPMENT OF A PROTOTYPE OF A PROGNOSTIC SYSTEM FOR MODELING THE TEMPERATURE DISTRIBUTION OVER THE DEPTH OF THE WELL BORE IN THE PROCESS OF STEAM CYCLING TREATMENTS**

*Stepanov Vladimir A.<sup>1</sup>, Kultysheva Svetlana N.<sup>2</sup>*

<sup>1</sup> Perm State University

Str. Bukireva, 15, Perm, Russia, 614990, svamail@list.ru

<sup>2</sup> Perm National Research Polytechnic University

Str. Komsomolsky prospect, 29, Perm, Russia, 614990, skultysheva@mail.ru

The article presents a description of the results of the development of an engineering software module for predicting the parameters of the coolant along the wellbore in the process of steam cycling treatments on the example of the



Permian-Carboniferous deposit of the Usinskoye field. Variants of neural network architectures were designed and their preliminary training, validation and testing were performed.

**Keywords:** steam cycling, geophysical surveys, neural network, training and validation set, test well.

УДК 004.032.26

## НЕЙРОСЕТЕВАЯ СИСТЕМА ПРОГНОЗИРОВАНИЯ СТОИМОСТИ ДОМОВ

*Ковалева Софья Сергеевна*

Национальный исследовательский университет «Высшая школа экономики»,  
614107, Россия, г. Пермь, ул. Бульвар Гагарина, 37А, sskovaleva@edu.hse.ru

В статье представлено описание разработки нейросетевой системы для прогнозирования результатов стоимости домов относительно некоторых их характеристик. С помощью разработанной интеллектуальной системы проведено исследование предметной области, выявлены закономерности, имеющие практического значения.

**Ключевые слова:** искусственный интеллект, нейросетевые технологии, прогнозирование, жилые дома, недвижимость.

**Введение.** Прогнозирование стоимости жилья является достаточно актуальной темой, так как именно прогноз может позволить снизить определенные риски, связанные с покупкой жилья. Прогнозирование стоимости жилья может значительно упростить как покупку нового жилья, так и продажу имеющейся недвижимости для ее владельцев. Эксперименты, связанные с нейросетевым программированием в области оценки стоимости недвижимости, показывают, что прогнозирование часто показывает успешные результаты и дает конкретные конечные выводы по взаимосвязям между жильем и характеризующими его факторами [1-4].

Анализ литературных источников показал, что в основном прогнозированием стоимости жилья на рынке занимаются люди вручную, не используя искусственный интеллект. Спрогнозированные результаты опираются на текущее и предположительное будущее состояние рынка недвижимости в целом. Так, в одной из проанализированных научных работ [5] приводится пример прогнозирования стоимости недвижимости с помощью регрессионных моделей. Расчет корреляции между параметрами, а также процесс построения регрессионной модели вручную занимает большое количество времени и является более трудоемким в сравнении с проведением прогнозирования с помощью искусственного интеллекта.

Основная цель настоящей работы заключается в поиске данных о стоимости жилой недвижимости, а также в создании и обучении нейросетевой модели на этих данных. Конечным результатом будет являться нейросетевая система, способная прогнозировать стоимость жилья.

Для создания нейросетевой системы были выбраны следующие параметры: X1 – количество спален, X2 – количество ванных комнат, X3 – размер жилой площади, X4 – размер общей площади участка, X5 – количество этажей, X6 – год постройки жилья.

Некоторые значения представлены в дробном виде, хотя ожидается, что они таковыми быть не должны. Так, количество ванных комнат в выборке встречается в виде вещественных чисел. Дело в том, что ванная комната предполагает наличие именно ванны. В наборе данных представлены дробные количества, к примеру, 1.5 ванные, так как в доме имеется полноценная ванная комната (к примеру, на втором этаже) и отдельный гостевой туалет (на первом этаже) без душа или ванны. Под значением 0.25 имеется в виду отдельная прачечная. Этажи также указаны в виде дробей. Если в доме 1.5 этажа, то в нем есть подвал или чердак. Целое значение предполагает целое количество этажей, а не этаж и подвал, и чердак.

Обучающее множество было найдено на сайте Kaggle [6]. Перед переходом к проектированию и обучению нейросетевой модели, была выполнена очистка исходного множества от противоречивых примеров, выбросов и дубликатов, которые были удалены при работе с языком «Python». Выбросы были найдены посредством ящичковых диаграмм. Так, в данных был обнаружен столбец, заполненный пустыми значениями и не оказывающий никакого влияния на работу. Таким образом, объем итогового множества включает в себя порядка 406 примеров. Данное множество было разделено на обучающее, доминирующее и тестирующее в соотношении примерно 70%, 20% и 10% соответственно.

Проектирование, обучение, тестирование нейронной сети выполнялись с помощью программы «Нейросимулятор 5» [7] по методике [8]. В результате оптимизации спроектированная нейронная сеть представляла собой персептрон, который имеет шесть входных нейронов, один выходной и два скрытых слоя – один с двумя нейронами, другой – с одним.

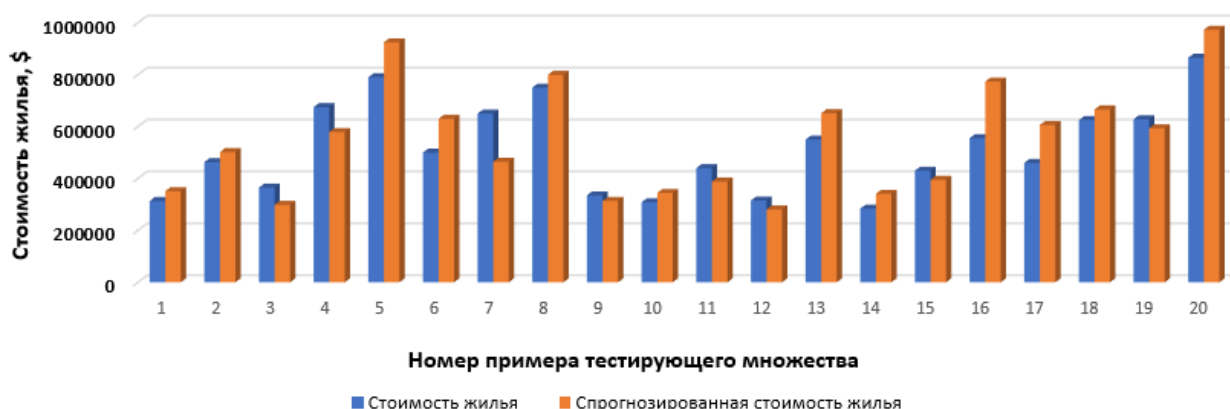
Для оценки корректной работы спроектированной нейронной сети использовалось тестирующее множество, состоящее из 20 примеров. Средняя относительная ошибка тестирования составила 12,3%, что можно считать приемлемым результатом. На рисунке 1 представлена гистограмма, демонстрирующая разницу между фактическим и прогнозируемым нейросетью стоимости жилья.

Из результатов, изображенных на рисунке 1, можно сделать вывод об адекватной работе спроектированной нейронной сети.

Оценка значимости входных параметров была выполнена с помощью программы «Нейросимулятор 5», результат отображен на рисунке 2.

Как видно из рисунка 2, наиболее значимыми являются размер жилой площади. Скорее всего, размер жилой площади покупаемой недвижимости

является наиболее значимым также и для покупателей. Год постройки является следующим по важности параметром, но ни один не опережает по важности размер жилой площади.



**Рисунок 1. Спрогнозированные результаты стоимости жилья**

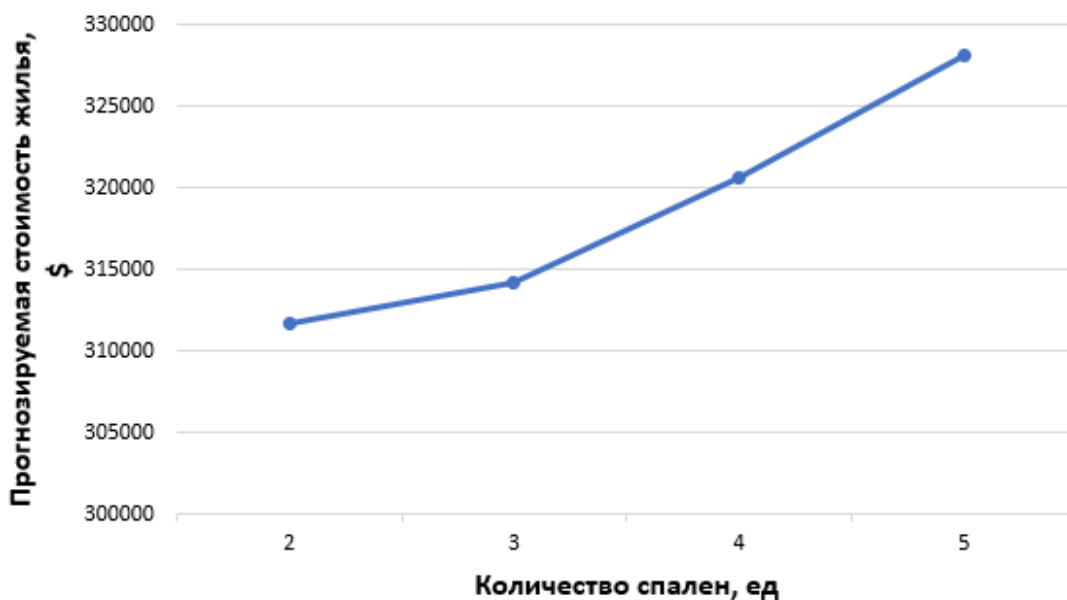


**Рисунок 2. Значимость входных параметров нейросетевой модели**

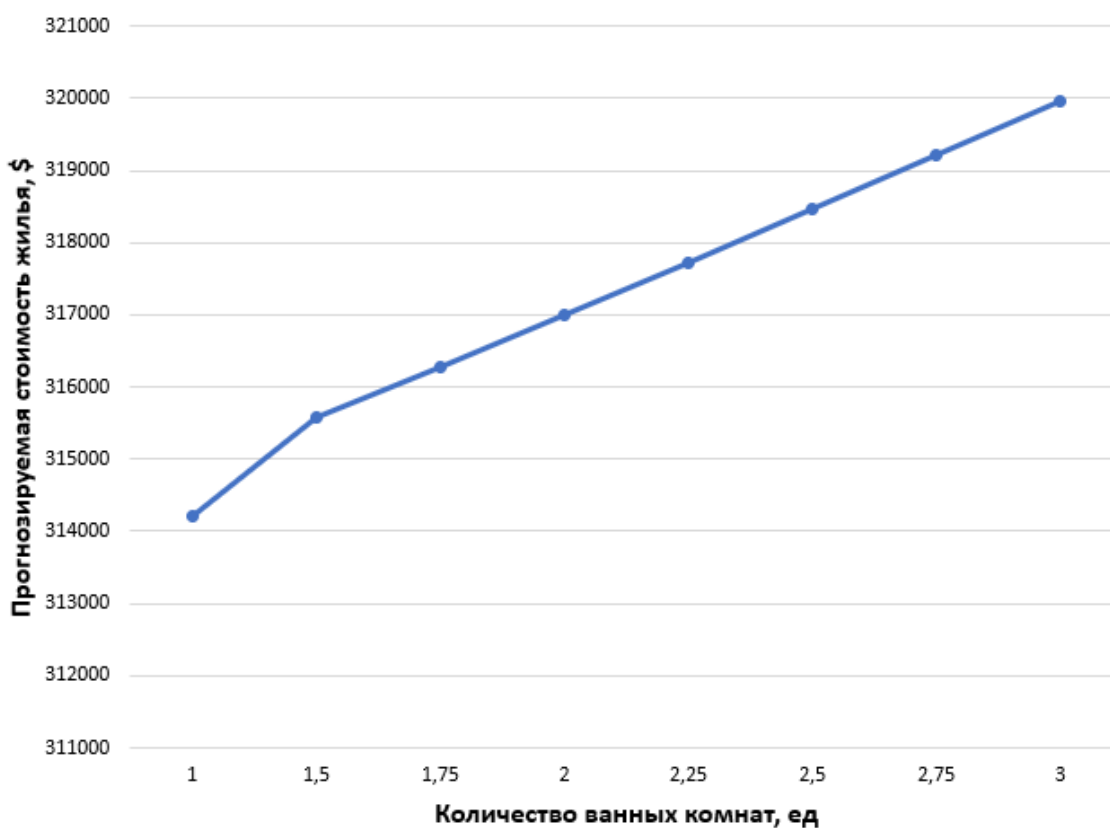
Далее было проведено исследование полученных зависимостей между входными параметрами и стоимости жилья. Исследование производилось с помощью метода «замораживания» [1-2, 6], суть которого заключается в варьировании значения одного параметра и фиксации значений всех других параметров. При работе с каждым из параметров была произведена фиксация всех данных, кроме этого параметра – он оставался изменяемым.

На рисунке 3 показан график зависимости прогнозируемой стоимости жилья от количества спален в нем. Как видно, с повышением количества комнат в доме растет и его стоимость.

На рисунке 4 продемонстрирована зависимость прогнозируемой стоимости жилья от количества ванных комнат в доме. Можно заметить, что чем больше в доме ванных, тем выше его стоимость.

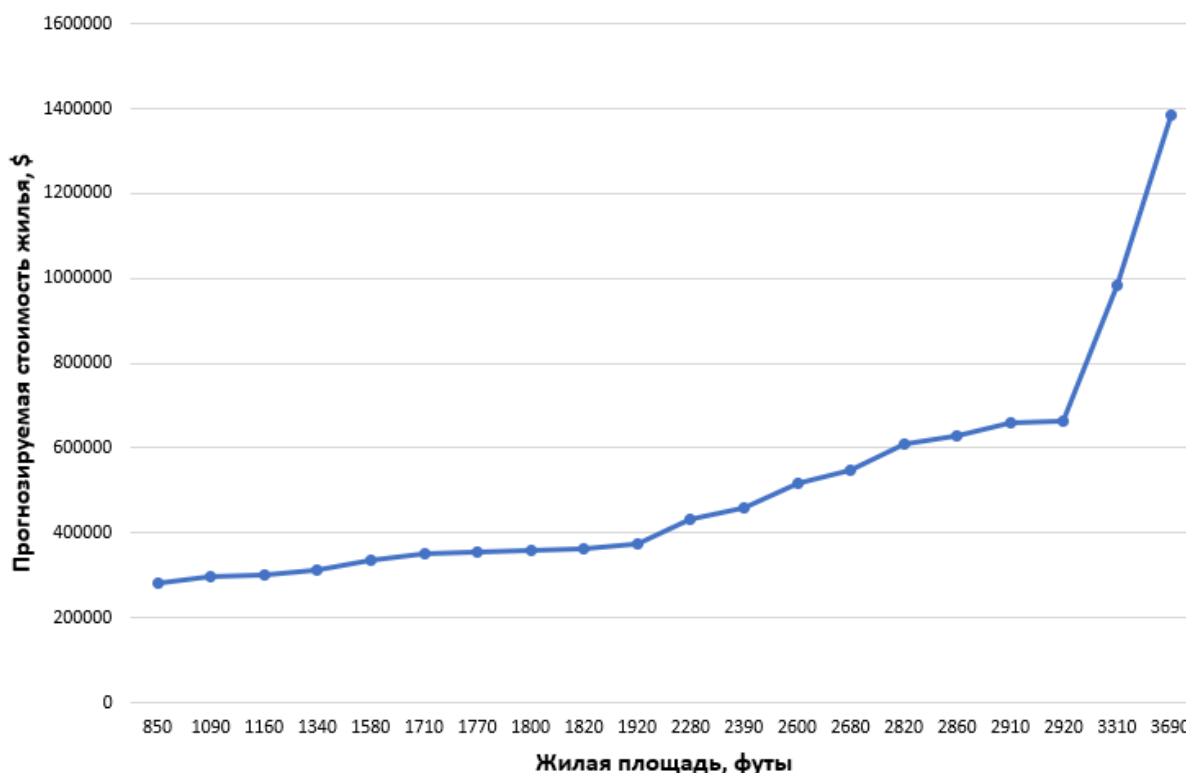


**Рисунок 3. Зависимость прогнозируемой стоимости жилья от количества спален**



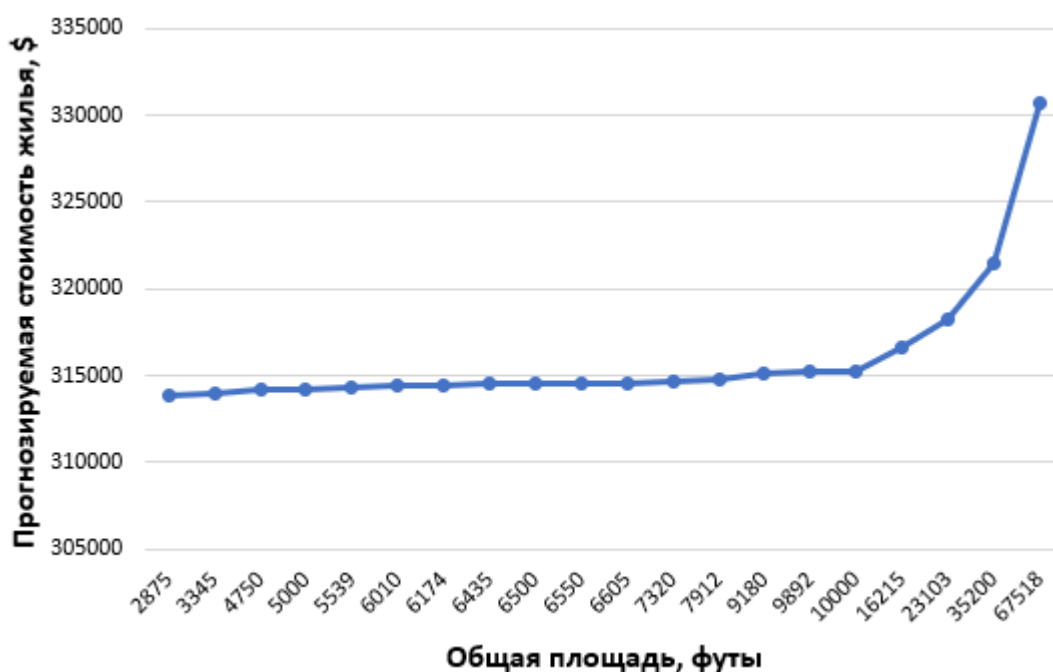
**Рисунок 4. Зависимость прогнозируемой стоимости жилья от количества ванных комнат**

На рисунке 5 изображен график зависимости прогнозируемой стоимости жилья от размера жилой площади. Как видно из графика, в большей степени, стоимость жилья возрастает с увеличением жилой площади. После отметки в приблизительно 3100 футов стоимость на дома значительно возрастает.



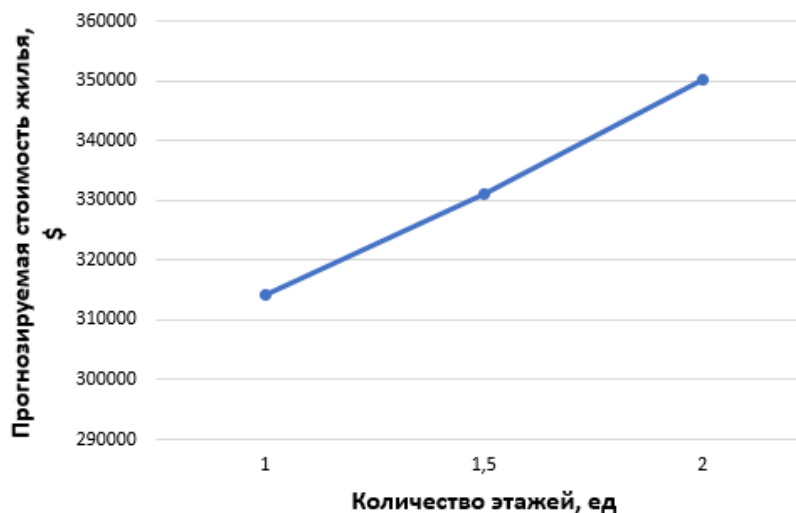
**Рисунок 5. Зависимость прогнозируемой стоимости жилья от размера жилой площади**

На рисунке 6 продемонстрирована зависимость прогнозируемой стоимости жилья от общей площади участка. Как видно из графика, с увеличением площади в промежутке между 5000 и около 10000 футов стоимость жилья почти не возрастает. Однако, когда общая площадь участка начинает превышать 10000 футов, стоимость жилья начинает возрастать значительно.

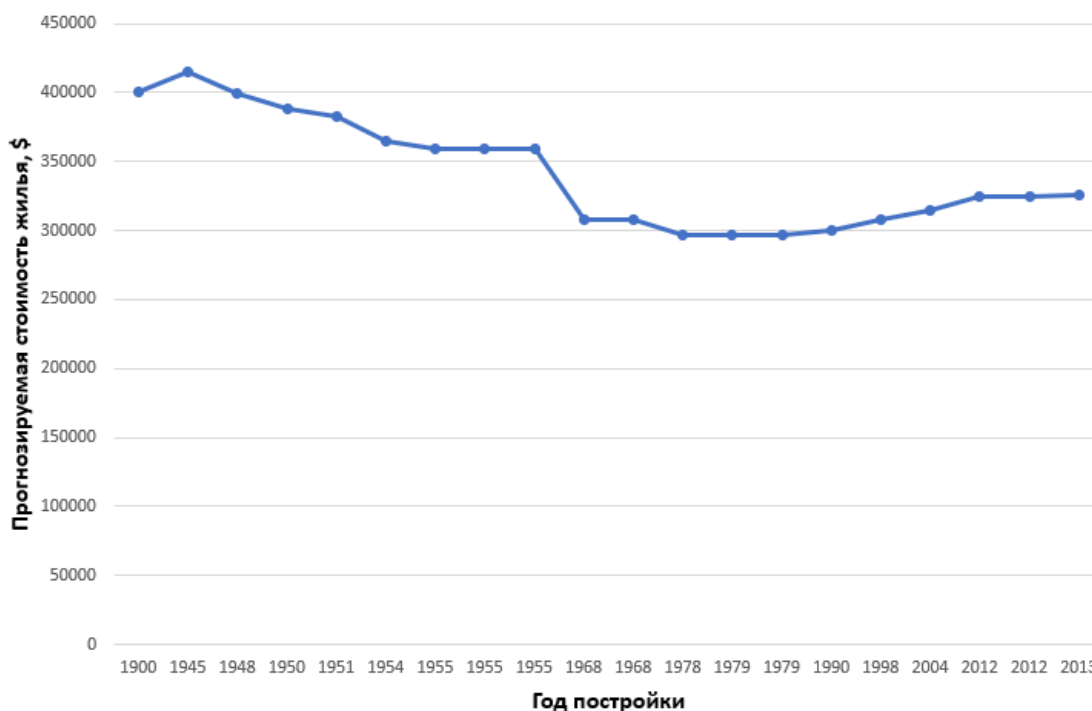


**Рисунок 6. Зависимость прогнозируемой стоимости жилья от общей площади**

На рисунке 7 продемонстрирована зависимость прогнозируемой стоимости жилья от количества этажей. С увеличением количества этажей в доме его стоимость возрастает.



**Рисунок 7. Зависимость прогнозируемой стоимости жилья от количества этажей**



**Рисунок 8. Зависимость прогнозируемой стоимости жилья от года постройки**

На рисунке 8 продемонстрирована зависимость прогнозируемой стоимости жилья от года постройки дома. Как можно заметить, старые дома стоят выше, чем более новые. Набор данных подразумевает, что у всех представленных домов хороший ремонт и дополнительных вложений при въезде такие дома не требуют. Дома, построенные до середины XX века стоят значительно дороже, чем те, что были построены больше. Падение в стоимости можно заметить между домами 1955 и 1968 годами постройки. Возможно такие значения стоимости связаны с материалами, из которых

изготавливались дома ранее и сейчас, а также с возможными интересными интерьерами, сохранившимися в более старых домах.

Полученные результаты исследований не противоречат реальности, что подтверждает, что спроектированную нейронную сеть можно считать пригодной для прогнозирования стоимости жилья.

Помимо работы с «Нейросимулятором 5» была также построена нейросеть на языке «Python» с помощью модуля «Keras». Модель, аналогичная той, что была построена в «Нейросимуляторе 5» обучалась не так хорошо. Чтобы добиться нужного результата была построена совершенно другая модель. В результате оптимизации спроектированная нейронная сеть представляла собой персептрон, который имеет шесть входных нейронов, один выходной и три скрытых слоя – один с шестью нейронами, второй – с четырьмя и третий с двумя.

Построенная на языке «Python» модель оказалась гораздо хуже построенной с помощью «Нейросимулятора 5». Отклонение в прогнозируемых ей результатах превышало иногда пятьсот тысяч долларов, что показало, что работа модели на языке «Python» не является допустимой для адекватного прогнозирования.

**Заключение.** В ходе работы была построена нейросетевая модель прогнозирования стоимости жилья. Спроектированная нейросетевая модель учитывает 6 параметров: количество спален, количество ванных комнат, жилая площадь дома, общая площадь участка, количество этажей, год постройки дома. Методом сценарного прогнозирования построены графики зависимостей прогнозируемого победителя матча от изменения входных параметров. Для проведения данного прогноза был использован метод «заморожки» данных. Применение такого набора параметров в модели позволяет с проводить прогноз стоимости жилья.

Прогнозирование с помощью языка «Python» и модуля «Keras» показала результаты в разы хуже, чем при использовании нейросетевой модели, созданной в «Нейросимуляторе 5».

### **Библиографический список**

1. Ясницкий Л.Н., Ясницкий В.Л. Разработка и применение комплексных нейросетевых моделей массовой оценки и прогнозирования стоимости жилых объектов на примере рынков недвижимости Екатеринбурга и Перми. Имущественные отношения в Российской Федерации. 2017. № 3 (186). С. 68-84.
2. Ясницкий Л.Н., Ясницкий В.Л. Методика создания комплексной экономико-математической модели массовой оценки стоимости объемов недвижимости на примере квартирного рынка города Перми. Вестник Пермского университета. Серия: Экономика. 2016. № 2 (29). С. 54-69.
3. Alexeev A.O., Alexeeva I.E., Yasnitsky L.N., Yasnitsky V.L. Self-adaptive intelligent system for mass evaluation of real estate market in cities. Advances in Intelligent Systems and Computing. 2018. Т. 850. С. 81-87.
4. Yasnitsky L.N., Yasnitsky V.L., Alekseev A.O. The complex neural network model for mass appraisal and scenario forecasting of the urban real estate market value that adapts itself to space and time. Complexity. 2021. Т. 2021. С. 5392170.

5. <https://cyberleninka.ru> – исследование на тему: «Прогнозирование рыночной стоимости коммерческой недвижимости на основе показателей экономического развития территории». [Электронный ресурс]. Режим доступа: <https://cyberleninka.ru/article/n/prognozirovanie-rynochnoy-stoimosti-kommercheskoy-nedvizhimosti-na-osnove-pokazateley-ekonomicheskogo-razvitiya-territorii/viewer>

6. <https://www.kaggle.com> – данные по жилой недвижимости. [Электронный ресурс]. Режим доступа: <https://www.kaggle.com/shree1992/housedata>

7. Черепанов Ф.М., Ясницкий Л.Н. Нейросимулятор 4.0. Свидетельство о регистрации программы для ЭВМ RUS 2014612546. Заявка № 2014610341 от 15.01.2014.

8. Ясницкий Л.Н. Интеллектуальные системы: учебник. М.: Лаборатория знаний, 2016. 221 с.

## NEURAL NETWORK SYSTEM FOR PREDICTING COST OF REAL ESTATE

*Kovaleva Sofia S.*

National Research University “Higher School of Economics”  
St. Bulvar Gagarina, 37A, Perm, Russia, 614107, sskovaleva@edu.hse.ru

The article describes the development of a neural network system for predicting the cost of real estate in United States. The system allows to accurately predict the price of real estate, considering different parameters of the house and territory. With the help of the developed intellectual system, a study of the subject area was carried out, patterns of practical importance were identified.

**Keywords:** artificial intelligence, neural network, prediction, real estate, houses.

УДК 004.032.26

## НЕЙРОСЕТЕВАЯ СИСТЕМА ПРОГНОЗИРОВАНИЯ УРОВНЯ СТРЕССА ЧЕЛОВЕКА ВО СНЕ

*Сырвачева Екатерина Романовна*

Национальный исследовательский университет «Высшая школа экономики»,  
614990, Россия, г. Пермь, ул. Бульвар Гагарина, 37, ersyvacheva@edu.hse.ru

В статье представлено описание разработки нейросетевой системы для прогнозирования уровня стресса человека во время сна. Система позволяет с большой точностью предсказать уровень человеческого стресса во сне и через него путем мониторинга его физиологических данных. С помощью разработанной интеллектуальной системы проведено исследование предметной области, выявлены закономерности, имеющие практическое значение.



**Ключевые слова:** искусственный интеллект, нейросетевые технологии, прогнозирование, стресс, сон.

**Введение.** Стресс определяется как состояние умственного или эмоционального напряжения из-за неизбежных или сложных обстоятельств, также известных как стрессоры. Стресс также можно определить как специфическую нагрузку на организм человека, вызванную различными стрессорами. Стрессовые факторы заставляют организм человека выделять гормоны стресса. Отслеживание изменения стресса в зависимости от качества сна находится в центре внимания этой работы.

Чем лучше качество сна, тем ниже уровень стресса. Сон можно определить как активный период, который помогает в развитии оптимального здоровья и благополучия путем восстановления и укрепления организма человека. Важно понимать, как качество сна влияет на уровень стресса.

На сегодняшний день существует очень мало исследований на тему прогнозирования уровня стресса в зависимости от качества сна, в основном, исследователи прогнозируют стрессоустойчивость человека или стрессовые реакции.

Основная цель настоящей работы заключается в сборе множества данных об физиологических данных человека во время сна, а также создание и обучение нейросетевой модели на основе этих данных. Конечным результатом должна быть нейросетевая система, способная прогнозировать уровень стресса человека во время сна.

Для создания нейросетевой системы были выбраны следующие параметры:

X1 – уровень храпа (в децибелах),

X2 – частота дыхания (количество вдохов в минуту),

X3 – температура тела (°F),

X4 – движение конечностей (по индексу периодических движений конечностей (PLMI), который представляет собой частоту периодических движений конечностей во сне в час от общего времени сна),

X5 – кислород в крови,

X6 – движение глаз,

X7 – часы сна,

X8 – частота сердцебиения (уд/мин).

Выходной параметр – прогноз уровня стресса во время сна и на следующий день по следующей шкале:

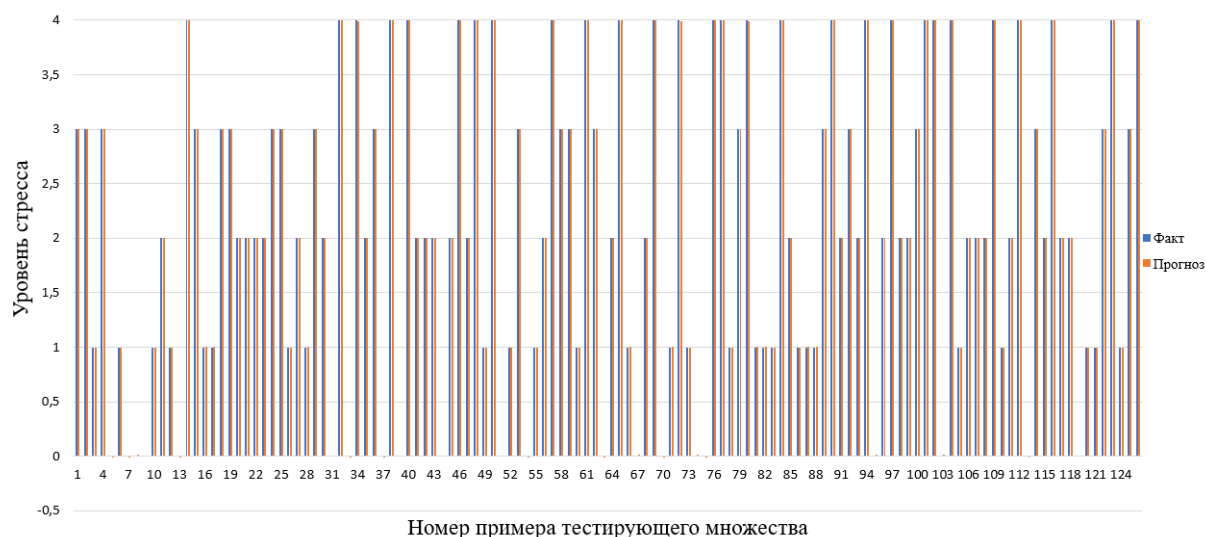
- 0 – низкий/нормальный
- 1 – средне-низкий,
- 2 – средний,
- 3 – средневысокий,
- 4 – высокий.

Обучающее множество [1, 2] было взято с сайта Kaggle [3] (система организации конкурсов по исследованию данных, а также социальная сеть специалистов по обработке данных и машинному обучению). Перед переходом к проектированию и обучению нейросетевой модели, была выполнена очистка

исходного множества от противоречивых примеров, выбросов, дубликатов. Объем итогового множества включает в себя 630 примеров. Данное множество было разделено на обучающее, валидирующее и тестирующее в соотношении 70%, 10% и 20% соответственно.

Проектирование, обучение, тестирование нейронной сети выполнялись с помощью программы «Нейросимулятор 5» [4] по методике [5]. В результате оптимизации спроектированная нейронная сеть представляла собой персептрон, который имеет восемь входных нейронов, один выходной и один скрытый слой с тремя нейронами.

Для оценки корректной работы спроектированной нейронной сети использовалось тестирующее множество, состоящее из 126 примеров. Средняя относительная ошибка тестирования составила 0,046, что можно считать приемлемым результатом. На рисунке 1 представлена гистограмма, демонстрирующая разницу между фактическим значением уровня стресса и значением, прогнозируемым нейросетью.



**Рисунок 3. Результат тестирования нейронной сети**

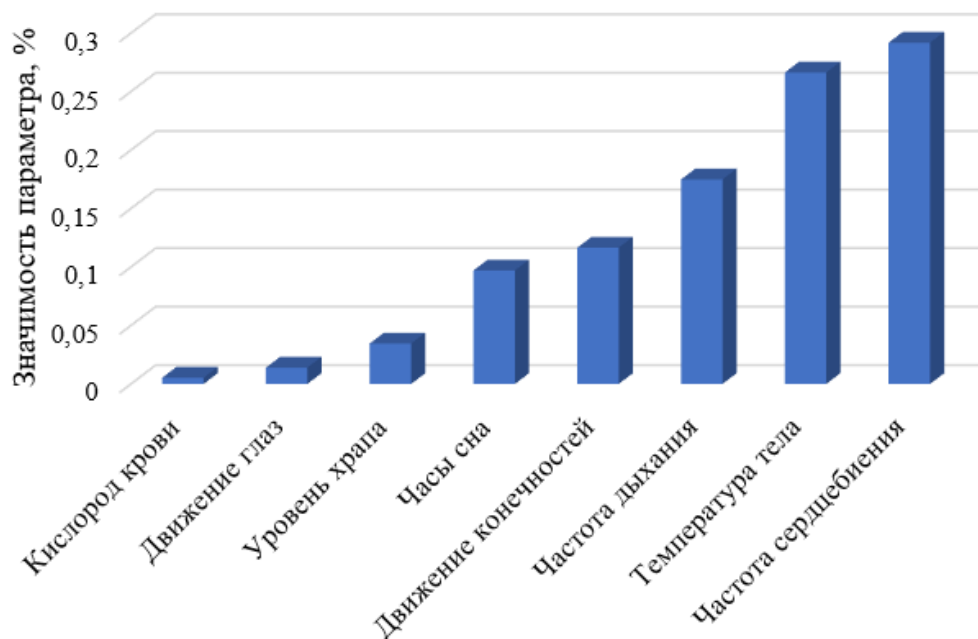
Из результатов, изображенных на рисунке 1, можно сделать вывод о хорошей работе спроектированной нейронной сети.

Оценка значимости входных параметров была выполнена с помощью программы «Нейросимулятор 5», результат отображен на рисунке 2.

Как видно из рисунка 2, все параметры имеют ту или иную значимость. Наиболее значимыми параметрами являются частота сердцебиения, температура тела и частота дыхания. Как и ожидалось, данные параметры являются наиболее влиятельными, так как оказывают серьезное влияние на тело и здоровье человека.

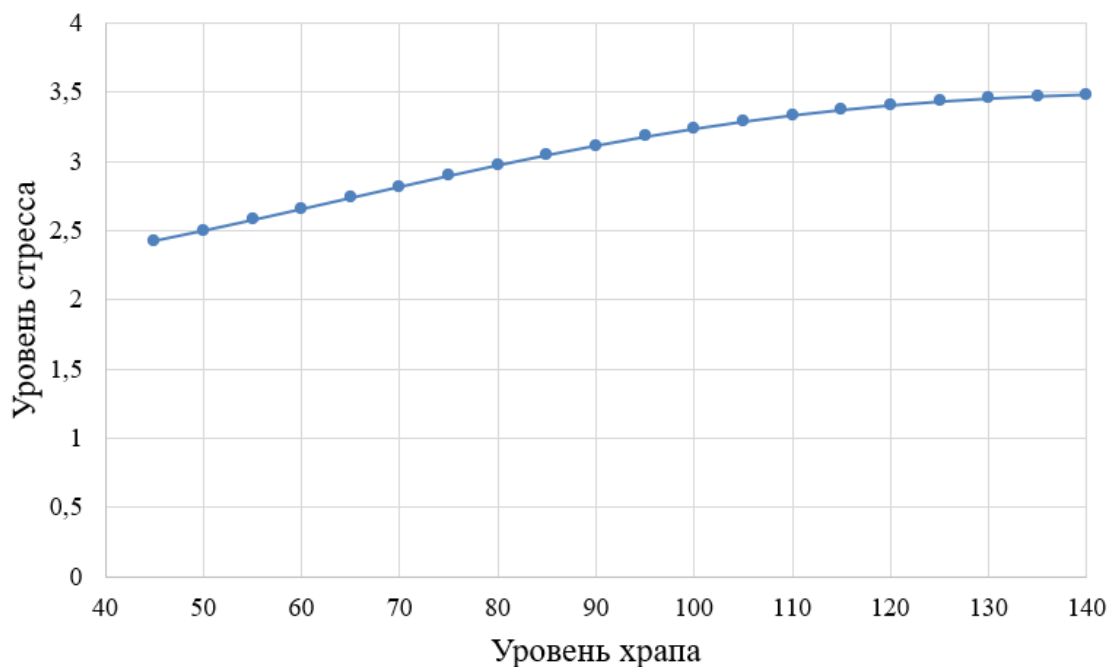
Далее было проведено исследование полученных зависимостей между входными параметрами и выходным параметром, то есть уровнем стресса. Исследование производилось с помощью метода «замораживания» [1-3, 8], суть которого заключается в варьировании значения одного параметра и фиксации значений всех других параметров. Для этой цели был отобран

«нейтральный пример», по которому нейросеть не может с уверенностью определить уровень стресса.



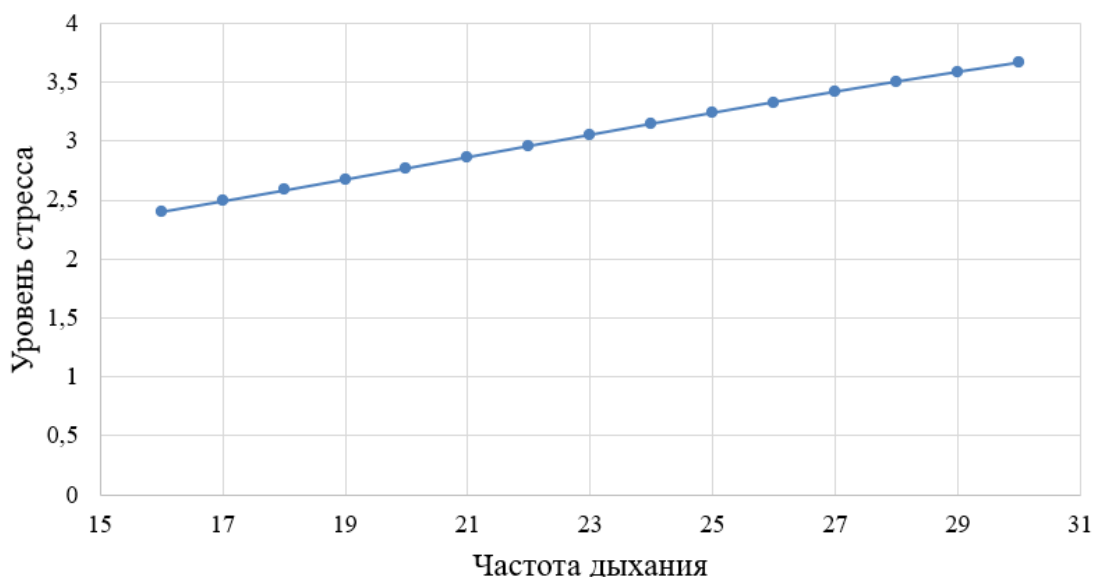
**Рисунок 2. Значимость входных параметров нейросетевой модели**

На рисунке 3 показан график зависимости прогнозируемого уровня стресса от уровня храпа человека во сне. В том случае, когда у человека высокий уровень храпа, нейросеть прогнозирует увеличение уровня стресса.



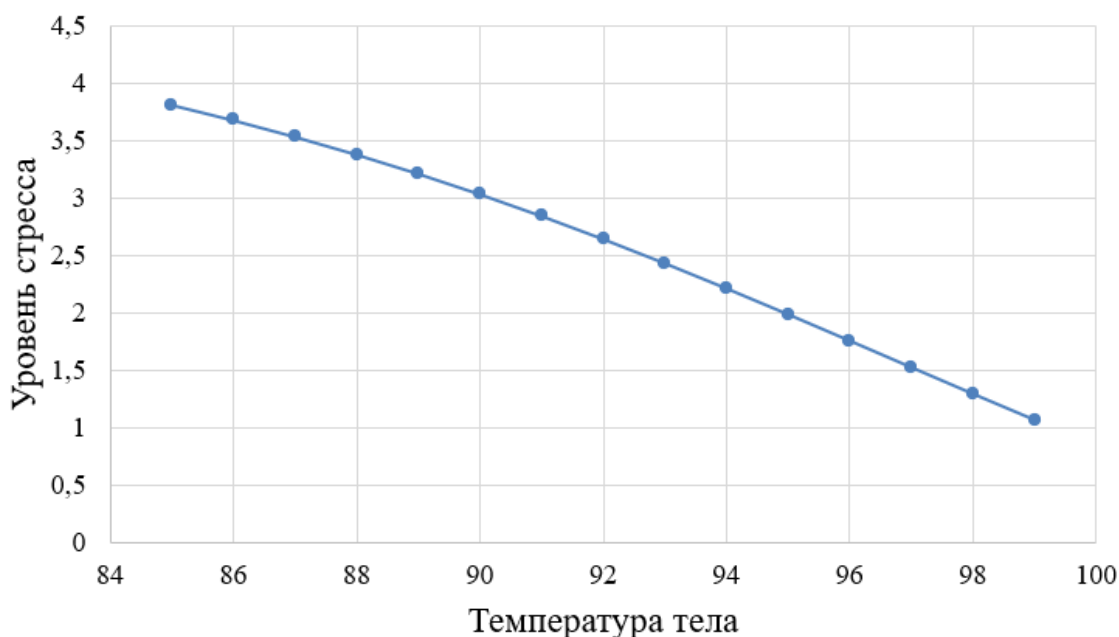
**Рисунок 3. Зависимость прогнозируемого уровня стресса от уровня храпа человека во сне**

На рисунке 4 продемонстрирована зависимость прогнозируемого уровня стресса от частоты дыхания человека во время сна. Можно заметить, что чем выше частота дыхания, тем выше прогнозируемый уровень стресса.



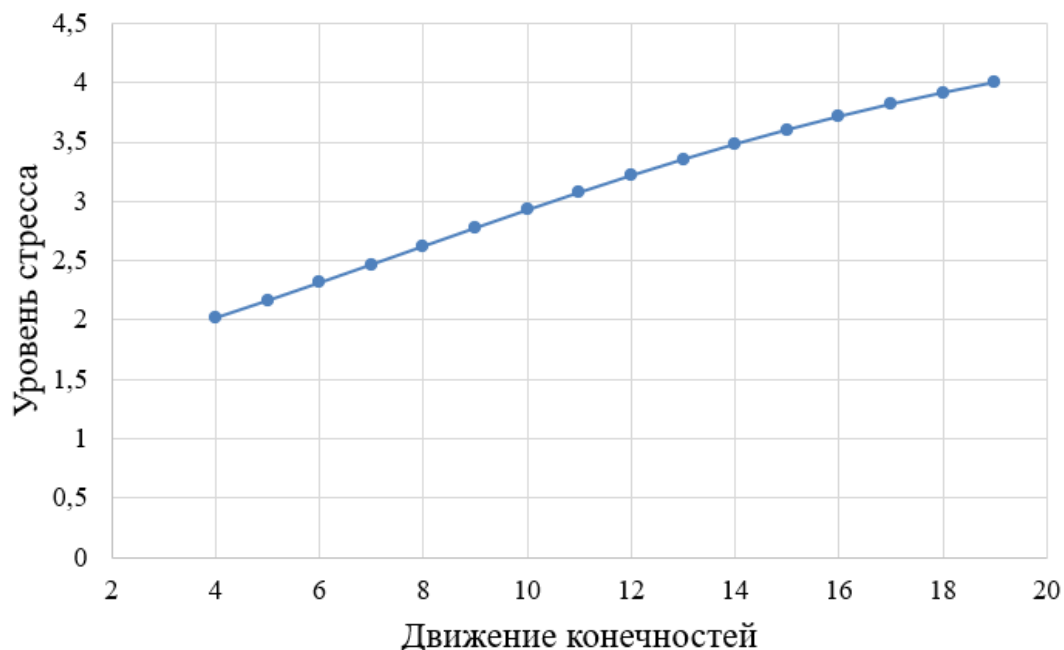
**Рисунок 4. Зависимость прогнозируемого уровня стресса от частоты дыхания**

На рисунке 5 изображен график зависимости уровня стресса от температуры тела. Как видно из графика, чем выше температура тела, тем ниже уровень стресса человека. Температура тела варьируется от 85 до 99 градусов Фаренгейта, что составляет от 29 до 37 градусов Цельсия. 29 градусов Цельсия – это пониженная температура тела, соответственно чем выше температура, тем она ближе к нормальной и уровень стресса ниже.



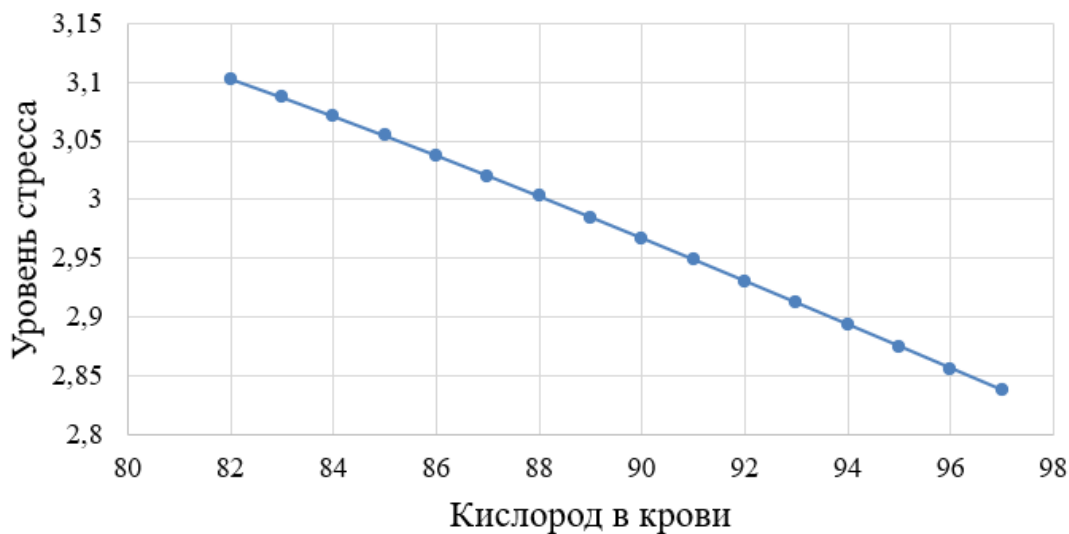
**Рисунок 5. Зависимость прогнозируемого уровня стресса от температуры тела**

На рисунке 6 продемонстрирована зависимость прогнозируемого уровня стресса от движения конечностей во сне. Как видно из графика, с увеличением уровня движения конечностей повышается уровень стресса.



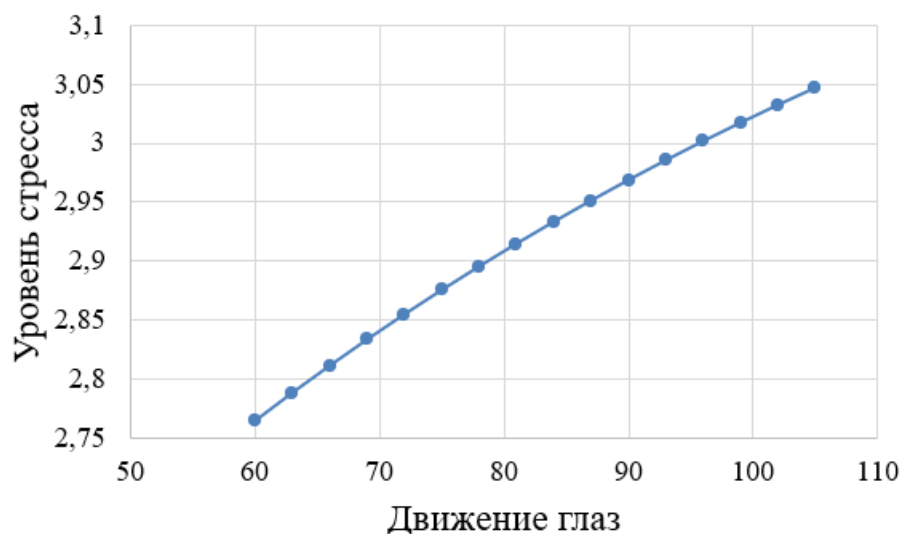
**Рисунок 6. Зависимость прогнозируемого уровня стресса от движения конечностей**

На рисунке 7 продемонстрирована зависимость прогнозируемого уровня стресса от уровня кислорода в крови. Кислород жизненно необходим всем клеткам тела, если уровень кислорода в крови у человека снижается, это грозит серьезными последствиями для организма. Как видно из диаграммы на рисунке 8, чем выше уровень кислорода в крови человека, тем, соответственно, ниже уровень его стресса.



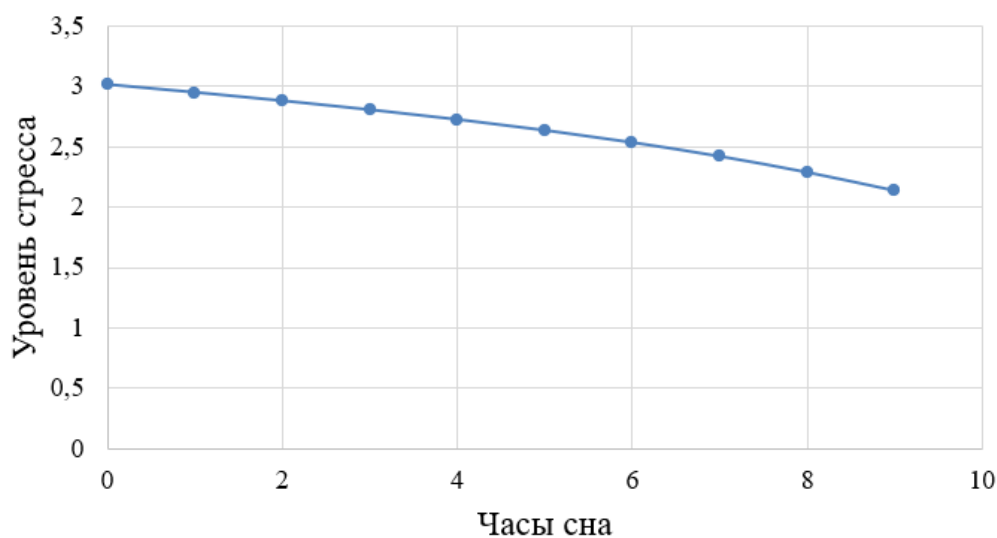
**Рисунок 7. Зависимость прогнозируемого уровня стресса от уровня кислорода в крови**

На рисунке 8 продемонстрирована зависимость прогнозируемого уровня стресса от количества движений глаз человека во время сна. С увеличением количества движений глаз повышается прогнозируемый уровень стресса человека.



**Рисунок 8. Зависимость прогнозируемого уровня стресса от движения глаз во время сна**

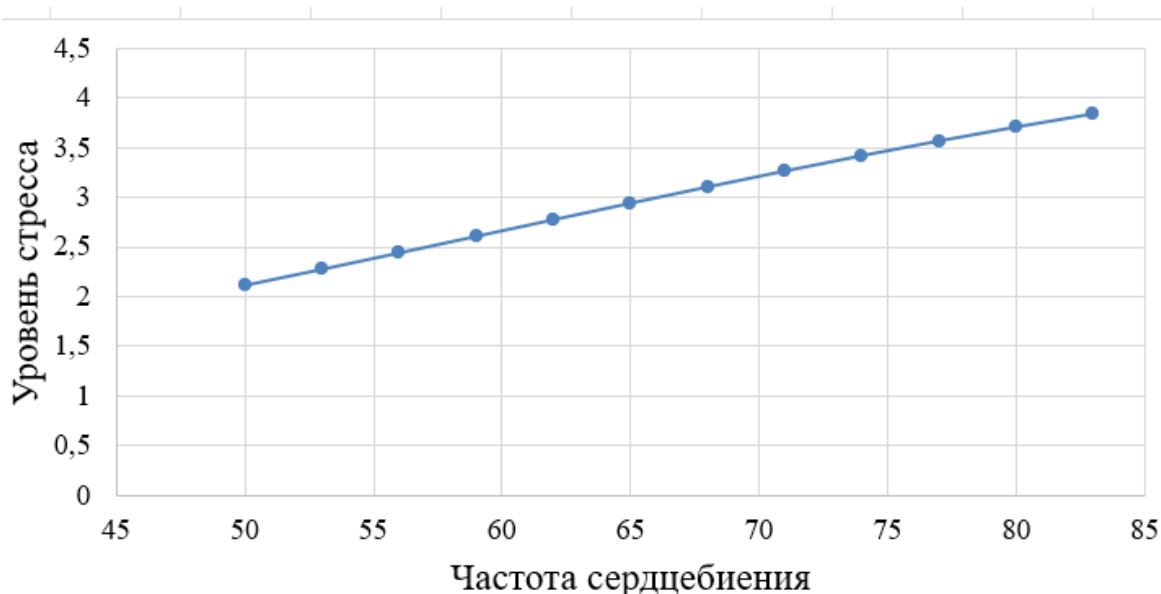
На рисунке 9 показан график зависимости прогнозируемого уровня стресса от количества часов сна человека. В том случае, когда у человек спал недостаточное количество времени, нейросеть прогнозирует увеличение уровня стресса.



**Рисунок 9. Зависимость прогнозируемого уровня стресса от количества часов сна**

На рисунке 10 продемонстрирована зависимость прогнозируемого уровня стресса от частоты сердцебиения человека во время сна, то есть пульса. Можно заметить, что чем выше пульс человека, тем выше прогнозируе-

мый уровень стресса. Во время сна снижается активность всех органов, уменьшается частота сердечных сокращений, замедляется обмен веществ, а вместе с этим и падает частота ударов. Учащенный пульс нередко вызывает нарушения сна.



**Рисунок 10. Зависимость прогнозируемого уровня стресса от частоты сердцебиения**

Полученные результаты исследований не противоречат реальности, что подтверждает, что спроектированную нейронную сеть можно считать пригодной для прогнозирования уровня стресса человека на основании показателей, полученных во время сна.

**Заключение.** Построена система нейросетевого прогнозирования уровня человеческого стресса через сон. Спроектированная нейросетевая модель учитывает 8 параметров: уровень храпа, частота дыхания, температура тела, движение конечностей, кислород в крови, движение глаз, часы сна, частота сердцебиения. Методом сценарного прогнозирования построены графики зависимостей прогнозируемого уровня стресса от изменения входных параметров. Применение такого набора параметров в модели позволяет с высокой точностью прогнозировать уровень стресса.

### **Библиографический список**

1. L. Rachakonda, A. K. Bapatla, S. P. Mohanty, and E. Kougianos, “SaY-oPillow: Blockchain-Integrated Privacy-Assured IoMT Framework for Stress Management Considering Sleeping Habits”, IEEE Transactions on Consumer Electronics (TCE), Vol. 67, No. 1, Feb 2021, pp. 20-29.
2. L. Rachakonda, S. P. Mohanty, E. Kougianos, K. Karunakaran, and M. Ganapathiraju, “Smart-Pillow: An IoT based Device for Stress Detection Considering Sleeping Habits”, in Proceedings of the 4th IEEE International Symposium on Smart Electronic Systems (iSES), 2018, pp. 161--166.

3. Kaggle.com – Your Home For Data Science. [Электронный ресурс]. Режим доступа: <https://www.kaggle.com/>

4. Черепанов Ф.М., Ясницкий Л.Н. Нейросимулятор 4.0. Свидетельство о регистрации программы для ЭВМ RUS 2014612546. Заявка № 2014610341 от 15.01.2014.

5. Ясницкий Л.Н. Введение в искусственный интеллект. М.: Издательский центр «Академия», 2005. – 176 с.

## NEURAL NETWORK SYSTEM FOR PREDICTING HUMAN STRESS THROUGH SLEEP

*Syrvacheva Ekaterina R.*

National Research University “Higher School of Economics”,  
614990, Russia, Perm, st. Gagarin Boulevard, 37, [ersyvacheva@edu.hse.ru](mailto:ersyvacheva@edu.hse.ru)

The article presents a description of the development of a neural network system for predicting the level of human stress during sleep. The system makes it possible to predict with great accuracy the level of human stress in and through sleep by monitoring its physiological data. With the help of the developed intellectual system, a study of the subject area was carried out, patterns of practical importance were identified.

**Keywords:** artificial intelligence, neural network technologies, forecasting, stress, sleep.

УДК 30.1+519.23

## НЕЙРОСЕТЕВАЯ СИСТЕМА ПРОГНОЗИРОВАНИЯ ЗАПОЛНЕННОСТИ ЗАЛА НА СПЕКТАКЛЕ В ТЕАТРЕ-ТЕАТРЕ

**Аухадиев Михаил Русланович**

Пермский государственный национальный исследовательский университет,  
ПМИ. 614990, Россия, г. Пермь, ул. Букирева, 15.  
[tyver9@gmail.com](mailto:tyver9@gmail.com)

В статье представлено описание разработки нейросетевой системы для прогнозирования заполненности зала в Театр Театре. Система позволяет с большой точностью предсказать на сколько будет заполнен зрительный зал в Театр Театре на основании данных о спектакле. С помощью разработанной интеллектуальной системы проведено исследование предметной области, выявлены закономерности, имеющие практического значения.

**Ключевые слова:** искусственный интеллект, нейросетевые технологии, прогнозирование, театр, зрители.



**Введение:** прогнозирование заполненности зала, хоть и не является исследованием претендующем на научную значимость, сможет помочь, как театру, понять какая нагрузка ляжет на “обслуживающий” персонал на спектакле, какие спектакли способны собирать аншлаги и от чего зависит популярность спектакля. так и обычным любителям культурно провести вечер в театре, не знаящим стоит купить билет заранее или можно повременить и дожидаться скидок, либо же не попасть в ситуации, когда в желании попасть на спектакль следить за сайтом театра в надежде, что в какой-то момент наконец появится свободное место.

Основная цель настоящей работы заключается в сборе множества публичных данных о различных спектаклях в репертуаре Театр Театра и количестве купленных билетов на них, а также создание и обучение нейросетевой модели на этих данных. Конечным результатом должна быть нейросетевая система, способная корректно прогнозировать процент заполненности зала.

Для создания нейросетевой системы были выбраны следующие параметры:

- X1- название спектакля
- X2 – дата показа (месяц)
- X3 – дата показа (день недели)
- X4 – время начала показа
- X5 – возрастное ограничение
- X6 – продолжительность спектакля в минутах
- X7 – сколько уже лет показывают спектакль на сцене
- X8 – жанр спектакля

Выходной параметр – процент занятых мест в зрительном зале.

Так как данные о количестве занятых мест на спектаклях нельзя найти в интернете обучающее множество было собрано после разговора с администрацией и получения от нее для данного проекта данных о заполняемости зала за 2019 год. Изначально в множестве было собрано 337 примера. Так как театр периодически проводит акции и скидки в множество попали противоречащие примеры, заполненность зала в которых зависела не столько от заданных входных данных, сколько от этих самых акций, так что перед началом обучения перед переходом к проектированию и обучению нейросетевой модели была необходима очистка исходного множества от противоречивых примеров и выбросов. Обучив нейронную сеть на основе обучающего и валидирующего множеств, была проведена проверка обучающего множества (вместо тестового) для поиска выбросов в обучающем множестве. Результат этих действий представлен на рисунке 1.

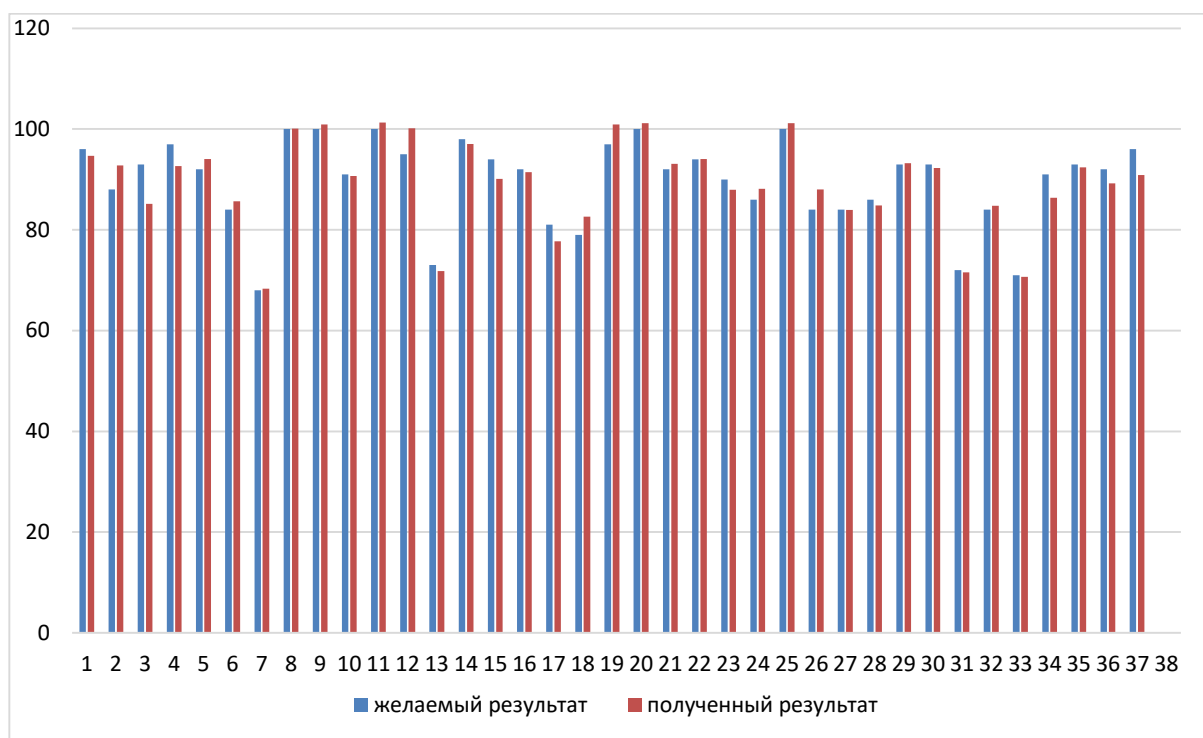
Результат обучения на исходном множестве: обучив систему на основе исходного множества, был получена неточная система с высоким процентом ошибок, гистограмма разницы между правильным результатом и полученным и на основе тестового множества представлена на рисунке 1. Самыми некорректными являются результаты тестов на примерах 9, 70 и 31.

Было принято решение очистить множество от некорректных примеров: удалялись примеры с самыми высокими погрешностями обучения, после чего

система заново обучалась и проводилась оценка результатов тестирования на общем множестве до получения удовлетворительных погрешностей [1].

Таким образом, объем итогового очищенного множества включает в себя 304 пример. Данное множество было разделено на обучающее, валидирующее и тестирующее в соотношении 13%, 75% и 12% соответственно.

Проектирование, обучение, тестирование нейронной сети выполнялись с помощью программы «Нейросимулятор 5» [2] по методике [3]. В результате оптимизации спроектированная нейронная сеть представляла собой персептрон, имеющий восемь, один выходной и один скрытый слой.



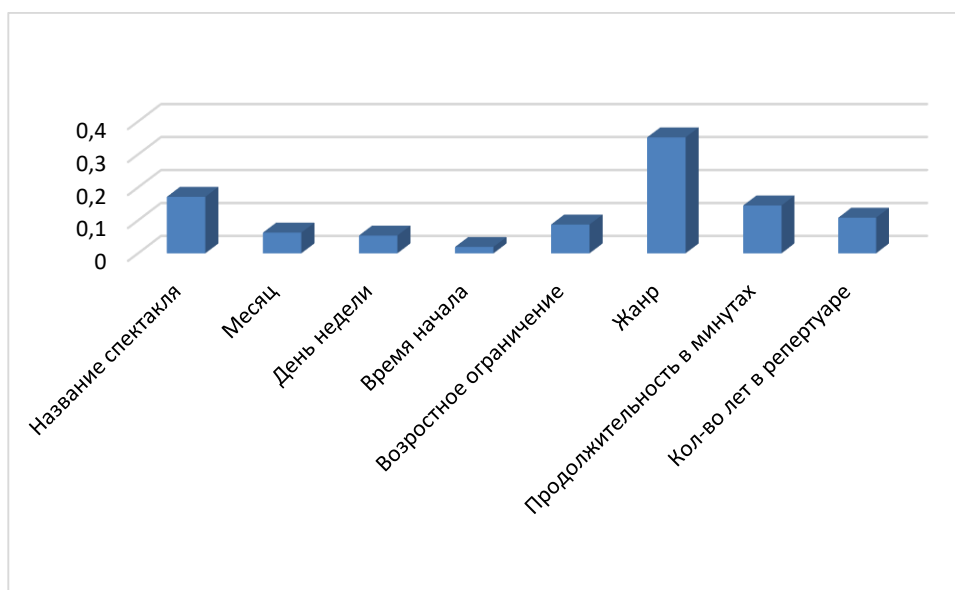
**Рисунок 1. Результат работы нейронной сети на тестовом множестве**

Для оценки корректной работы спроектированной нейронной сети использовалось тестирующее множество, состоящее из 21 примера. Средняя относительная ошибка тестирования составила 1.8%, что можно считать приемлемым результатом. На рисунке 2 представлена гистограмма, демонстрирующая разницу между фактическим и прогнозируемым нейросетью процентом заполненности зала.

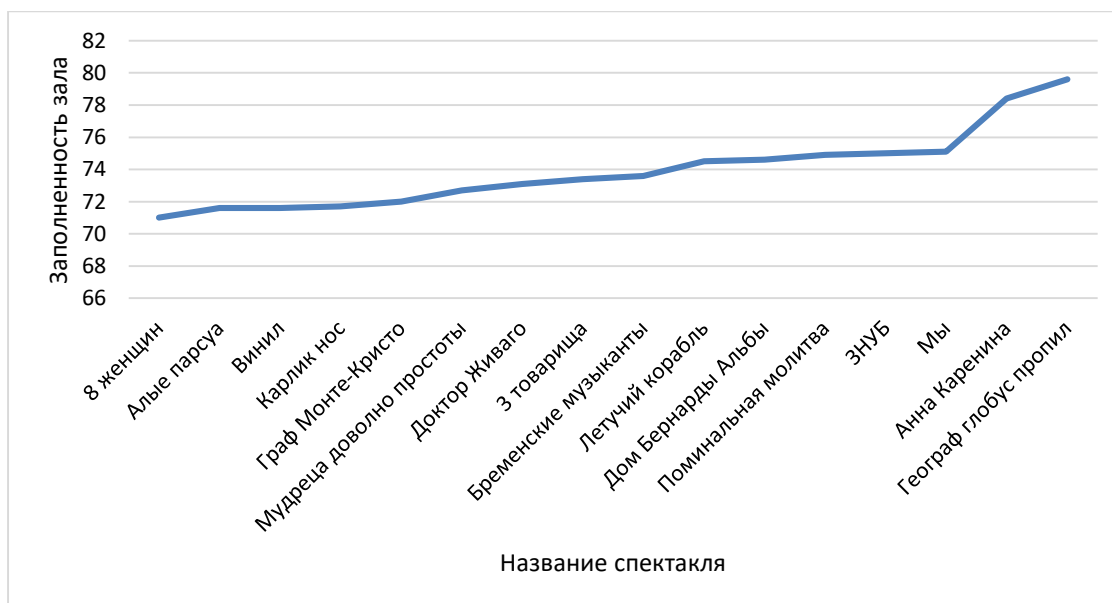
Далее было проведено исследование полученных зависимостей между входными параметрами и результатом прогнозирования. Исследование производилось с помощью метода «замораживания»[3], суть которого заключается в варьировании значения одного параметра и фиксировании значений всех других параметров. Для этой цели был отобран «нейтральный пример» для которого нейронная сеть не может дать однозначно высокий или низкий результат.

Как видно из рисунка 3 самыми значимыми параметрами оказались жанр и название спектакля, что не оставляет вопросов, так как люди в большинстве своем предпочитают что-то более легкое для восприятия, веселое, а

именитые спектакли способны привлекать новых зрителей год за годом одним своим именем. Неожиданно, что самым малозначимым оказалось время начала, так как большинство спектаклей проходят по будням, когда люди зачастую находятся на работе и не способны прийти в театр.



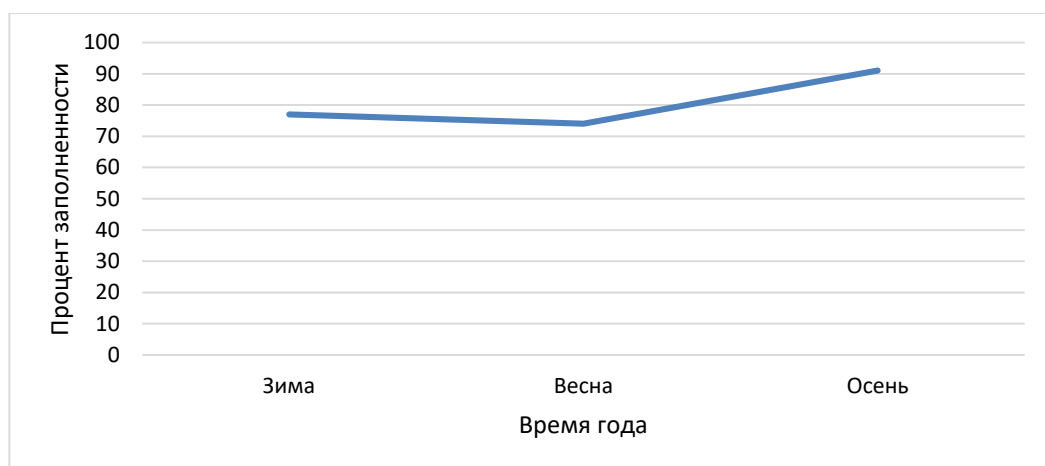
**Рисунок 2. Значимость входных параметров нейросетевой модели**



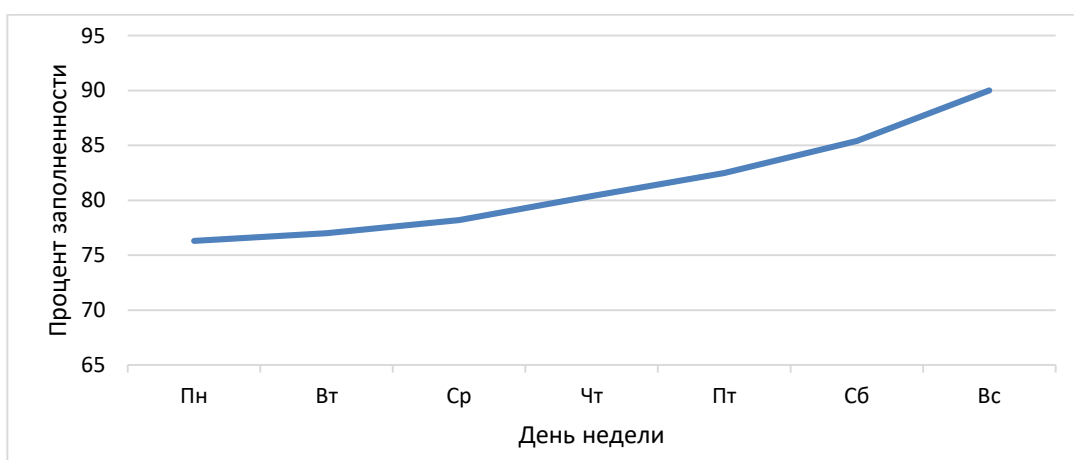
**Рисунок 3. Зависимость прогнозируемого результата от именитости спектакля**

На рисунке 4 мы видим, что лучше все собирает зрителей спектакль “Географ глобус пропил”. Вероятно это связано с тем, что он поставлен по достаточно известному произведению, но ставится крайне редко.

По данному графику мы видим, что лучше всего зрители ходят на спектакли осенью, а хуже всего весной.

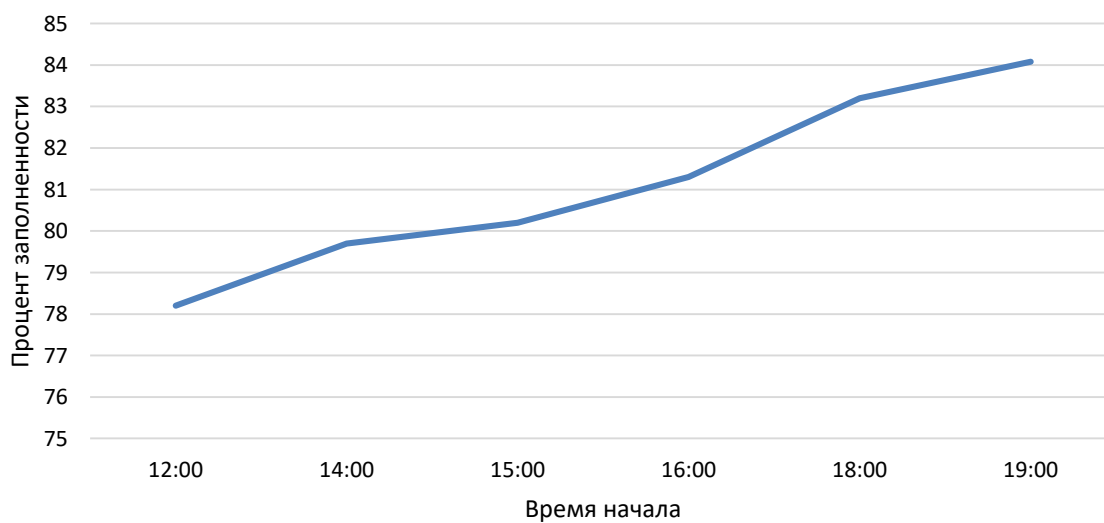


**Рисунок 4. Зависимость прогнозируемого результата от времени года показа**



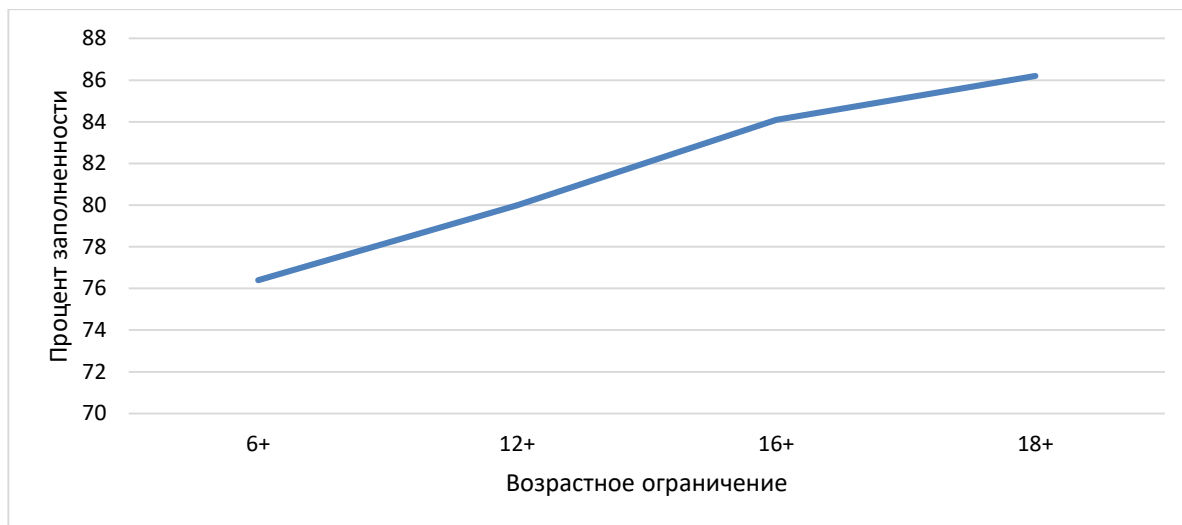
**Рисунок 5. Зависимость прогнозируемого результата от дня недели показа**

По данному графику мы видим, что больше всего зрители ходят в театр **ПО ВЫХОДНЫМ.**



**Рисунок 6. Зависимость прогнозируемого результата от времени начала спектакля**

На данном график мы видим, что к вечеру в театр ходит больше зрителей.



**Рисунок 7. Зависимость прогнозируемого результата от возрастного ограничения спектакля**

На данном графике мы видим, что спектакли, ориентированные на детей, собирают меньше зрителей.



**Рисунок 8. Зависимость прогнозируемого результата от жанра спектакля**



**Рисунок 9. Зависимость прогнозируемого результата продолжительности спектакля**

На данном графике мы видим, что драма собирает значительно меньше зрителей, чем мюзиклы и музыкальные сказки.

На данном графике мы видим, что продолжительность спектакля не оказывает значительного влияния на заполняемость зала



**Рисунок 10. Зависимость прогнозируемого результата от того, сколько лет спектакль находится в репертуаре театра**

На данном графике мы видим, что самые новые спектакли больше привлекают зрителей, с годами же интерес к спектаклю может падать.

**Заключение:** Построена система нейросетевого прогнозирования степени заполненности зала на спектакле в Театре Театре. Спроектированная нейросетевая модель учитывает 8 параметров: время года, когда ставится спектакль, день недели, что это именно за спектакль, время начала, возрастное ограничение, продолжительность спектакля, жанр, сколько лет прошло с премьеры. Методом сценарного прогнозирования построены графики зависимостей прогнозируемого процента занятых мест от изменения входных параметров. Применение такого набора параметров в модели позволяет с высокой точностью прогнозировать результат.

### **Библиографический список**

1. Yasnitsky L.N. Algorithm for searching and analyzing abnormal observations of statistical information based on the Arnold – Kolmogorov – Hecht-Nielsen theorem // International Journal of Advanced Trends in Computer Science and Engineering. 2020. 9(2). Pp. 1814-1819. DOI: 10.30534/ijatcse/2020/139922020 <http://www.warse.org/IJATCSE/static/pdf/file/ijatcse139922020.pdf>

2. Черепанов Ф.М., Ясницкий Л.Н. Нейросимулятор 4.0. Свидетельство о регистрации программы для ЭВМ RUS 2014612546. Заявка № 2014610341 от 15.01.2014.

3. Ясницкий Л.Н. Введение в искусственный интеллект. М.: Издательский центр «Академия», 2005. – 176 с.

# NEURAL NETWORK SYSTEM FOR PREDICTION OF HALL OCCUPANCY AT THE PERFORMANCE IN THE THEATER-THEATRE

*Aukhadiev Mikhail R.*

Perm State National Research University,  
614990, Russia, Perm, st. Bukirev, 15.  
tyver9@gmail.com

The article presents a description of the development of a neural network system for predicting the occupancy of the hall in the Theater Theatre. The system allows you to predict with great accuracy how much the auditorium in the Theater will be filled based on performance data. With the help of the developed intellectual system, a study of the subject area was carried out, patterns of practical importance were identified.

**Keywords:** artificial intelligence, neural network technologies, forecasting, theatre, spectators.

УДК 004.89

## НЕЙРОННЫЕ СЕТИ И УПРАВЛЕНИЕ КАЧЕСТВОМ ПРОИЗВОДСТВЕННЫХ ПРОЦЕССОВ

*Силина Александра Андреевна, Постных Денис Витальевич*

Пермский техникум промышленных  
и информационных технологий им. Б.Г. Изгагина

*Мезенцев Алексей Сергеевич*

Группа компаний ITPS, г.Пермь

*Черепанов Федор Михайлович*

Институт информационных систем, г.Пермь

*Морозов Андрей Александрович, Голдобин Максим Алексеевич*

ОДК-Пермские моторы

*Ясницкий Виталий Леонидович*

Пермский национальный исследовательский политехнический университет

*Ясницкий Леонид Нахимович*

Пермский государственный национальный исследовательский университет  
yasn@psu.ru

Сообщается о положительном опыте применения нейросетевых технологий для моделирования производственного процесса, что позволило определить регламент на технологические параметры, обеспечивающий

минимальную вероятность образования брака и максимально высокое качество производимых изделий.

**Ключевые слова:** искусственный интеллект, нейросетевые технологии, прогнозирование, качество изделий, регламент, рудное сырье.

Во все времена люди боролись за качество производимых изделий. Применение для этих целей математических методов восходит к цивилизациям Египта, Китая, Греции, Римской империи.

В 20 веке пользовались популярностью методы, основанные на решении краевых задач математической физики [1, 2]. Однако, несмотря на высокую сложность и красоту математического аппарата, они имели весьма серьезный недостаток. Это – низкая точность моделирования, причиной которой являлась необходимость ввода грубых физических допущений, связанных с ограничениями на количество входных параметров математических моделей и, как следствие, невозможностью учета всего многообразия явлений, сопровождающих реальные производственные процессы.

Новые возможности открывает применение методов машинного обучения, в частности – нейросетевые технологии. Они не имеют ограничения на количество входных параметров. Отпадает необходимость ввода упрощающих гипотез. Нужно только правильно сформулировать математическую задачу, включив в число входных параметров как можно большее их количество, не упустив даже самые, казалось бы, несущественные факторы, способные оказывать влияние на прогнозируемые величины.

Надо отметить, что применение методов машинного обучения для решения проблем качества производимых изделий уже происходит во всем мире. Так в недавно опубликованной статье [3] дан обзор 127 научных работ, посвященных обнаружению дефектов, прогнозированию их возникновения, мониторингу и управлению качеством изделий. Однако, среди этих работ не встречается ни одной, в которой бы нейронные сети использовались для назначения регламента на технологические параметры, обеспечивающего минимум брака и максимум качества изделий. Именно решению этой важной в практическом отношении проблемы посвящена настоящая работа.

В качестве моделируемого технологического процесса мы взяли процесс изготовления керамических изделий ответственного назначения. Было выделено около сотни входных параметров. Это параметры, характеризующие: рудное сырье, его состав и процесс приготовления сырьевой массы; процесс прессования керамического изделия в металлической форме; процесс обжига изделия, процесс его упаковки и пр.

В качестве выходного параметра модели был выбран параметр, характеризующий отклонение геометрической формы изготавливаемого изделия от заданной формы – коробление в наиболее ответственном сечении.

Было очень важно собрать достаточное количество примеров поведения моделируемого производственного процесса, допустив при этом минимальное количество ошибок – так называемых статистических выбросов. В нашем случае эту задачу успешно решили студенты Пермского техникума



промышленных и информационных технологий имени Б.Г. Изгагина. За время производственной практики ими были собраны около пяти тысяч данных о составе и параметрах используемого рудного сырья и производственных процессах ОДК Пермские моторы.

Собранный набор данных (dataset) содержал некоторое количество статистических выбросов, для обнаружения и ликвидации которых использовался метод [4].

Применение Нейросимулятора 5.0 [5] позволило создать оптимальную нейронную сеть, показавшую значение коэффициента детерминации с реальными данными величины коробления керамического изделия  $R^2 = 92\%$ .

Виртуальные компьютерные эксперименты, выполненные над нейросетевой математической моделью с применением метода замораживания входных параметров, позволили построить кривые зависимости величины коробления от наиболее значимых технологических параметров. На основании этих кривых был найден регламент на эти параметры, обеспечивающий минимальную вероятность брака и максимальное качество производимых изделий.

Еще раз отметим, что, как следует из обзора [4] для назначения регламента на технологические параметры, нейронные сети, по-видимому, до сих пор не использовались.

Исследование выполнено за счет гранта Российского научного фонда № 22-61-00096, <https://rscf.ru/project/22-61-00096/>

### Библиографический список

1. Ясницкий Л.Н. Суперпозиция базисных решений в методах типа трэффтца // Известия Академии наук СССР. Механика твердого тела. 1989. № 2. С. 95-101.
2. Самойлович Ю.А., Ясницкий Л.Н. Сопряженная задача теплообмена, гидродинамики и затвердевания расплава // Инженерно-Физический журнал. – 1981. – Т.ХLI. – №6. – С. 1109-1118.
3. Selvaraj S.K., Раджу А., Ришикеш М. Р., Чадхе У., Paramasivam V. A Review on Machine Learning Models in Injection Molding Machines // Advances in Materials Science and Engineering, 2022, vol. 2022, article number 1949061. DOI: 10.1155/2022/1949061
4. Yasnitsky L.N. Algorithm for searching and analyzing abnormal observations of statistical information based on the Arnold – Kolmogorov – Hecht-Nielsen theorem // International Journal of Advanced Trends in Computer Science and Engineering. 2020. 9(2). Pp. 1814-1819. DOI: 10.30534/ijatcse/2020/139922020
5. Нейросимулятор 5.0 : Свидетельство Роспатент о государственной регистрации программы для ЭВМ № 2014618208 от 12.07.2014 г. / Черепанов Ф.М., Ясницкий Л.Н. (РФ). <https://cloud.mail.ru/public/Jr29/RzgouF6b9>.

# NEURAL NETWORKS AND QUALITY MANAGEMENT OF PRODUCTION PROCESSES

*Silina Alexandra A., Postnykh Denis V.*

Perm College of Industrial and Information Technologies n.a. B.G. Izgagin

*Mezentsev Aleksey S.*

ITPS group of companies, Perm

*Cherepanov Fedor M.*

Institute of Information Systems, Perm

*Morozov Andrei A., Goldobin Maxim A.*

UEC-Perm Motors

*Yasnitsky Vitaly L.*

Perm National Research Polytechnic University

*Yasnitsky Leonid N.*

Perm State National Research University

yasn@psu.ru

It is reported about the positive experience of using neural network technologies for modeling the production process, which made it possible to determine the regulations for technological parameters that ensure the minimum probability of marriage and the highest possible quality of manufactured products.

**Keywords:** artificial intelligence, neural network technologies, forecasting, product quality, regulations, ore raw materials.

УДК 004.032.26

## НЕЙРОСЕТЕВАЯ СИСТЕМА ПРОГНОЗИРОВАНИЯ УРОВНЯ УДОВЛЕТВОРЕННОСТИ ПОСЕТИТЕЛЕЙ ГОСТИНИЦ

*Поселенцева Диана Юрьевна*

Пермский государственный национальный исследовательский университет,  
614990, Россия, г. Пермь, ул. Букирева, 15, octopus.diana@gmail.com

В статье представлено описание разработки нейросетевой системы для прогнозирования удовлетворенности посетителей отелей. Система позволяет с большой точностью предсказать, будет ли удовлетворен посетитель гостиницы в зависимости от характеристик этой гостиницы и качества услуг, которые она предоставляет своим постояльцам.

**Ключевые слова:** искусственный интеллект, нейросетевые технологии, прогнозирование, гостиничный бизнес, пользовательский отзыв, онлайн-отзывы.

**Введение.** На сегодняшний день мировая индустрия туризма и гостеприимства находится в жестких условиях конкурентной борьбы за каждого клиента и вынуждена крайне стремительно реагировать на внешние изменения условий хозяйствования [1]. Гостиничный бизнес является одной из наиболее быстро развивающихся отраслей, на которую приходится 6% мирового внутреннего валового продукта и около 5% всех налоговых поступлений. Развитие гостиничного бизнеса активно стимулирует развитие других отраслей и направлений деятельности: международного бизнеса, услуг в области красоты и здоровья, транспортной индустрии, торговли, информационных технологий, строительства, сельского хозяйства, и т.п. [2]. Поэтому каждый из владельцев подобного бизнеса задумывается о том, каким образом можно обеспечить максимальный комфорт своим постояльцам, ведь от этого зависит не только благополучие клиентов, но и финансовое благосостояние владельца гостиницы. Каждый из отелей и гостиничных комплексов имеет свои методы привлечения потенциальных клиентов, имеют свой богатый и уникальный список предоставляемых услуг. При этом не всегда владельцам гостиницы удается в полной мере удовлетворить потребности своих клиентов, в виду различных причин: недостаточное количество финансовых средств для увеличения уровня обслуживания посетителей, человеческий фактор и т.п. Поэтому для поддержания конкурентоспособности гостиничного предприятия, необходимо определиться с содержанием и последовательностью конкретных действий для достижения поставленных целей, что и определяет в широком смысле процесс планирования и прогнозирования деятельности гостиниц. И если процесс планирования деятельности в большей мере ложится на плечи руководителя гостиничного комплекса, то с задачей прогнозирования ему вполне могут помочь современные информационные технологии, такие как нейросетевое моделирование. Как показал опыт Пермской научной школы искусственного интеллекта [3], правильно обученная нейросетевая система может добиться большой точности в прогнозировании, что в свою очередь может в значительной мере увеличить прибыль как владельцам небольших отелей, так и руководителям больших гостиничных комплексов по всему миру, и, соответственно, повысить уровень предоставляемых отелями услуг.

При анализе исследуемой области выяснилось, что работ на тему исследования факторов, влияющих на удовлетворенность клиента гостиницей, в которой он остановился, достаточно много [4-7]. Так, Мохамад Алаа Ктет и Елена Кметь в своей работе [4] исследовали отзывы посетителей гостиниц г. Владивостока на одном из самых известных сайтов бронирования Booking.com. Для своего исследования они выбрали 12700 положительных и отрицательных отзывов от 8476 гостей, что помогло им создать представление о детерминантах и факторах, наиболее влияющих на мнение гостей об отелях. Исследование авторы статьи проводили посредством текстовой аналитики и факторного анализа собранных отзывов, в результате чего авторы сделали следующие выводы:

1. Три главных фактора, влияющих на *удовлетворенность* посетителя: наличие удобств и персонала, доступность отеля и вид из номера.

2. Три главных фактора, влияющих на *неудовлетворенность* посетителя: качество еды, отсутствие удобств и окон в номере, состояние ванной комнаты.

Основная цель настоящей работы заключается в сборе большого множества отзывов посетителей различных отелей, а также создание и обучение нейросетевой модели на этих данных. Конечным результатом должна быть нейросетевая система, способная прогнозировать удовлетворенность посетителя отеля в больше чем 70% случаев.

Для создания нейросетевой системы были выбраны следующие параметры: X1 – пол посетителя, X2 – возраст посетителя, X3 – цель приезда, X4 – тип поездки, X5 – тип бронирования, X6 – уровень Wi-Fi соединения, X7 – удобство прибытия и выезда из отеля, X8 – удобство онлайн-бронирования номера, X9 – местоположение отеля, X10 – качество еды, X11 – уют номера, X12 – развлечения в общей комнате, X13 – удобство въезда и выезда из номера, X14 – другие услуги, X15 – чистота номера. Каждый из параметров X3-X15 оценивался посетителями отеля значениями от 1 до 5, определяющими степень удовлетворенности сервисом. При этом 1 означало «ужасный сервис», а 5 – «отличный сервис».

Обучающее множество было взято из социальной сети специалистов по обработке данных и машинному обучению Kaggle [8] и включала выборку отзывов посетителей множества европейских отелей, вынужденных поселиться в том или ином отеле. Перед переходом к проектированию и обучению нейросетевой модели, были выполнены следующие шаги:

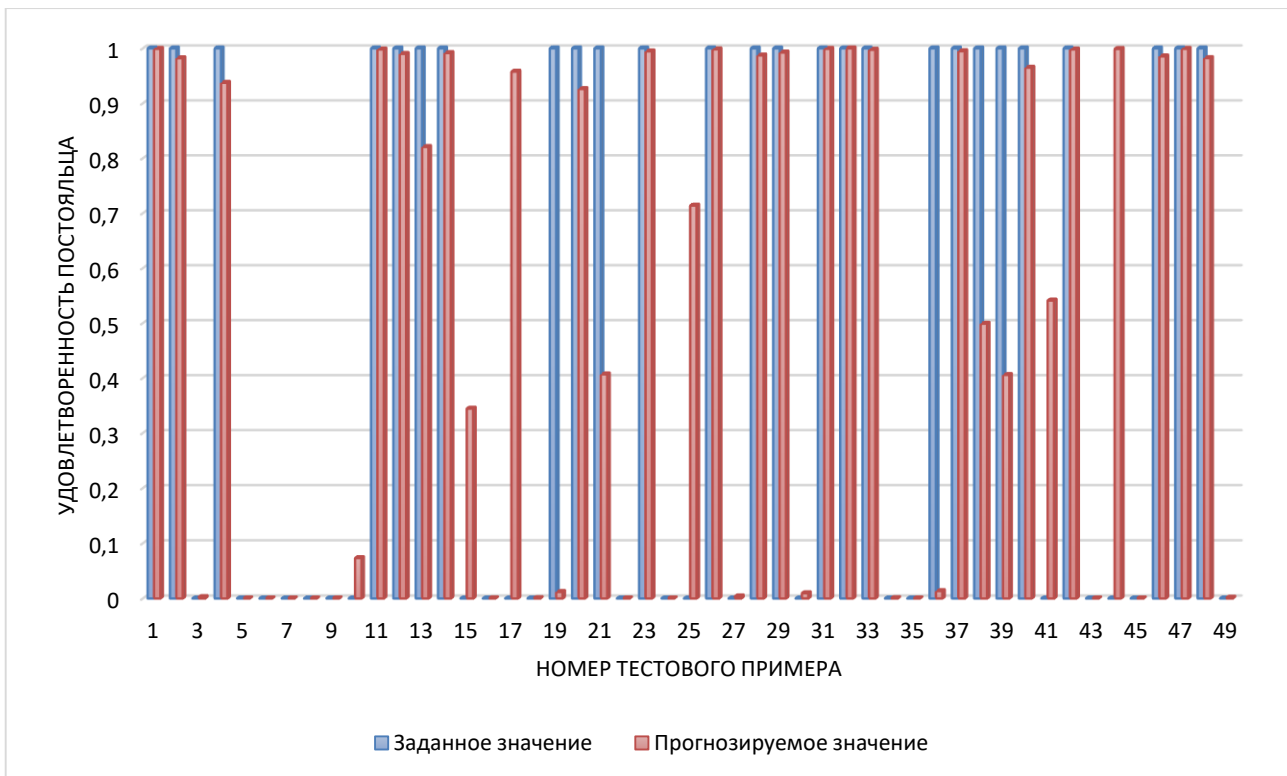
1. Очистка исходного множества от противоречивых примеров,
2. Очистка исходного множества от выбросов,
3. Очистка исходного множества от дубликатов.

Например, некорректными примерами считались такие, в которых отсутствовала информация по одному из рассматриваемых параметров или же при негативных отзывах клиентов по каждому из параметров клиент оказывался удовлетворен своим посещением.

Таким образом, объем итогового множества включает в себя 103904 примера отзывов различных посетителей гостиниц Европы. Данное множество было разделено на обучающее и тестовое множества в соотношении 70% и 30%, соответственно. Таким образом, в обучающее множество попало 72733 примера, а в тестовое – 31171 пример.

Проектирование, обучение и тестирование нейросетевой модели выполнялись с помощью программы «Нейросимулятор 5» [9] по методике [10]. В результате оптимизации спроектированная нейронная сеть представляла собой персептрон, который имеет *15 входных нейронов, один скрытый слой с 8-ью нейронами и 1 выходной нейрон*.

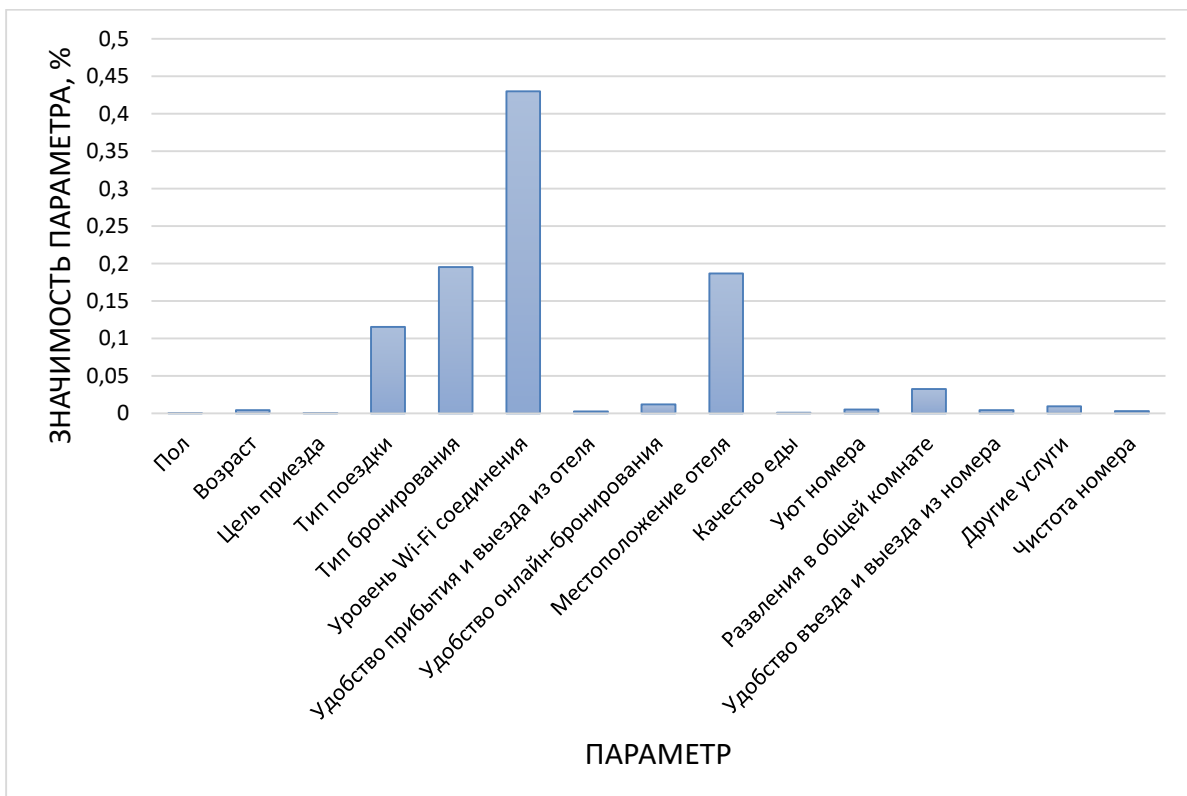
Средняя относительная ошибка тестирования составила 21,37%, что можно считать приемлемым результатом. На рисунке 1 представлена гистограмма, демонстрирующую разницу между фактическим и прогнозируемым нейросетью результатом.



**Рисунок 1. Результат тестирования обученной нейросетевой модели**

Из результатов, изображенных на рисунке 1, можно сделать вывод об адекватной работе спроектированной нейронной сети.

Оценка значимости входных параметров была выполнена с помощью программы «Нейросимулятор 5», результат отображен на рисунке 2.

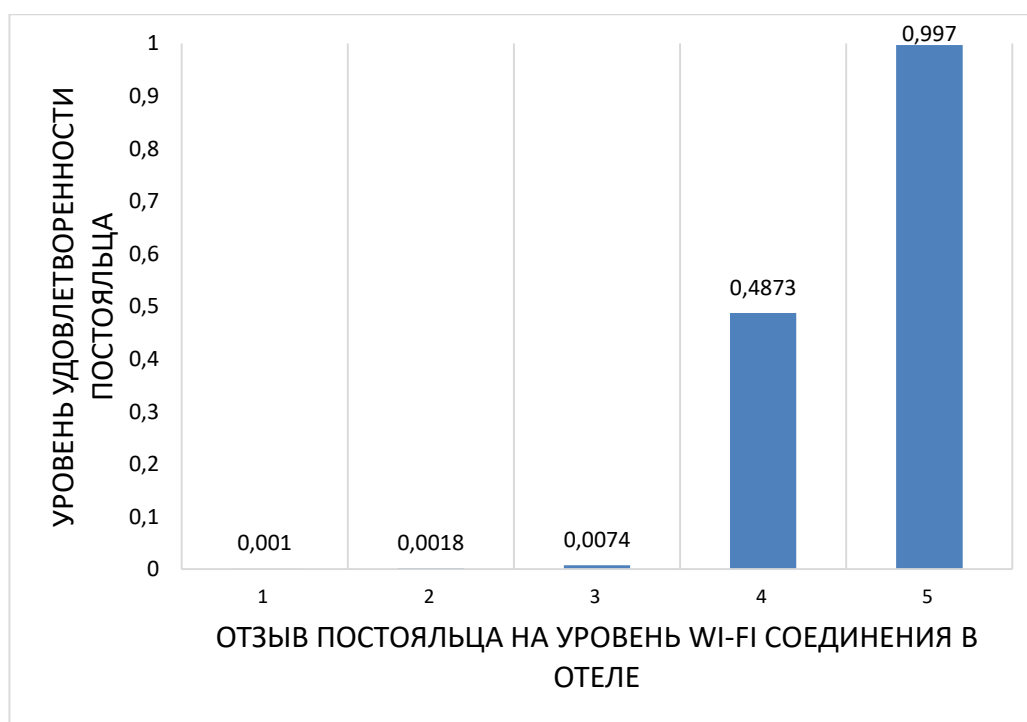


**Рисунок 2. Значимость параметров нейросетевой модели**

На основании данных, продемонстрированных на рисунке 2, наиболее значимыми параметрами нейросетевой модели являются X6 (**уровень Wi-Fi соединения**), X5 (**тип бронирования**), X9 (**местоположение отеля**) и X4 (**тип поездки**). Наиболее значимым параметром является качество Wi-Fi соединения, что вполне ожидаемо, ведь в настоящее время человек уже плохо представляет свою жизнь без Интернета и информационных технологий.

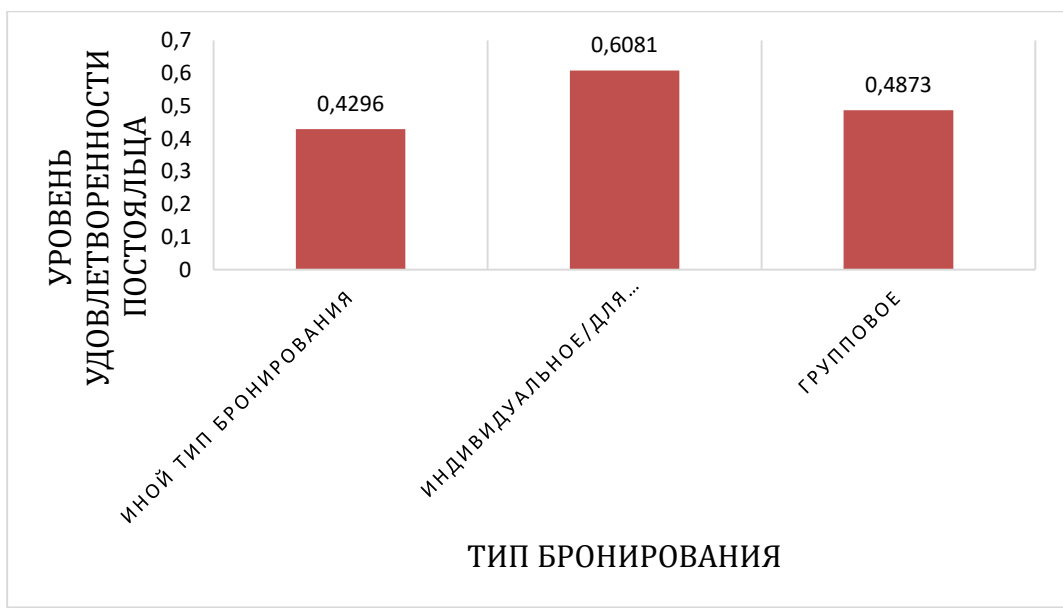
Далее было проведено исследование полученных зависимостей между входными параметрами и удовлетворенностью постояльцев отелей. Исследование проводилось с помощью метода «замораживания» [3, 10], суть которого заключается в варьировании значения одного параметра и фиксации значений всех других параметров. Для этой цели был отобран «нейтральный пример», про который нейросеть не может с уверенностью сказать, будет ли удовлетворен постоялец услугами отеля.

На рисунке 3 показан график зависимости прогнозируемого значения удовлетворенности посетителей отеля от их оценки качества Wi-Fi соединения по пятибалльной шкале. Заметна тенденция, подтверждающая высокую значимость данного параметра в спроектированной нейросетевой модели: чем выше оценивает посетитель качество Wi-Fi соединения в отеле, в котором он остановился, тем выше уровень удовлетворенности постояльца.



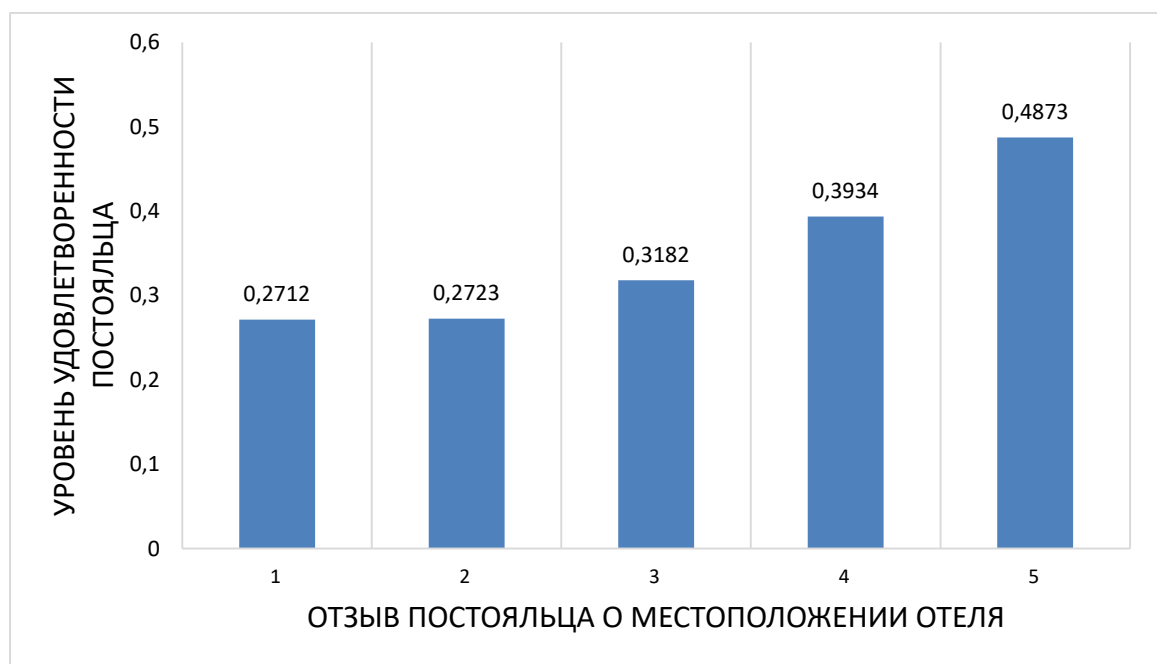
**Рисунок 3. Зависимость прогнозируемого значения от оценки посетителями уровня Wi-Fi соединения**

На рисунке 4 продемонстрирована зависимость прогнозируемого уровня удовлетворенности посетителя отеля от типа бронирования. По результатам данного исследования выяснилось, что постояльцы были удовлетворены посещением тех номеров в отелях, которые сдаются тем людям, которые приехали либо одни, либо со своей второй половиной.



**Рисунок 4. Зависимость прогнозируемого значения от типа бронирования номера**

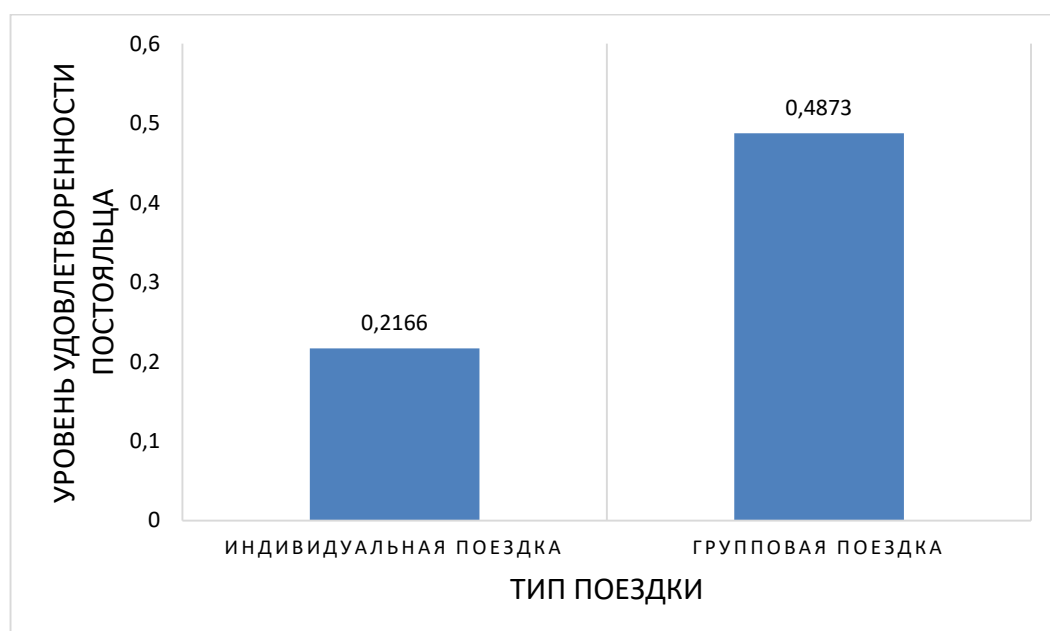
На рисунке 5 продемонстрирована зависимость прогнозируемого уровня удовлетворенности посетителей отеля от их оценки географического места положения постояльцами по пятибалльной шкале. Опять же прослеживается тенденция плавного роста прогнозируемого значения при последовательном увеличении оценки посетителя. Это говорит о том, что для постояльцев крайне важно, где территориально находится гостиница, в которой они остановились. Удобство расположения отеля гарантирует комфорт его постояльцам, что, соответственно, повышает их уровень удовлетворенности.



**Рисунок 5. Зависимость прогнозируемого значения от оценки посетителями местоположения отеля**

На рисунке 6 продемонстрирована зависимость прогнозируемого уровня удовлетворенности посетителей отеля от типа их поездки. Поездка могла быть индивидуальной или же с организованной группой (деловой, туристической и т.п.). Согласно полученным результатам, люди, поездка которых не была индивидуальной, были удовлетворены отелем больше, чем те, кто приехал в отель в гордом одиночестве.

Такие результаты могут натолкнуть на следующую мысль: вероятно, стоит организовывать какие-то групповые мероприятия именно для тех посетителей, которые приехали в отель не одни, а с друзьями, близкими, родственниками и т.п. Организация подобных мероприятий поможет привлечь больше людей, отдающих предпочтение путешествиям в компании близких людей, что в свою очередь положительно скажется на их удовлетворенности как от отеля, так и от поездки в целом, и, соответственно, поможет увеличить прибыль владельцу гостиничного комплекса.



**Рисунок 6. Зависимость прогнозируемого значения от типа поездки**

Полученные результаты исследований не противоречат реальности, что подтверждает, что спроектированную нейронную сеть можно считать пригодной для прогнозирования уровня удовлетворенности постояльцев отелей и гостиничных комплексов по всему миру.

**Заключение.** Построена система нейросетевого прогнозирования уровня удовлетворенности постояльцев отелей и гостиничных комплексов. Спроектированная нейросетевая модель учитывает 15 параметров, таких как: пол посетителя, возраст посетителя, цель приезда, тип поездки, тип бронирования, уровень Wi-Fi соединения, удобство прибытия и выезда из отеля, удобство онлайн-бронирования номера, местоположение отеля, качество еды, уют номера, развлечения в общей комнате, удобство въезда и выезда из номера, другие услуги, чистота номера. Методом сценарного прогнозирования построены графики зависимостей прогнозируемого уровня удовлетво-



ренности постояльца отеля от изменения входных параметров. Применение такого набора параметров в модели позволяет с высокой точностью прогнозировать уровень удовлетворенности постояльца отеля. В процессе исследования выявлен параметр, имеющий наибольшую значимость – качество Wi-Fi соединения в отеле.

### Библиографический список

1. Рогова И.А. Гостиничный бизнес в современных условиях: актуальные тенденции и проблемы развития // Вестник ГУУ. 2018. №6. URL: <https://cyberleninka.ru/article/n/gostinichnyy-biznes-v-sovremennyh-usloviyah-aktualnye-tendentsii-i-problemy-razvitiya> (дата обращения: 04.12.2021).

2. Федоров Р.Г. Гостиничный бизнес как составляющая современной индустрии туризма // Молодой ученый. – 2013. – №4. – С. 307-311.

3. Ясницкий Л.Н., Бржевская А.С., Черепанов Ф.М. О возможностях применения методов искусственного интеллекта в сфере туризма // Сервис plus. – 2010 – №4. – С.111-115.

4. Ктет М.А., Кметь Е.Б. Факторы, влияющие на удовлетворенность и неудовлетворенность клиентов гостиниц г. Владивостока // Вестник ТГЭУ. 2019. №2 (90). URL: <https://cyberleninka.ru/article/n/factory-vliyayuschie-na-udovletvorennost-i-neudovletvorennost-klientov-gostinits-g-vladivostoka> (дата обращения: 05.12.2021).

5. Морозова Л.С., Земскова А.А. Влияние качества обслуживания на эффективность деятельности предприятий индустрии гостеприимства // Сервис в России и за рубежом. 2017. №2 (72). URL: <https://cyberleninka.ru/article/n/vliyanie-kachestva-obslyuzhivaniya-na-effektivnost-deyatelnosti-predpriyatij-industrii-gostepriimstva> (дата обращения: 05.12.2021).

6. Морозова Л.С., Трусевич И.В., Кузнецова Е.В. Исследование взаимосвязи показателей качества обслуживания и бизнес-процессов гостиницы // Сервис в России и за рубежом. 2014. №3 (50). URL: <https://cyberleninka.ru/article/n/issledovanie-vzaimosvyazi-pokazateley-kachestva-obslyuzhivaniya-i-biznes-protsesov-gostinitsy> (дата обращения: 05.12.2021).

7. Сердюкова Н.К., Романова Л.М., Сердюков Д.А. Программа лояльности гостиничного предприятия как инструмент коммуникационной стратегии в условиях цифровой трансформации // Вестник Академии знаний. 2020. №4 (39). URL: <https://cyberleninka.ru/article/n/programma-loyalnosti-gostinichnogo-predpriyatiya-kak-instrument-kommunikatsionnoy-strategii-v-usloviyah-tsifrovoy-transformatsii> (дата обращения: 05.12.2021).

8. Kaggle – социальная сеть специалистов по обработке данных и машинному обучению. [Электронный ресурс]. Режим доступа: <https://www.kaggle.com>.

9. Черепанов Ф.М., Ясницкий Л.Н. Нейросимулятор 4.0. Свидетельство о регистрации программы для ЭВМ RUS 2014612546. Заявка № 2014610341 от 15.01.2014.

10. Ясницкий Л.Н. Интеллектуальные системы: учебник. М.: Лаборатория знаний, 2016. 221 с.

# NEURAL NETWORK SYSTEM FOR PREDICTING THE LEVEL OF SATISFACTION OF HOTEL GUESTS

*Poselentseva Diana Y.*

Perm State University

Str. Bukireva, 15, Perm, Russia, 614990, octopus.diana@gmail.com

The article describes the development of a neural network system for predicting the satisfaction of hotel visitors. The system allows predicting with great accuracy whether a hotel visitor will be satisfied, depending on the characteristics of this hotel and the quality of services it provides to its guests.

**Keywords:** artificial intelligence, neural network technologies, forecasting, hospitality, user feedback, online reviews.

УДК 004.032.26

## НЕЙРОСЕТЕВАЯ СИСТЕМА ПРОГНОЗИРОВАНИЯ ИНДЕКСА ПРОИЗВОДИТЕЛЬНОСТИ ГРАФИЧЕСКИХ ПРОЦЕССОРОВ

*Шилов Илья Алексеевич*

Пермский государственный национальный исследовательский университет, ПМИ. 614990, Россия, г. Пермь, ул. Букирева, 15, ishilov56@gmail.com

В статье представлено описание разработки нейросетевой системы для прогнозирования индекса производительности графического процессоров, используемых в электронно-вычислительных машинах. Система позволяет с большой точностью вычислить индекс производительности графического процессора на основе данных о его характеристиках. С помощью разработанной интеллектуальной системы проведено исследование предметной области, выявлены закономерности, имеющие практического значения.

**Ключевые слова:** искусственный интеллект, нейросетевые технологии, прогнозирование, графические процессоры.

**Введение.** Графический процессор – это важный компонент ЭВМ, необходимый при использовании устройства как для развлечения, так и профессиональной работы. При выборе нового графического процессора одним из важнейших факторов является его производительность. Однако, её адекватная оценка покупателем порой невозможна, особенно если графический процессор появился в продаже только недавно. Поэтому создаются рейтинги производительности, производящие оценку графического процессора относительно других графических процессоров. Такой рейтинг может помочь по-

купателям покупать графические процессоры более обдуманно. Однако, создание такого рейтинга требует обладание соответствующим графическим процессором, что не всегда является возможным.

При анализе литературных источников выяснилось, что работы на тему прогнозирования индекса производительности графических процессоров не проводились. Существующие оценки производительности создавались посредством проведения физического тестирования, но при этом не было создано специальной системы, которая могла бы прогнозировать индексы производительности графических процессоров исключительно на основе данных о технических характеристиках.

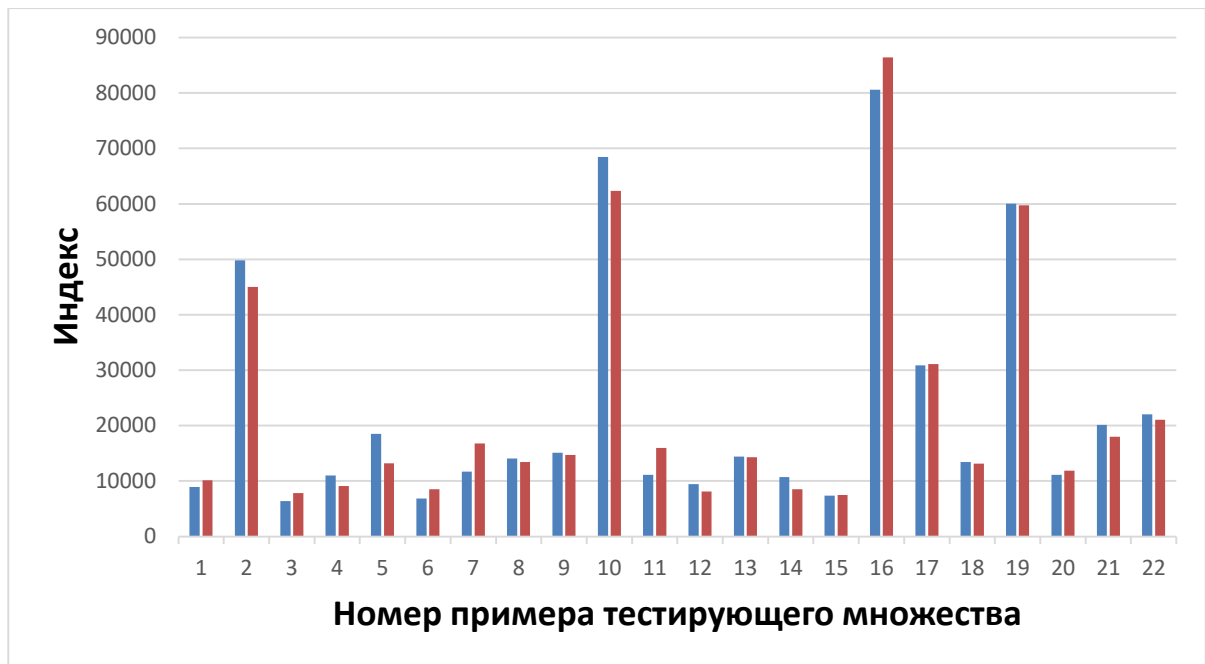
Основная цель настоящей работы заключается в сборе множества данных о характеристиках графических процессоров и сравнительного рейтинга их производительности, а также создание и обучение нейросетевой модели на этих данных. Конечным результатом должна быть нейросетевая система, способная прогнозировать индекс производительности графического процессора на основе его технических характеристик.

Для создания нейросетевой системы были выбраны следующие параметры: X1 – минимальная частота (МГц), X2 – максимальная частота, достижимая с помощью “разгона” графического процессора (МГц), X3 – размер памяти (Мбайт), X4 – пропускная способность памяти на один контакт (Мбит/с), X5 – ширина шины (бит), X6 – количество универсальных шейдерных блоков, X7 – количество текстурных блоков (TLU), X8 – количество блоков растеризации (ROP), X9 – количество ядер для выполнения трассировки лучей (RT-ядер), X10 – количество тензорных ядер. Выходной параметр – индекс производительности графического процессора.

Обучающее множество было вручную собрано с интернет-ресурсов. Перед переходом к проектированию и обучению нейросетевой модели, была выполнена очистка исходного множества от противоречивых примеров, выбросов, дубликатов. Например, некорректными примерами считались те, где у графического процессора не хватало информации о технических характеристиках. Таким образом, объем итогового множества включает в себя 220 примеров. Данное множество было разделено на обучающее и тестирующее в соотношении 90% и 10% соответственно. Большая часть данных была собрана с интернет-ресурсов [1-2].

Проектирование, обучение, тестирование нейронной сети выполнялись с помощью программы «Нейросимулятор 5» [3] по методике [4]. В результате оптимизации спроектированная нейронная сеть представляла собой персептрон, который имеет десять входных нейронов, один выходной и один скрытый слой с тремя нейронами.

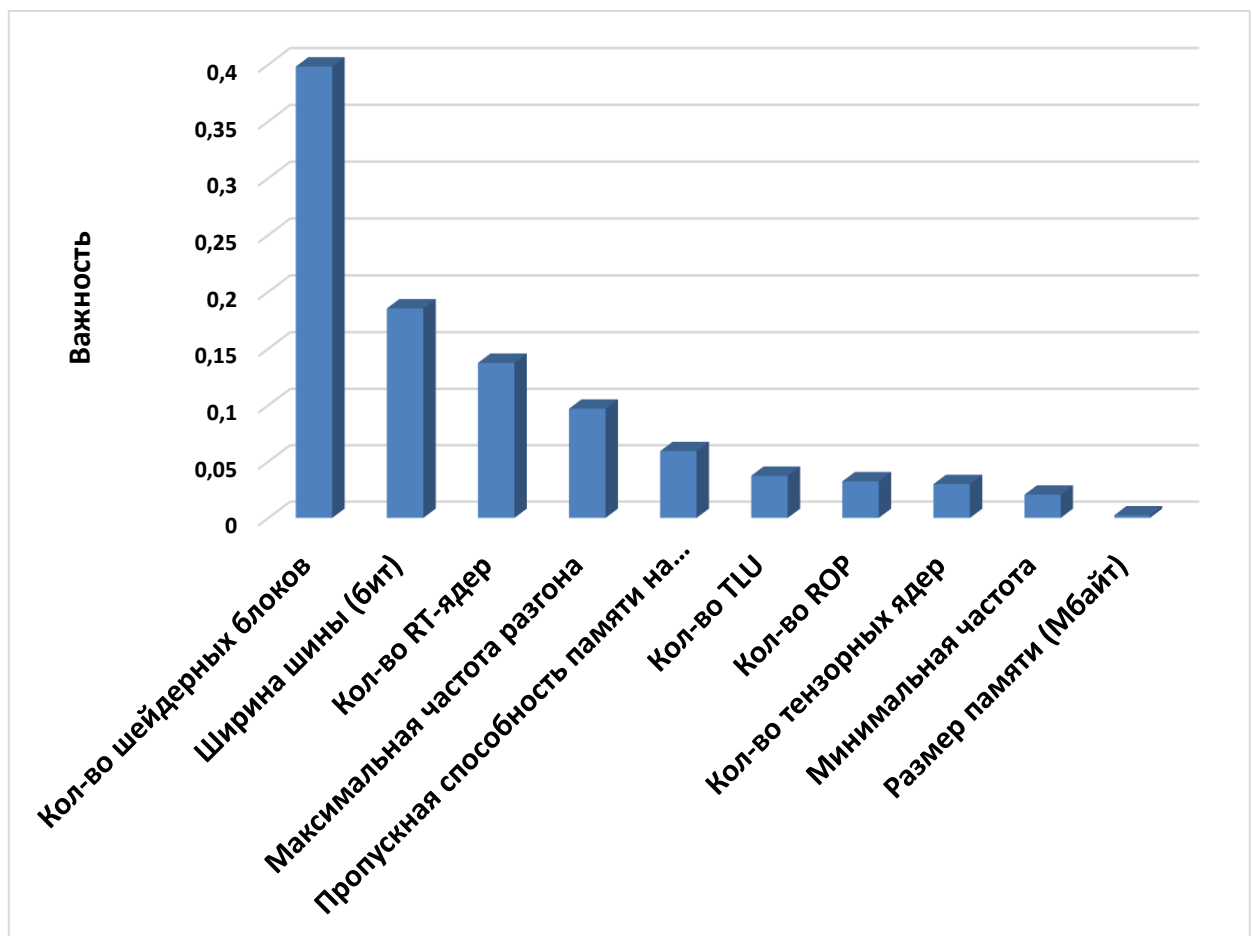
Для оценки корректной работы спроектированной нейронной сети использовалось тестирующее множество, состоящее из 22 примеров. Среднеквадратичная относительная ошибка тестирования составила 4%, что можно считать приемлемым результатом. На рисунке 1 представлена гистограмма, демонстрирующая разницу между фактическими и прогнозируемыми нейросетью индексами производительности.



**Рисунок 1. Результат тестирования нейронной сети**

Из результатов, изображенных на рисунке 1, можно сделать вывод об адекватной работе спроектированной нейронной сети.

Оценка значимости входных параметров была выполнена с помощью программы «Нейросимулятор 5», результат отображен на рисунке 2.



**Рисунок 2. Значимость входных параметров нейросетевой модели**

Как видно из рисунка 2, наиболее значимыми являются количество шейдерных блоков, ширина шины, количество ядер трассировки лучей и т.д. Наиболее влиятельным параметром является количество шейдерных блоков. Именно они отвечают за выполнение большинства вычислений, и поэтому от их количества существенно зависит производительность графического процессора.

Далее было проведено исследование полученных зависимостей между входными параметрами и индексом производительности. Исследование производилось с помощью метода «замораживания» [4], суть которого заключается в варьировании значения одного параметра и фиксировании значений всех других параметров. Для этой цели был создан «усреднённый графический процессор» со следующими характеристиками:

Минимальная частота: 1000 МГц

Максимальная частота, достижимая с помощью “разгона” графического процессора: 1200 МГц

Размер памяти: 4096 Мбайт

Пропускная способность памяти на один контакт: 7000 Мбит/с

Ширина шины: 256 бит

Количество универсальных шейдерных блоков: 2048

Количество текстурных блоков: 128

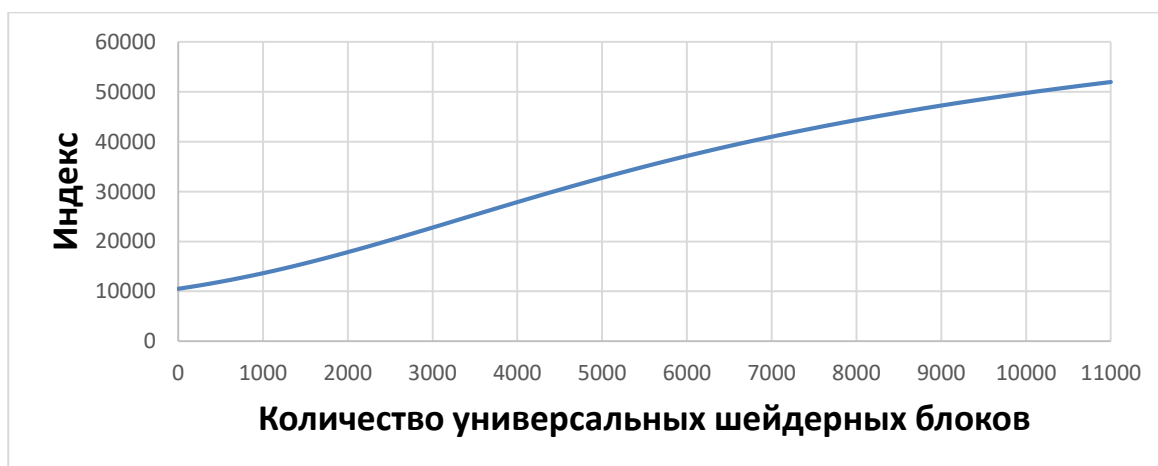
Количество блоков растеризации: 32

Количество ядер для выполнения трассировки лучей: 8

Количество тензорных ядер: 32

Всего было проведено исследование влияния следующих параметров: количество шейдерных блоков, ширина шины, количество ядер трассировки лучей, максимальная частота “разгона” и пропускная способность памяти на один контакт.

На рисунке 3 показан график зависимости прогнозируемого индекса производительности от количества универсальных шейдерных блоков. С увеличением числа универсальных шейдерных блоков происходит плавный, но при этом ощутимый рост прогнозируемого индекса производительности.

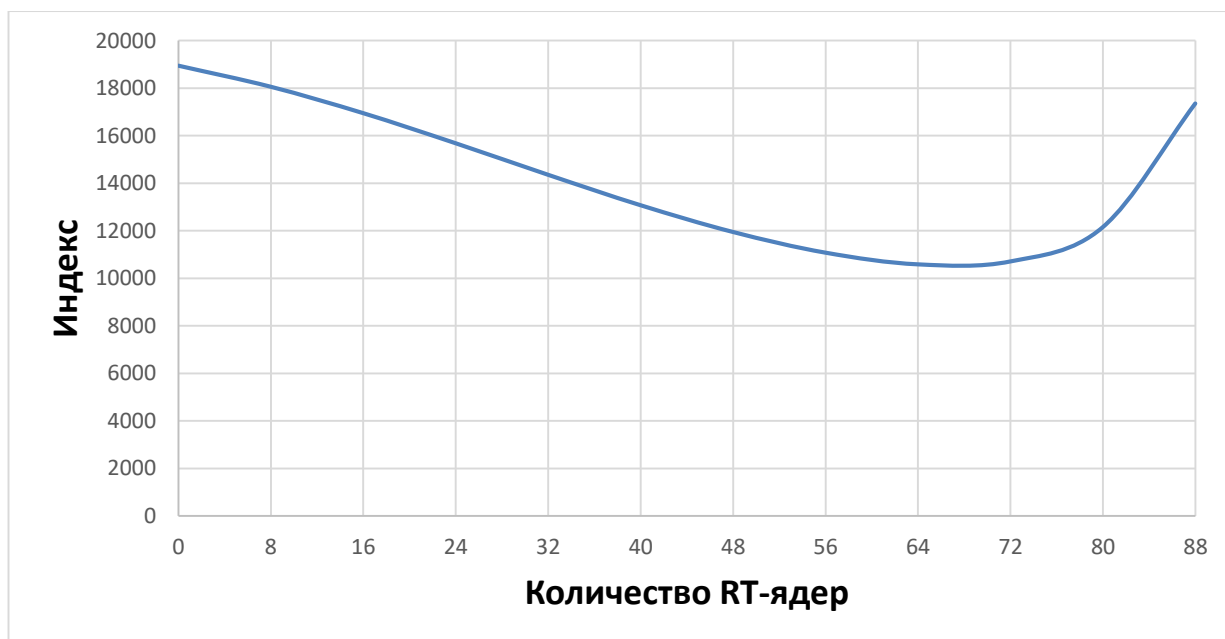


**Рисунок 3. Зависимость прогнозируемого индекса производительности от количества универсальных шейдерных блоков (X6)**

На рисунке 4 продемонстрирована зависимость прогнозируемого индекса производительности от ширины шины. При малых размерах шины (примерно до 2800 бит) происходят лишь малозначительные изменения индекса производительности, значительный скачок индекса производительности происходит только при большой ширине шины.



**Рисунок 4. Зависимость прогнозируемого индекса производительности от ширины шины (X5)**

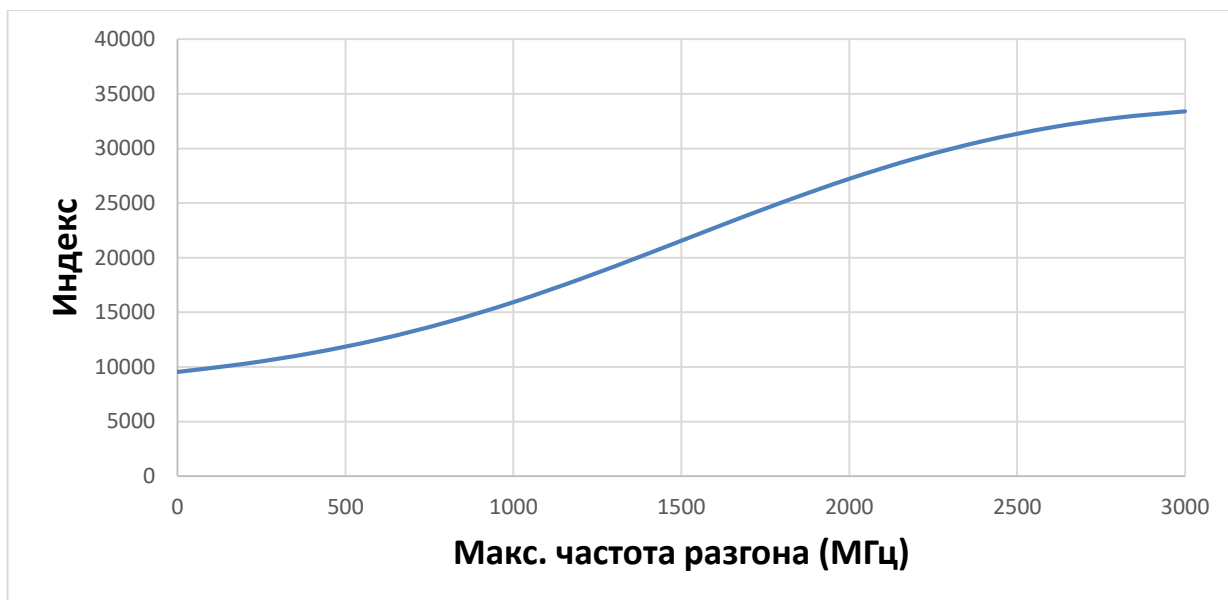


**Рисунок 5. Зависимость прогнозируемого индекса производительности от количества ядер трассировки лучей (X9)**

На рисунке 5 изображен график зависимости прогнозируемого индекса производительности от количества ядер трассировки лучей. Как видно из графика, увеличение количества ядер трассировки лучей снижает индекс производительности до определённого момента (пока RT-ядер меньше 68),

после чего начинается рост прогнозируемого индекса (при этом он так и не восстанавливается до начального значения).

На рисунке 6 продемонстрирована зависимость прогнозируемого индекса производительности от максимальной частоты “разгона”. С повышением максимальной частоты разгона происходит плавный рост индекса производительности.



**Рисунок 6. Зависимость прогнозируемого индекса производительности от максимальной частоты “разгона” (X2)**

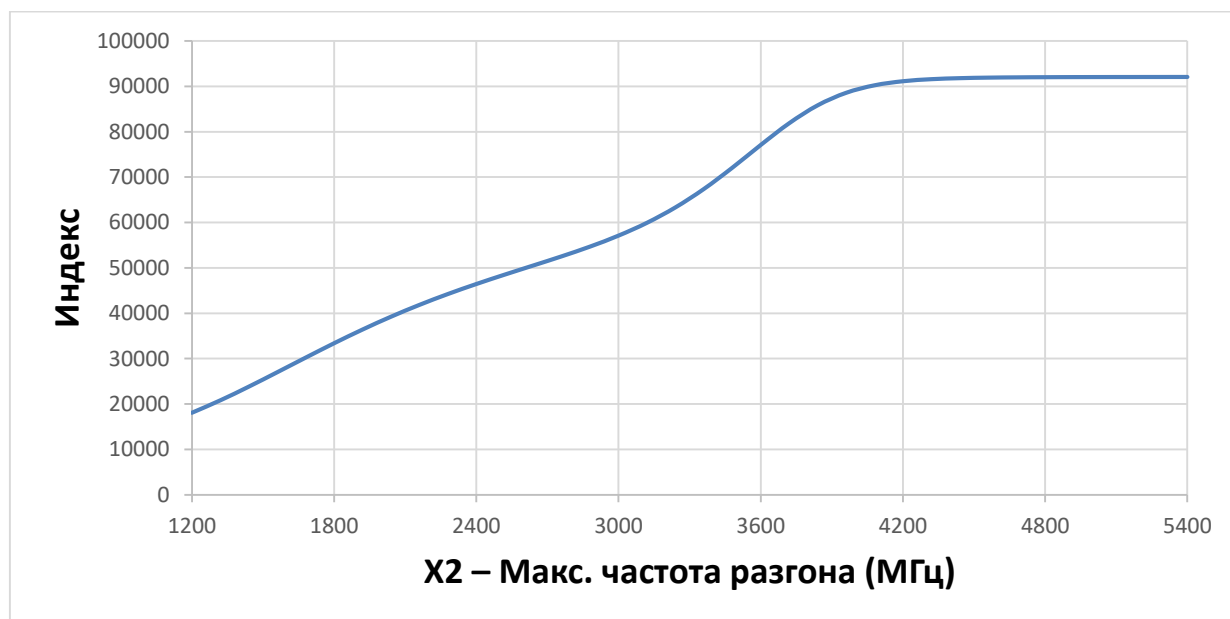


**Рисунок 7. Зависимость прогнозируемого индекса производительности от пропускной способности памяти на один контакт (X4)**

На рисунке 7 продемонстрирована зависимость прогнозируемого индекса производительности от пропускной способности памяти на один контакт. Так же как и в случае с максимальной частотой разгона, с ростом пропускной способности происходит плавный рост индекса производительности. При этом видно, что зависимость является практически линейной.

На рисунке 8 продемонстрирована одновременное влияние наиболее значимых параметров на прогнозируемый индекс производительности. Всего было проведено исследование влияния следующих параметров: количество шейдерных блоков, ширина шины, количество ядер трассировки лучей, максимальная частота “разгона” и пропускная способность памяти на один контакт. Каждый из параметров имеет шаг в 10% от начального значения (взятого из «усреднённого графического процессора»).

Из графика видно, что увеличение значений всех параметров ведёт к плавному росту индекса производительности до определённого значения (примерно 92000), после которого рост индекса практически полностью прекращается.



**Рисунок 8. Зависимость прогнозируемого индекса производительности от параметров с наибольшим весом (X2, X4, X5, X6, X9)**

Полученные результаты исследований не противоречат реальности, что подтверждает, что спроектированную нейронную сеть можно считать пригодной для прогнозирования индекса производительности графических процессоров.

**Заключение.** Построена система нейросетевого прогнозирования индекса производительности графических процессоров. Спроектированная нейросетевая модель учитывает 10 параметров: минимальная частота (МГц), максимальная частота, достижимая с помощью “разгона” (МГц), размер памяти (Мбайт), пропускная способность памяти на один контакт (Мбит/с), ширина шины (бит), количество универсальных шейдерных блоков, количество текстурных блоков (TLU), количество блоков растеризации (ROP), количество ядер для выполнения трассировки лучей (RT-ядер), количество тензорных ядер. Методом сценарного прогнозирования построены графики зависимостей прогнозируемого индекса производительности от изменения входных параметров. Применение такого набора параметров в модели позволяет с высокой точностью прогнозировать индексы производительности гра-



фических процессоров. Данный набор параметров может быть изменен для оценки индексов производительности любых графических процессоров, у которых доступна информация о технических характеристиках.

### Библиографический список

1. Chaunikam.info – Быстродействие и спецификации видеокарт. [Электронный ресурс]. Режим доступа: [https://www.chaunikam.info/gpu\\_specif.html](https://www.chaunikam.info/gpu_specif.html)
2. Techpowerup.com – GPU Specs Database. [Электронный ресурс]. Режим доступа: <https://www.techpowerup.com/gpu-specs/>
3. Черепанов Ф.М., Ясницкий Л.Н. Нейросимулятор 4.0. Свидетельство о регистрации программы для ЭВМ RUS 2014612546. Заявка № 2014610341 от 15.01.2014.
4. Ясницкий Л.Н. Введение в искусственный интеллект. М.: Издательский центр «Академия», 2005. – 176 с.

## NEURAL NETWORK SYSTEM FOR PREDICTING BENCHMARK INDEX FOR GRAPHICAL PROCESSING UNITS

*Shilov Ilya A.*

Perm State University

Str. Bukireva, 15, Perm, Russia, 614990, [ishilov56@gmail.com](mailto:ishilov56@gmail.com)

The article describes the development of a neural network system for predicting benchmark index for graphics processing units used in computers. The system allows to accurately predict benchmark index of graphical processing unit based on the data of processing unit's specifications. With the help of the developed intellectual system, a study of the subject area was carried out, patterns of practical importance were identified.

**Keywords:** artificial intelligence, neural network, prediction, graphics processing unit.

# **ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ В ПЕДАГОГИКЕ**

## НЕЙРОСЕТЕВАЯ СИСТЕМА ПРОГНОЗИРОВАНИЯ ВЕРОЯТНОСТИ ПОСТУПЛЕНИЯ В МАГИСТРАТУРУ

*Кобыхно Марина Евгеньевна*

Пермский государственный национальный исследовательский университет,  
ПМИ. 614990, Россия, г. Пермь, ул. Букирева, 15,  
kobyhno.marina@mail.ru

В статье представлено описание разработки нейросетевой системы для прогнозирования вероятности поступления в магистратуру. Система позволяет с большой точностью предсказать оценку вероятности поступления в магистратуру в зарубежный ВУЗ на основе данных о человеке. С помощью разработанной интеллектуальной системы проведено исследование предметной области, выявлены закономерности, имеющие практическое значение.

**Ключевые слова:** искусственный интеллект, нейросетевые технологии, прогнозирование, поступление, магистратура.

**Введение.** Программы бакалавриата по своей сути призваны дать базовое образование в определённой профессии и определённом предмете. Получив диплом бакалавра, человек уже разбирается в основах предмета и профессии и может начинать работать в качестве квалифицированного специалиста. Однако, этого достаточно только для старта карьеры и начала работы в специальности. Для продвижения по карьерной лестнице необходимы более глубокие знания как в основной сфере деятельности, так и в смежных областях. Знание вероятности поступления в магистратуру может помочь в улучшении качества образования или даже в смене профессиональной деятельности.

При анализе литературных источников выяснилось, что работ на тему прогнозирования вероятности поступления на программы магистратуры не было проведено.

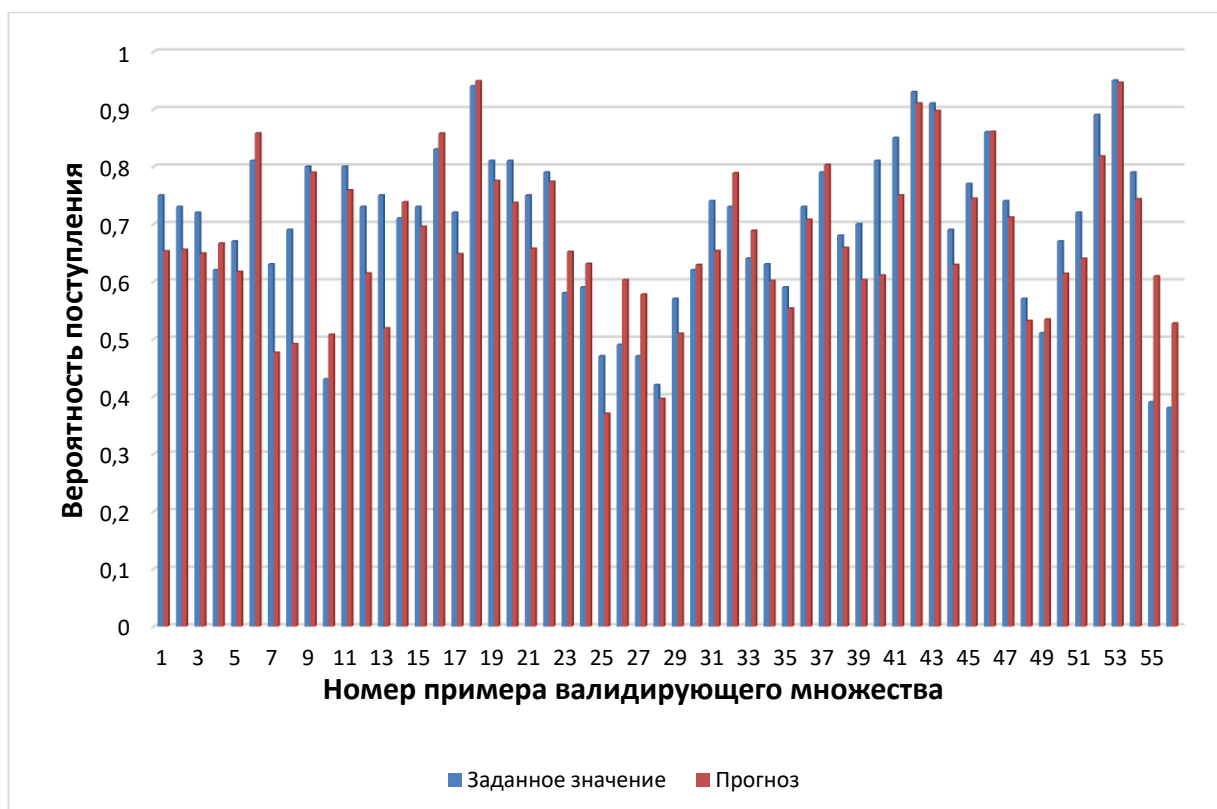
Основная цель настоящей работы заключается в сборе множества данных о людях, собирающихся подать заявку на магистерские программы в зарубежные ВУЗы, а также создание и обучение нейросетевой модели на этих данных. Конечным результатом должна быть нейросетевая система, способная прогнозировать вероятность успешного поступления на магистерские программы зарубежных ВУЗов.

Для создания нейросетевой системы были выбраны следующие параметры: X1 – баллы экзамена GRE (Graduate Record Examinations), X2 – баллы экзамена TOEFL (Test of English as a Foreign Language), X3 – рейтинг университета, X4 – оценка изложения цели поступления, X5 – оценка рекомендательного письма, X6 – средний балл бакалавриата, X7 – опыт исследовательской деятельности. Выходной параметр – вероятность поступления в магистратуру.

Обучающее множество было собрано с помощью программы-парсера с интернет-ресурсов. Перед переходом к проектированию и обучению нейросетевой модели, была выполнена очистка исходного множества от избыточности и дубликатов. Например, данные были подобраны так, чтобы было задействовано как можно больше различных средних баллов бакалавриата (от 6 до 10). Таким образом, объем итогового множества включает в себя 400 примеров. Данное множество было разделено на обучающее, валидирующее и тестирующее в соотношении 80%, 14% и 6% соответственно. Большая часть данных была собрана с интернет-ресурса [1].

Проектирование, обучение, тестирование нейронной сети выполнялись с помощью программы «Нейросимулятор 5» [2] по методике [3, 4]. В результате оптимизации спроектированная нейронная сеть представляла собой персептрон, который имеет семь входных нейронов, один выходной и один скрытый слой с пятью нейронами.

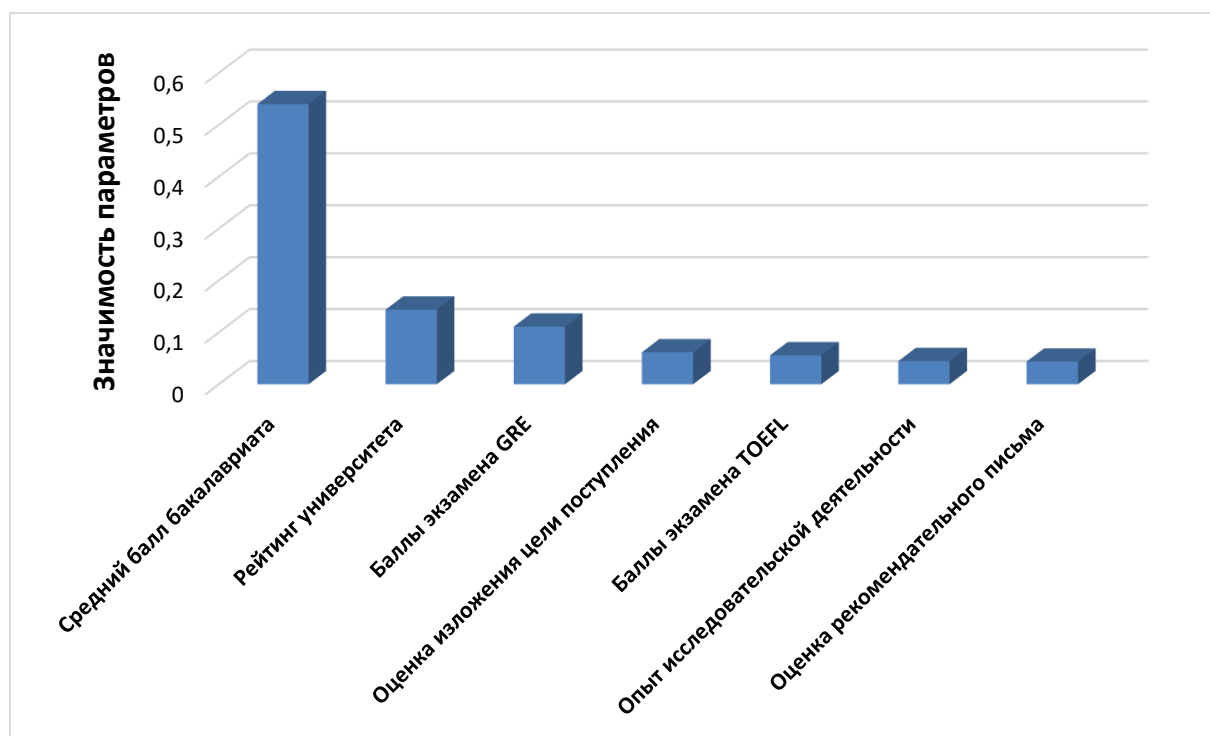
Для оценки корректной работы спроектированной нейронной сети использовалось валидирующее множество, состоящее из 56 примеров. Средняя относительная ошибка тестирования составила 9,5%, что можно считать приемлемым результатом. На рисунке 1 представлена гистограмма, демонстрирующая разницу между фактической и прогнозируемой нейронной сетью вероятностью.



**Рисунок 1. Результат тестирования нейронной сети**

Из результатов, изображенных на рисунке 1, можно сделать вывод об адекватной работе спроектированной нейронной сети.

Оценка значимости входных параметров была выполнена с помощью программы «Нейросимулятор 5», было найдено среднее значение значимости входных параметров обучающего и валидирующего множеств, результат отображен на рисунке 2.



**Рисунок 2. Значимость входных параметров нейросетевой модели**

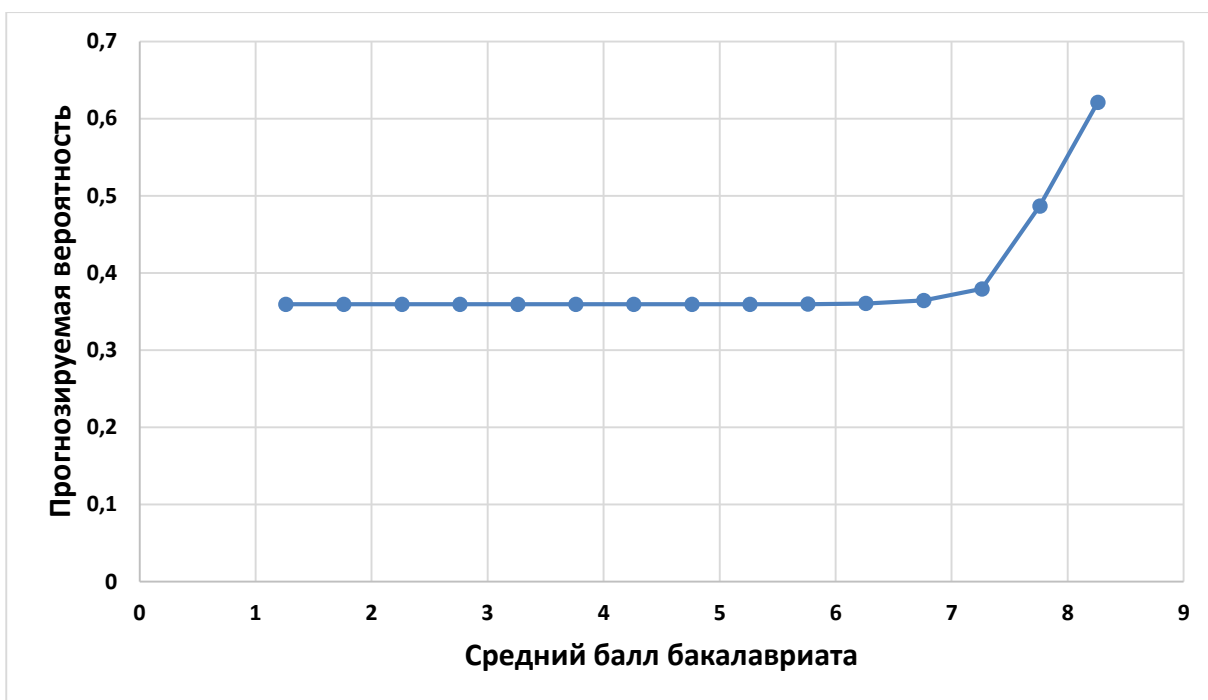
Как видно из рисунка 2, наиболее значимыми являются средний балл бакалавриата, рейтинг университета, баллы экзамена GRE и т.д. Наиболее влиятельным параметром является средний балл бакалавриата. Как и ожидалось оценка предыдущей деятельности играет очень важную роль в обучении на магистратуре.

Далее было проведено исследование полученных зависимостей между входными параметрами и вероятностью поступления. Исследование проводилось с помощью метода «замораживания», суть которого заключается в варьировании значения одного параметра и фиксировании значений всех других параметров. Для этой цели был отобран «нейтральный пример», про который нейросеть не может с уверенностью сказать, какая вероятность поступления в магистратуру у данного человека.

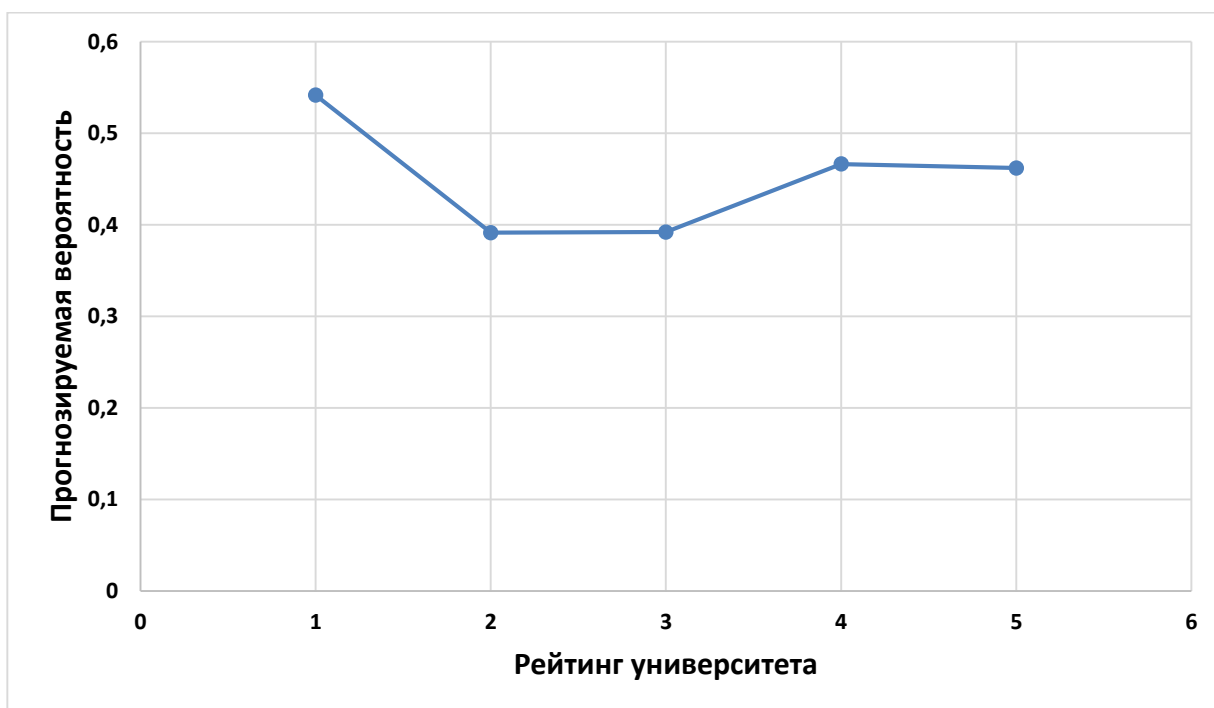
На рисунке 3 показан график зависимости прогнозируемой вероятности от среднего балла бакалавриата. В том случае, при более высоком среднем балле, нейросеть прогнозирует повышение вероятности поступления в магистратуру.

На рисунке 4 показан график зависимости прогнозируемой вероятности от рейтинга университета. Можно сделать вывод, что вероятность поступления в университет с более низким рейтингом выше, чем в университет с высоким рейтингом. Как видно из графика, вероятность поступления в университеты с рейтингом 2 и 3 ниже, чем в университеты с рейтингом 4 и 5.

Это связано с тем, что большее количество людей подает заявки в университеты со средним рейтингом, следовательно, возникает большая конкуренция.

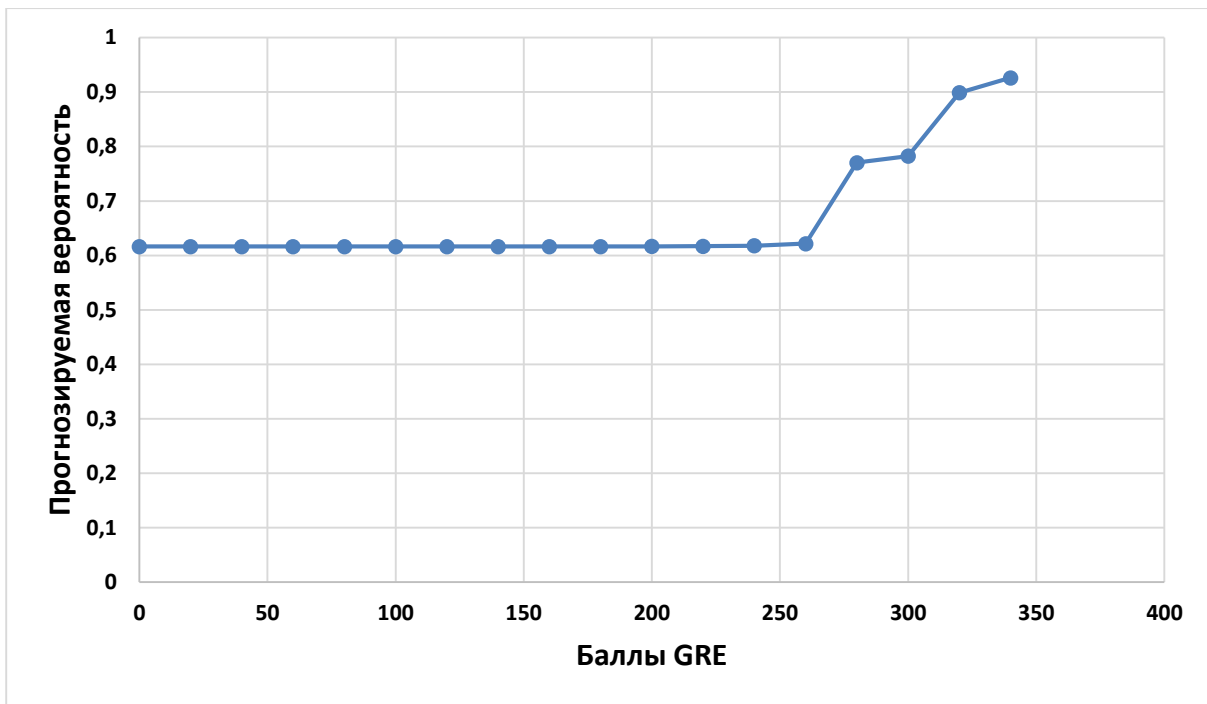


**Рисунок 3. Зависимость прогнозируемой вероятности поступления от среднего балла бакалавриата**



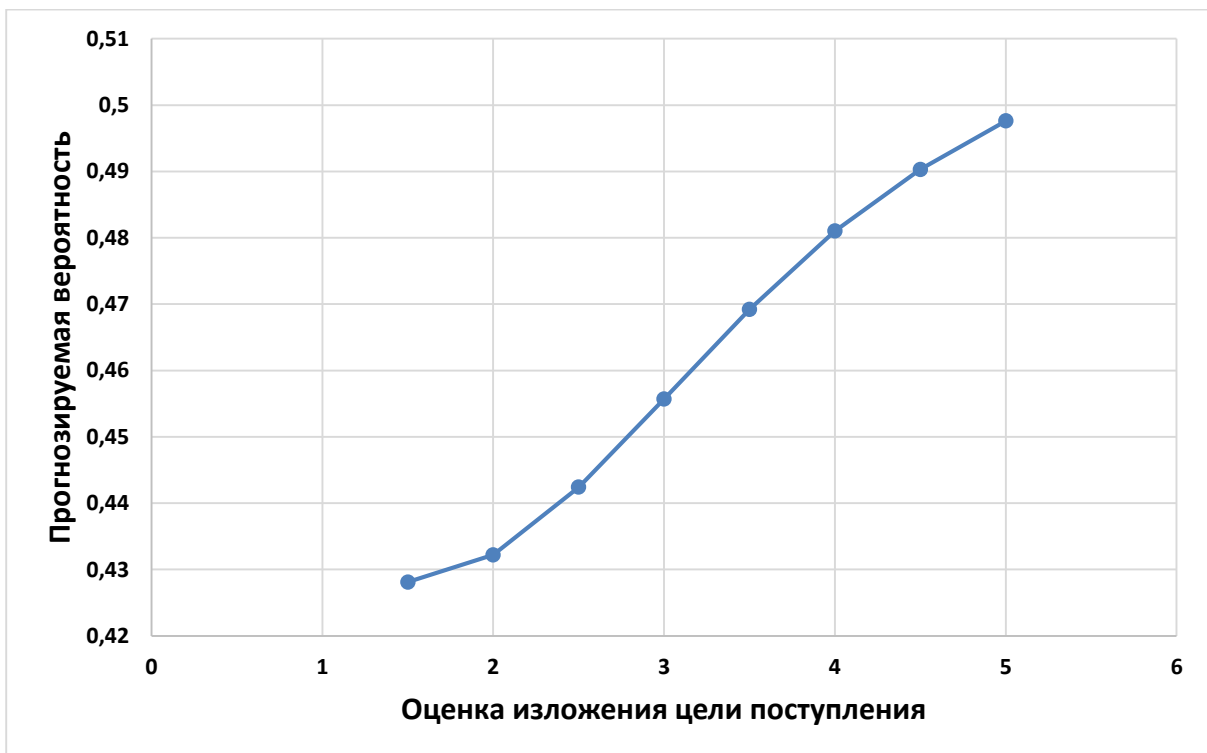
**Рисунок 4. Зависимость прогнозируемой вероятности поступления от рейтинга университета**

На рисунке 5 показан график зависимости прогнозируемой вероятности от баллов GRE [5]. Можно заметить, что при увеличении баллов GRE резко увеличивается вероятность поступления.



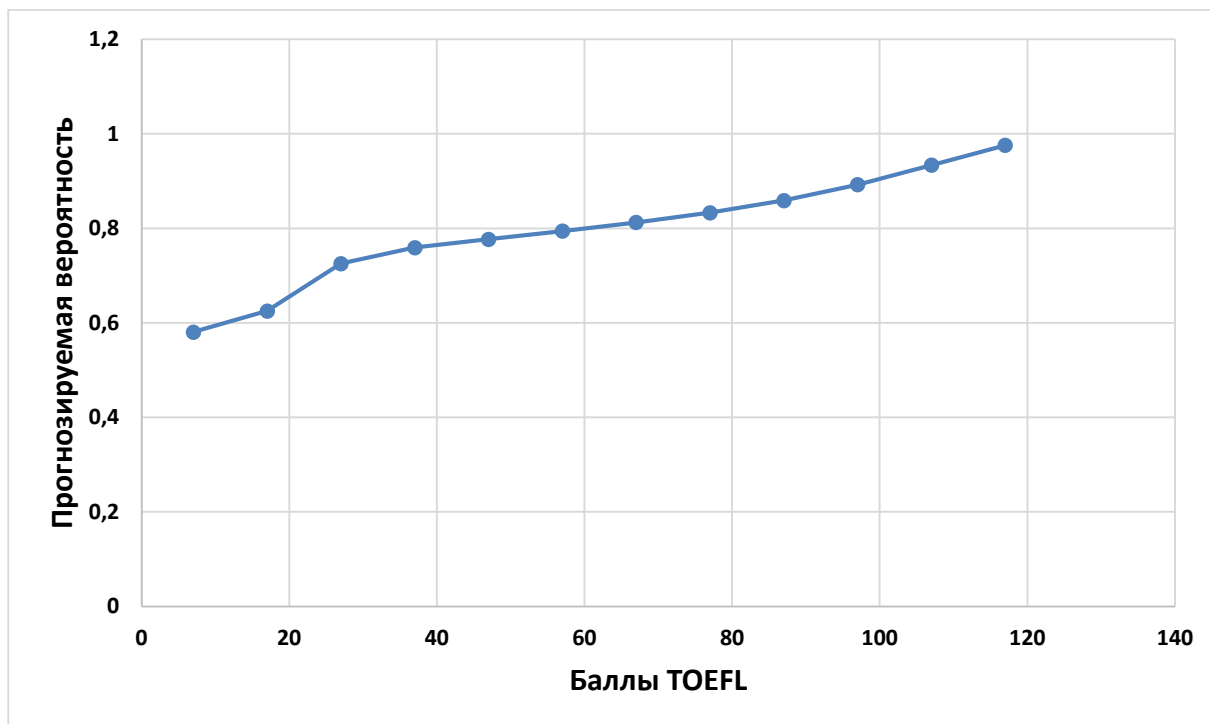
**Рисунок 5. Зависимость прогнозируемой вероятности поступления от баллов GRE**

На рисунке 6 продемонстрирована зависимость прогнозируемой вероятности поступления от оценки изложения цели поступления. Как видно из графика, оценка изложения цели оказывает сильное влияние на вероятность поступления, чем выше оценка, тем больше шансов поступить.



**Рисунок 6. Зависимость прогнозируемой вероятности поступления от оценки изложения цели поступления**

На рисунке 7 представлена зависимость вероятности поступления на магистерские программы от результатов экзамена TOEFL [6]. Как можно заметить из графика, при уменьшении количества баллов экзамена TOEFL уменьшается вероятность поступления.



**Рисунок 7. Зависимость прогнозируемой вероятности поступления от баллов TOEFL**

Выше были продемонстрированы графики, когда критерий оценки поступающего оказывал видимое влияние на поступление, в остальных случаях особого изменения вероятности поступления не обнаружено.

Полученные результаты исследований не противоречат реальности, что подтверждает, что спроектированную нейронную сеть можно считать пригодной для прогнозирования вероятности поступления в магистратуру в зарубежные ВУЗы.

**Заключение.** Построена система нейросетевого прогнозирования вероятности поступления в магистратуру в зарубежные ВУЗы. Спроектированная нейросетевая модель учитывает 7 параметров: баллы экзамена GRE, баллы экзамена TOEFL, рейтинг университета, оценку изложения цели поступления, оценку рекомендательного письма, средний балл бакалавриата, опыт исследовательской деятельности. Методом сценарного прогнозирования построены графики зависимости прогнозируемой вероятности поступления от изменения входных параметров. Применение такого набора параметров в модели позволяет с высокой точностью прогнозировать вероятность поступления в магистратуру в зарубежные ВУЗы.



## Библиографический список

1. Исходные данные. Набор данных о поступлении в магистратуру. [Электронный ресурс]. Режим доступа: <https://www.kaggle.com/datasets/mohansacharya/graduate-admissions>
2. Черепанов Ф.М., Ясницкий Л.Н. Нейросимулятор 4.0. Свидетельство о регистрации программы для ЭВМ RUS 2014612546. Заявка № 2014610341 от 15.01.2014.
3. Ясницкий Л.Н. Введение в искусственный интеллект. М.: Издательский центр «Академия», 2005. – 176 с.
4. Яснийкий Л.Н. Интеллектуальные системы. -М.: Лаборатория знаний, 2016. – 221с.
5. Экзамен GRE в магистратуру: структура и самоподготовка. [Электронный ресурс]. Режим доступа: [https://www.unipage.net/ru/exam\\_gre](https://www.unipage.net/ru/exam_gre)
6. Экзамены IELTS и TOEFL: зачем они нужны и как их сдают? [Электронный ресурс]. Режим доступа: <https://habr.com/ru/company/puzzleenglish/blog/408837/>

## NEURAL NETWORK SYSTEM FOR PREDICTING THE PROBABILITY OF ADMISSION TO THE MASTER'S PROGRAM

*Kobykhno Marina E.*

Perm State University

Str. Bukireva, 15, Perm, Russia, 614990, [kobyhno.marina@mail.ru](mailto:kobyhno.marina@mail.ru)

The article describes the development of a neural network system for predicting the probability of admission to a master's degree. The system allows you to predict with great accuracy the assessment of the probability of admission to a master's degree based on data about a person. With the help of the developed intellectual system, a study of the subject area was carried out, patterns of practical importance were identified.

**Keywords:** artificial intelligence, neural network technologies, forecasting, admission, master's degree.

## НЕЙРОСЕТЕВАЯ СИСТЕМА ПРОГНОЗИРОВАНИЯ ПОСТУПЛЕНИЯ АБИТУРИЕНТА В ВУЗ «УРАЛЬСКИЙ ФЕДЕРАЛЬНЫЙ УНИВЕРСИТЕТ»

*Рычков Андрей Владимирович*

Пермский государственный национальный исследовательский университет,  
КМБ. 614990, Россия, г. Пермь, ул. Букирева, 15,  
rychkovandrey1024@gmail.com

В статье представлено описание разработки нейросетевой системы для прогнозирования поступления абитуриента в вуз «Уральский Федеральный Университет» на направление «Компьютерная безопасность» в г. Екатеринбург. Система позволяет выполнять прогнозирование поступления в вуз с учетом результата сдачи ЕГЭ и индивидуальных достижений.

**Ключевые слова:** искусственный интеллект, нейросетевые технологии, прогнозирование, поступление в вуз, абитуриент, аттестат, студент, результаты ЕГЭ, высшее образование.

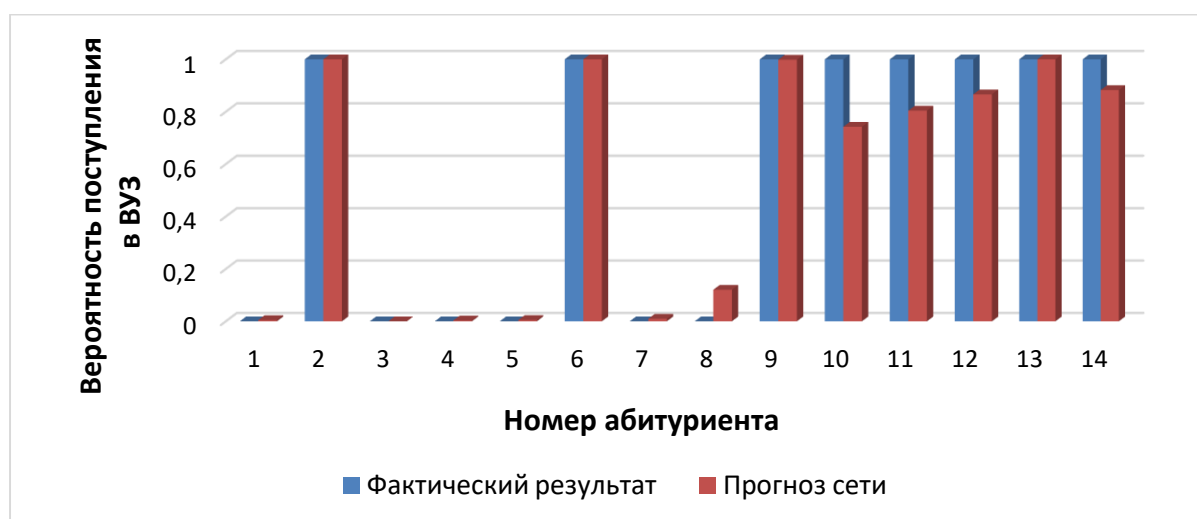
**Введение.** При поступлении абитуриента в вуз, очень часто возникает ситуация, когда требуется оценить вероятность поступления каждого из абитуриентов на выбранное им направление. Причем необходимость может возникнуть как со стороны приемной кампании вуза, так и со стороны самого абитуриента. В приемной кампании этим занимаются люди, которые могут по невнимательности допустить ошибку. Если же рассматривать сторону абитуриента, то стоит отметить, что проанализировать данные вручную будет достаточно проблематично, особенно если учитывать, что данные каждый год пополняются и в них несложно запутаться. Поэтому нейросеть для анализа вероятности поступления абитуриента в ВУЗ будет достаточно полезна как для абитуриентов, так и для высших учебных заведений.

При анализе литературных источников и списка абитуриентов за 2021 год [1-5], был выбран следующий набор показателей для нейросетевой системы: X1 – баллы ЕГЭ по русскому языку, X2 – баллы ЕГЭ по математике, X3 – баллы ЕГЭ по информатике, X4 – баллы за индивидуальные достижения, X5 – наличие оригинала аттестата. В качестве выходного параметра D1 получается значение – вероятность поступления абитуриента в ВУЗ.

Обучающее множество было собрано вручную с сайта Уральского Федерального Университета [3]. Объем итогового множества включает в себя 126 примеров. Данное множество было разделено на обучающее и тестирующее в соотношении 89% и 11% соответственно. Собранные данные охватывают период с июня по август 2021 года.

Проектирование, обучение, тестирование нейронной сети выполнялись с помощью программы «Нейросимулятор 5» [1] по методике [3]. В результате оптимизации спроектированная нейронная сеть представляла собой персептрон, который имеет пять входных нейронов, один выходной и один скрытый слой с двумя нейронами.

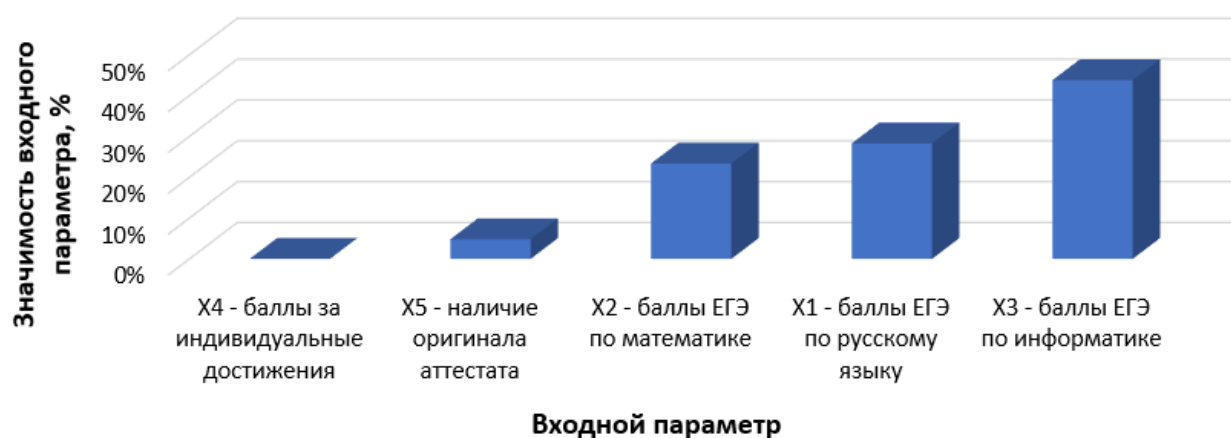
Для оценки корректной работы спроектированной нейронной сети использовалось тестирующее множество, состоящее из 14 примеров. Средняя относительная ошибка тестирования составила 9.6%, что можно считать неплохим результатом. На рисунке 1 представлена гистограмма, демонстрирующая разницу между фактическим и прогнозируемой нейросетью вероятностью поступления 14 абитуриентов из тестирующего множества.



**Рисунок 1. Результат тестирования нейронной сети**

Из результатов, изображенных на рисунке 1, можно сделать вывод об адекватной работе спроектированной нейронной сети.

Оценка значимости входных параметров была выполнена с помощью программы «Нейросимулятор 5», результат отображен на рисунке 2.



**Рисунок 2. Значимость входных параметров нейросетевой модели**

Как видно из рисунка 2, наиболее значимым параметром являются баллы по информатике, далее идут баллы за ЕГЭ по русскому языку, мате-

матике, а затем наличие оригинала аттестата и баллы за индивидуальные достижения.

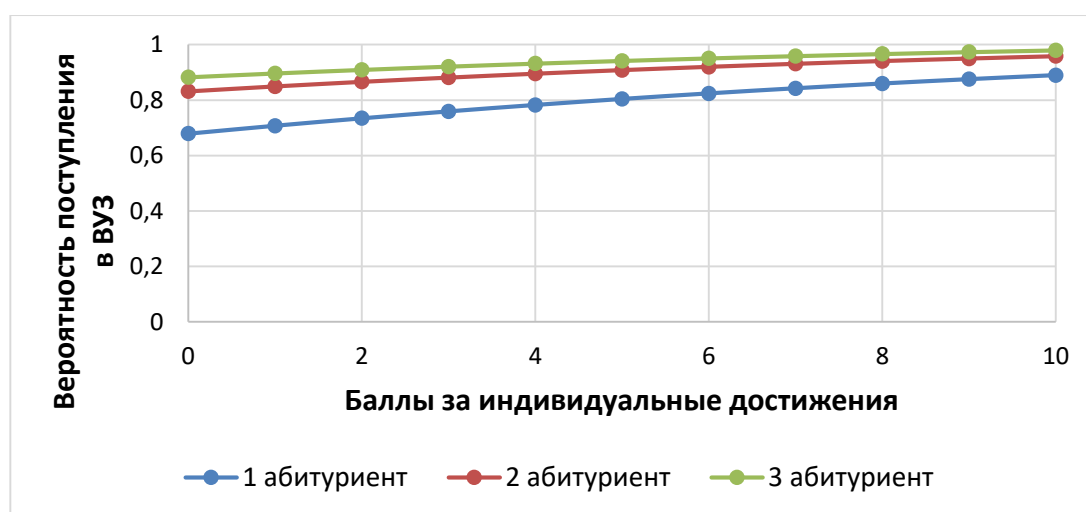
Далее было проведено исследование влияния наличия баллов за индивидуальные достижения. Исследование производилось с помощью метода «замораживания» [3], суть которого заключается в варьировании значения одного параметра и фиксировании значений всех других параметров. Для этой цели были отобраны 3 абитуриента, информация о параметрах которых показана в таблице 1.

**Таблица 1**

**Характеристики абитуриентов, выбранных для исследования**

№ абитуриента	Баллы за ЕГЭ по русскому языку	Баллы за ЕГЭ по математике	Баллы за ЕГЭ по информатике	Наличие оригинала аттестата
1	87	80	84	1
2	80	75	100	1
3	63	100	100	1

На рисунке 3 показан график зависимости вероятности поступления абитуриента от наличия индивидуальных достижений. Исходя из графика можно сказать, что наличие баллов за индивидуальные достижения увеличивают вероятность поступления.



**Рисунок 3. Зависимость вероятности поступления от наличия баллов за индивидуальные достижения**

Далее было проведено исследование влияния наличия оригинала аттестата на вероятность поступления абитуриента в ВУЗ. Для этого были отобраны 2 абитуриента, информация о параметрах которых показана в таблице 2.

На рисунке 4 продемонстрирована зависимость вероятности поступления абитуриента от наличия оригинала аттестата. Как следует из графика наличие оригинала аттестата увеличивает вероятность поступления.

Далее было проведено исследование влияния баллов ЕГЭ по каждому из предметов по отдельности. Для этого были отобраны 3 абитуриента, информация о параметрах которых показана в таблице 3.

Таблица 2

## Характеристики абитуриентов, выбранных для исследования

№ абитуриента	Баллы за ЕГЭ по русскому языку	Баллы за ЕГЭ по математике	Баллы за ЕГЭ по информатике	Баллы за индивидуальные достижения
1	80	100	75	2
2	63	100	100	0

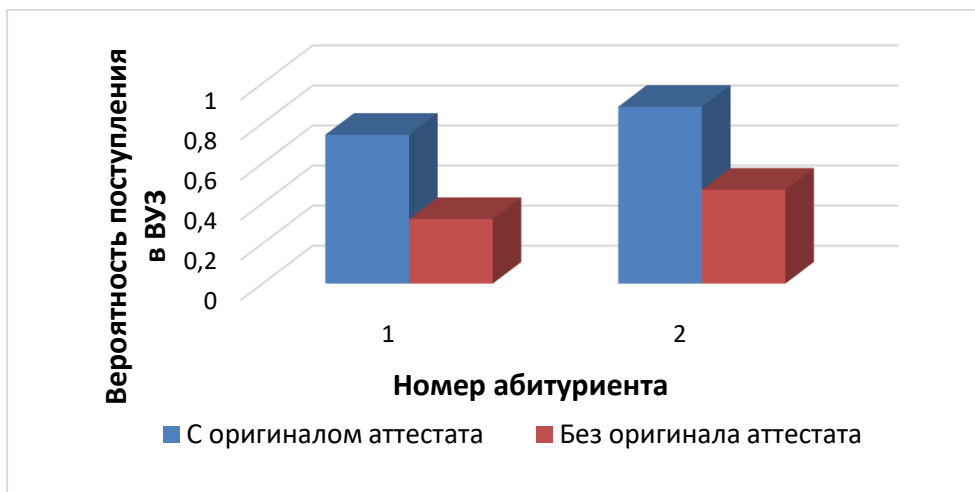


Рисунок 4. Зависимость вероятности поступления от наличия оригинала аттестата

Таблица 3

## Характеристики абитуриентов, выбранных для исследования

№ абитуриента	Баллы за ЕГЭ по русскому языку	Баллы за ЕГЭ по математике	Баллы за ЕГЭ по информатике	Баллы за индивидуальные достижения	Наличие оригинала аттестата
1	70	56	73	0	0
2	91	80	70	5	1
3	80	100	75	2	1

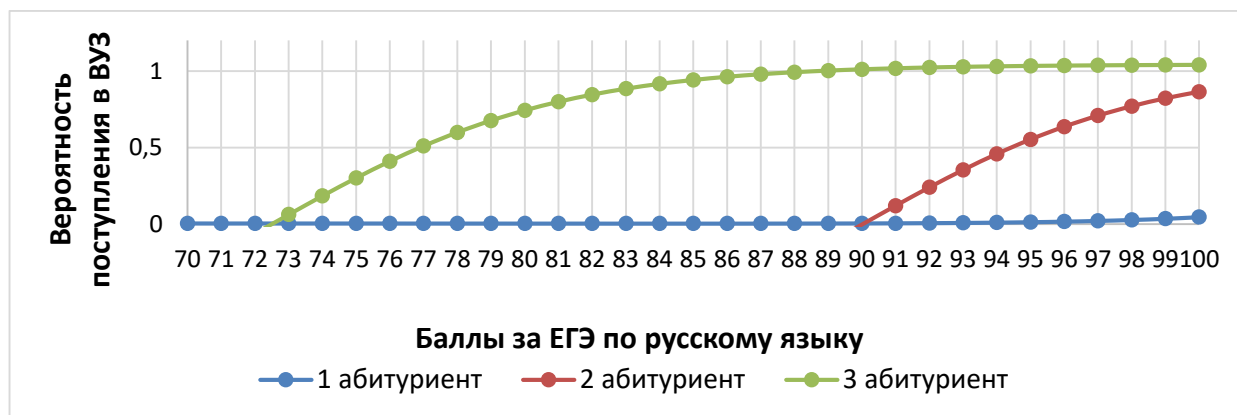


Рисунок 5. Зависимость вероятности поступления от баллов по русскому языку

На рисунках 5, 6 и 7 показаны графики зависимости вероятности поступления абитуриента от количества баллов ЕГЭ по различным предметам. Как следует из графика, увеличение количества баллов по соответствующим предметам увеличивает вероятность поступления.

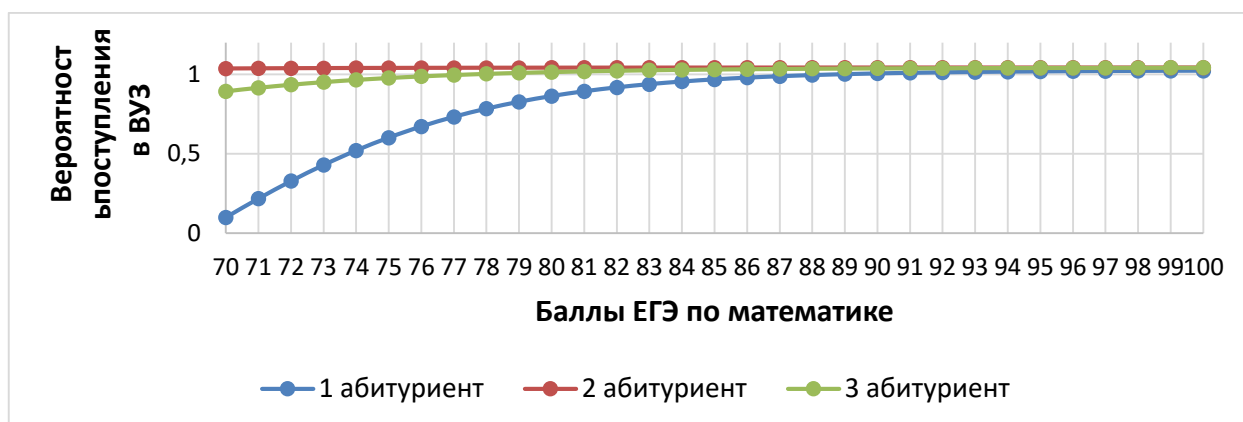


Рисунок 6. Зависимость вероятности поступления от баллов по математике

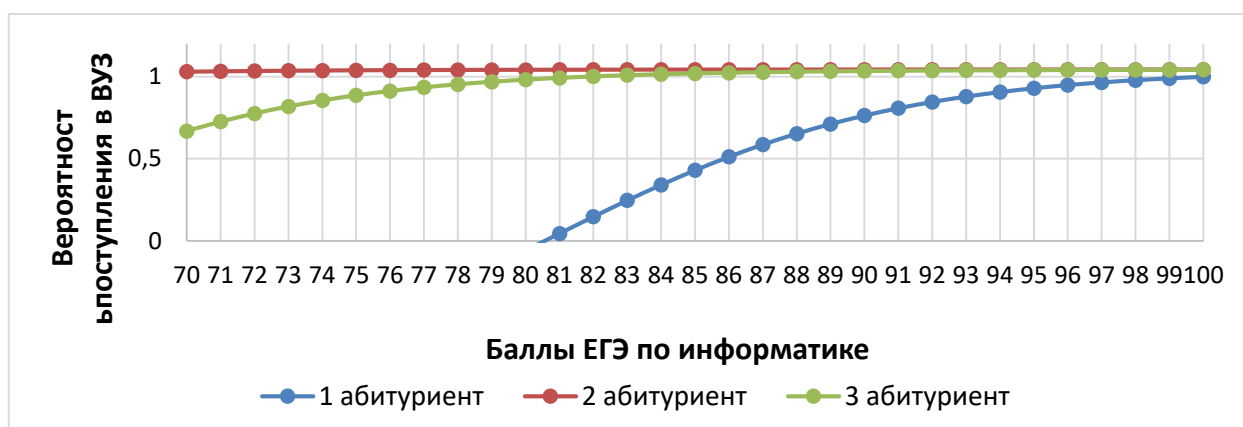


Рисунок 7. Зависимость вероятности поступления от баллов по информатике

Полученные результаты исследований не противоречат реальности, что подтверждает, что спроектированную нейронную сеть можно считать пригодной для прогнозирования вероятности поступления в Уральский Федеральный Университет на направление «Компьютерная безопасность».

**Заключение.** Построена система нейросетевого прогнозирования вероятности поступления абитуриента в Уральский Федеральный Университет на направление «компьютерная безопасность».

Спроектированная нейросетевая модель учитывает 5 параметров: баллы ЕГЭ по русскому языку, математике и информатике, баллы за индивидуальные достижения и наличие оригинала аттестата. Данный набор параметров может быть изменен для прогнозирования вероятности поступления абитуриента в другой ВУЗ или другое направление.

### Библиографический список

1. Черепанов Ф.М., Ясницкий Л.Н. Нейросимулятор 4.0. Свидетельство о регистрации программы для ЭВМ RUS 2014612546. Заявка № 2014610341 от 15.01.2014.

2. Ясницкий Л.Н. Интеллектуальные системы: учебник. М.: Лаборатория знаний, 2016. 221 с.
3. Urfu.ru – официальный сайт Уральского Федерального Университета. [Электронный ресурс]. Режим доступа: <https://urfu.ru/ru/alpha/full/>
4. 5fan.ru – система прогнозирования поступления абитуриента в вуз. [Электронный ресурс]. Режим доступа: <http://5fan.ru/wievjob.php?id=10753>
5. Система прогнозирования поступления абитуриента в ПГНИУ (2006 год). [Электронный ресурс]. Режим доступа: <http://permai.ru/files/projects/P09.pdf>

## NEURAL NETWORK SYSTEM FOR FORECASTING ADMISSION OF AN ENTRYENT TO THE URAL FEDERAL UNIVERSITY

*Rychkov Andrey V.*

Perm State University

Str. Bukireva, 15, Perm, Russia, 614990, [rychkovandrey1024@gmail.com](mailto:rychkovandrey1024@gmail.com)

The article describes the development of a neural network system for predicting the admission of an applicant to the Ural Federal University in the direction of "Computer security" in Yekaterinburg. The system makes it possible to forecast admission to a university, taking into account the result of passing the exam and individual achievements.

**Keywords:** artificial intelligence, neural network technologies, forecasting, admission to a university, applicant, certificate, student, USE results, higher education.

УДК 004.032.26

## НЕЙРОСЕТЕВАЯ СИСТЕМА ПРОГНОЗИРОВАНИЯ ЛУЧШЕГО УНИВЕРСИТЕТА

*Шилова Елена Андреевна*

Национальный исследовательский университет «Высшая школа экономики»,  
614070, Россия, г. Пермь, ул. Студенческая, 38, [eashilova\\_1@edu.hse.ru](mailto:eashilova_1@edu.hse.ru)

В статье представлено описание разработки нейросетевой системы для прогнозирования рейтинга университета в мире. Система позволяет с большой точностью предсказать рейтинг определенного университета на основании собранных данных о других университетах. С помощью разра-

ботанной интеллектуальной системы проведено исследование предметной области, выявлены закономерности, имеющие практической значение.

**Ключевые слова:** нейросетевые технологии, прогнозирование, рейтинг университета, качество образования, трудоустройство.

**Введение.** Рейтинг университетов – сложная, политическая и противоречивая практика. Существуют сотни различных национальных и международных систем ранжирования университетов, многие из которых противоречат друг другу. Этот набор данных содержит три глобальных рейтинга университетов из самых разных мест.

Мировой рейтинг университетов Times Higher Education широко известен как один из самых влиятельных и широко наблюдаемых университетских показателей. Центр мировых рейтингов университетов — менее известный список из Саудовской Аравии, основанный в 2012 году.

Как эти рейтинги соотносятся друг с другом?

Справедливы ли различные критические замечания в отношении этих рейтингов?

Рейтинг эффективности научных работ для университетов мира составлялся до 2012 года Тайваньским советом по оценке и аккредитации высшего образования [4]. Показатели были разработаны для измерения как долгосрочной, так и краткосрочной исследовательской деятельности исследовательских университетов.

В этом проекте использовалась библиометрия для анализа и ранжирования показателей 500 лучших университетов и 300 лучших университетов в шести областях. Исследование также предоставляет предметные рейтинги в области науки и техники. Он также входит в число 300 лучших университетов в десяти областях науки и техники. Рейтинг включал восемь показателей [7].

Это были: статьи, опубликованные за предыдущие 11 лет; цитирование этих статей, «текущие» статьи, текущие цитирования, среднее цитирование, количество «высокоцитируемых статей» и журнальных статей с высоким уровнем воздействия. Они представляли собой три критерия эффективности научных работ: продуктивность исследований, влияние исследований и превосходство исследований [5].

Основная цель настоящей работы заключается в сборе множества публичных данных об университетах, а также создание и обучение нейросетевой модели на этих данных. Конечным результатом должна быть нейросетевая система, способная прогнозировать рейтинг университета.

Для создания нейросетевой системы были выбраны следующие параметры:

- X1 – страна каждого университета (закодировано последовательно),
- X2 – ранг университета в каждой отдельной стране,
- X3 – уровень качества образования,
- X4 – трудоустройство выпускников,
- X5 – качество профессорско-преподавательского состава,
- X6 – публикации,



X7 – эффективность исследований,  
X8 – количество студентов в университете,  
X9 – ранг по патентам,  
X10 – общий балл, используемый для определения мирового рейтинга,  
X11 – год рейтинга.

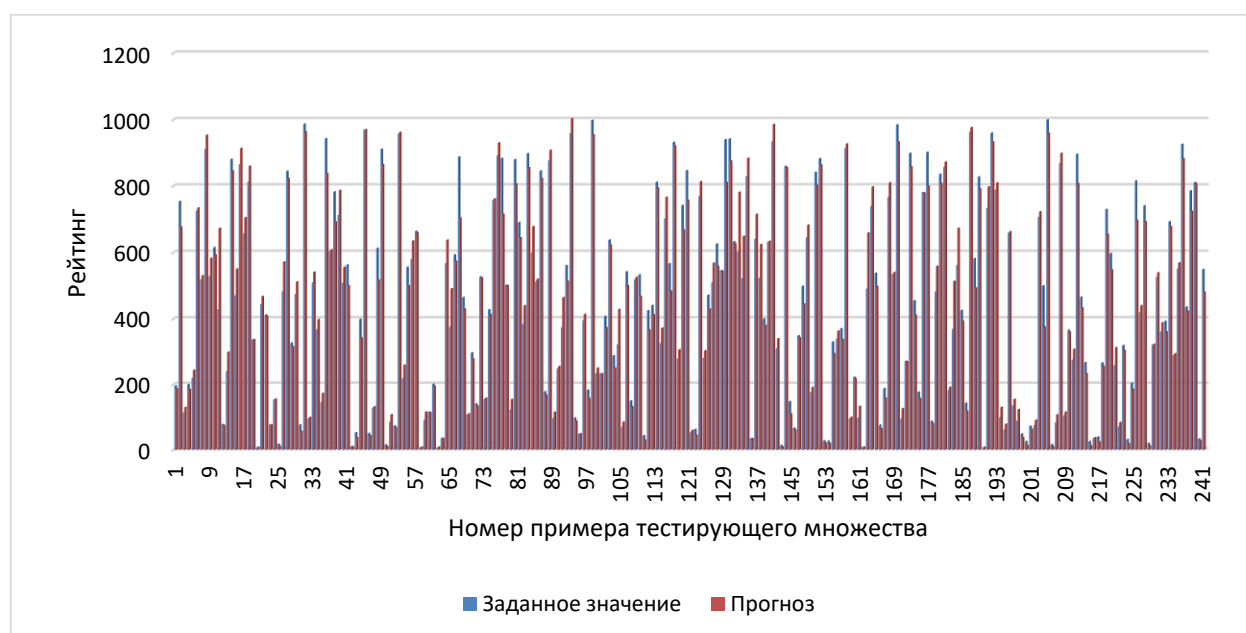
Выходной параметр:

D1 – мировой рейтинг для университета.

Обучающее множество было собрано с интернет-ресурсов Kaggle [3], а также вручную. Перед переходом к проектированию и обучению нейросетевой модели, была выполнена очистка исходного множества от противоречивых примеров, выбросов, дубликатов. Также были удалены лишние столбцы, которые не имели важного значения для проекта, и закодированы страны, университеты которых используются в исследовании. Таким образом, объем итогового множества включает в себя 1200 примеров. Данное множество было разделено на обучающее, валидирующее и тестирующее в соотношении 70%, 30% и 10% соответственно. Большая часть данных была собрана с интернет ресурсов [6].

Проектирование, обучение, тестирование нейронной сети выполнялись с помощью программы «Нейросимулятор 5» [1] по методике [2]. В результате оптимизации спроектированная нейронная сеть представляла собой персептрон, который имеет десять входных нейронов, один выходной и один скрытый слой с тремя нейронами.

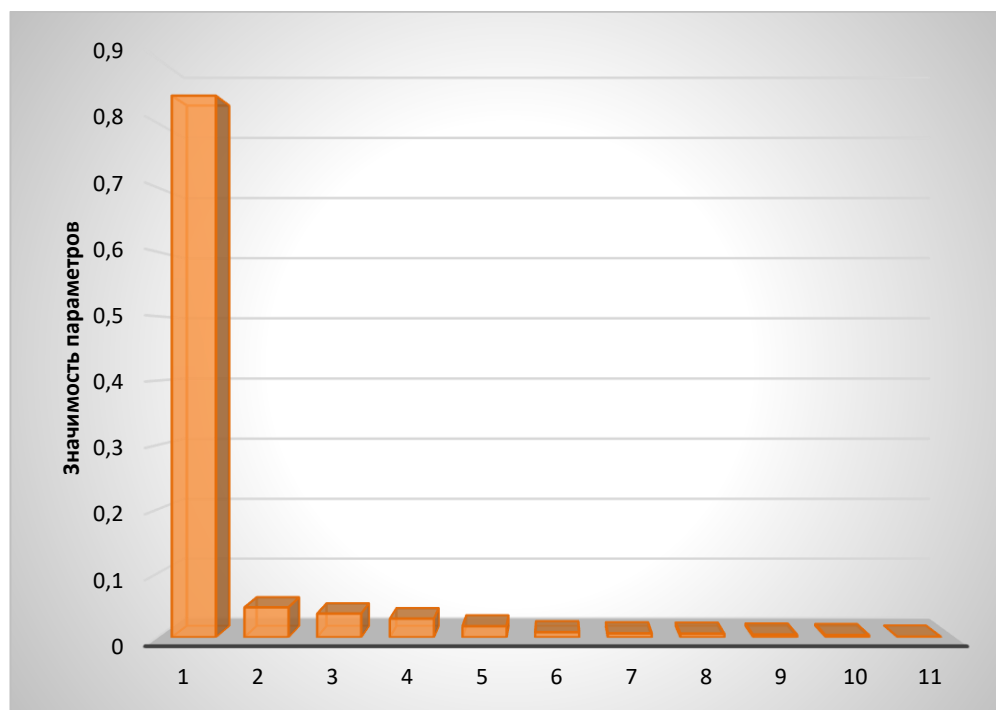
Для оценки корректной работы спроектированной нейронной сети использовалось тестирующее множество, состоящее из 1200 примеров. Средняя относительная ошибка тестирования составила 1.06%, что можно считать приемлемым результатом. На рисунке 1 представлена гистограмма, демонстрирующая разницу между фактическим и прогнозируемым нейросетью результатом рейтингов лучших университетов мира.



**Рисунок 1. Результат тестирования нейронной сети**

Из результатов, изображенных на рисунке 1, можно сделать вывод об адекватной работе спроектированной нейронной сети.

Оценка значимости входных параметров была выполнена с помощью программы «Нейросимулятор 5», результат отображен на рисунке 2.



**Рисунок 2. Значимость входных параметров нейросетевой модели**

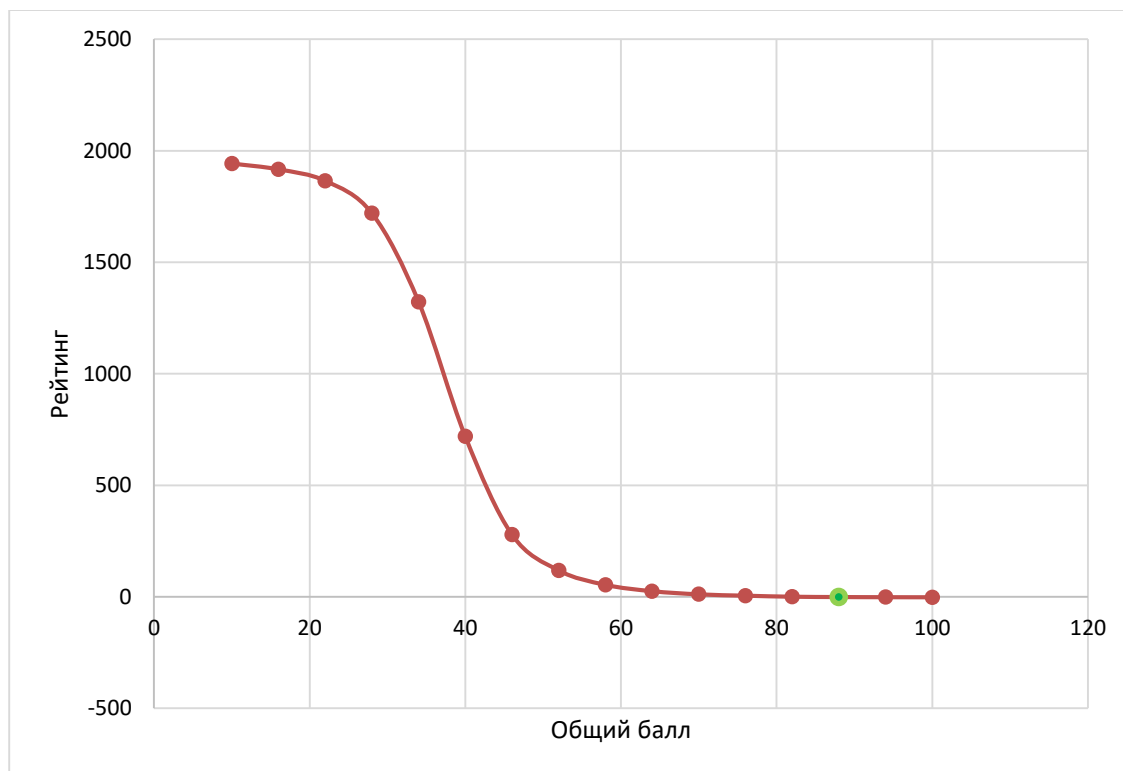
Как видно из рисунка 2, наиболее значимыми являются: общий балл, используемый для определения мирового рейтинга; публикации; трудоустройство выпускников; качество профессорско-преподавательского состава; влияние; год рейтинга; ранг по патентам; количество студентов в университете; уровень качества образования; страна каждого университета (закодировано последовательно); ранг университета в каждой отдельной стране.

На рисунке 3 показан график зависимости рейтинга университета от его общего балла. За основу был взят Принстонский университет, общий балл которого составляет 88,56. В том случае, когда общий балл университета стремится к отметке в 100, нейросеть прогнозирует значительное увеличение шанса стать выше в рейтинге.

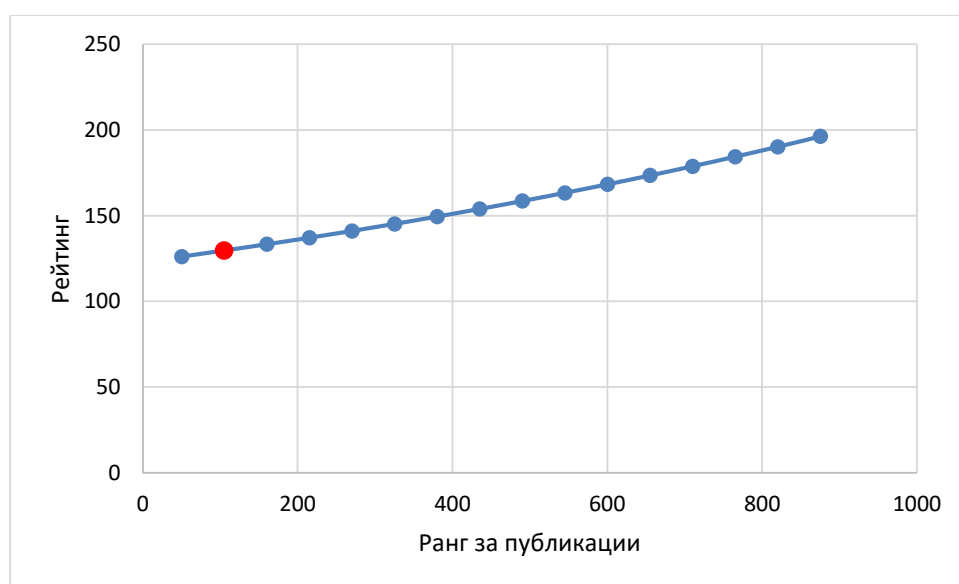
На рисунке 4 продемонстрирована зависимость рейтинга университета от публикаций. Можно заметить, значительную зависимость между ними. Чем выше качество публикаций, измеряемых количеством научных статей, опубликованных в ведущих журналах, тем выше рейтинг университета. Настоящий ранг университета равен 70. При увеличении ранга университет имеет возможность подняться в рейтинге.

На рисунке 5 изображен график зависимости рейтинга от трудоустройства выпускников. Как видно из графика, качество и частота трудоустройства

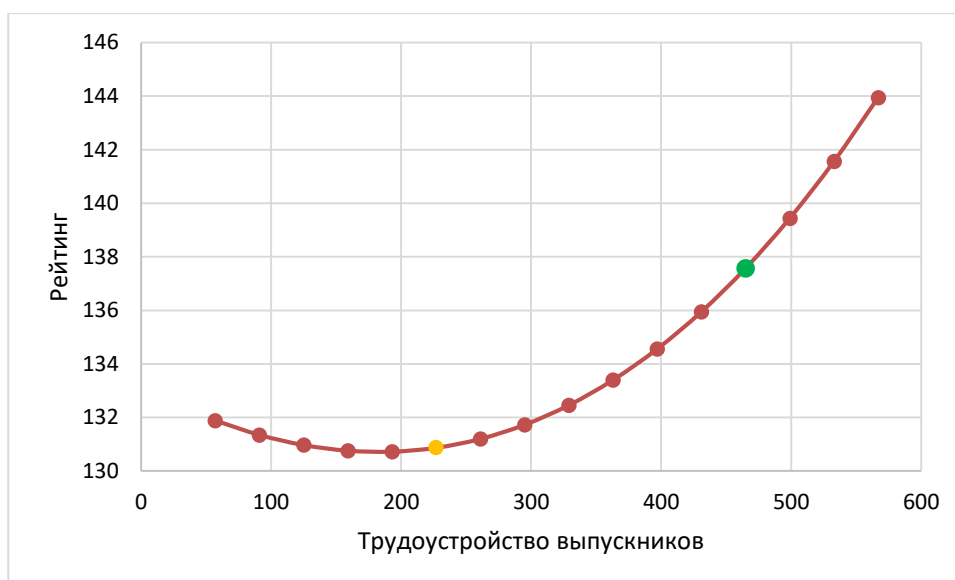
хорошо влияет на рейтинг университета в целом. Фактическое трудоустройство выпускников университета равно 478. Занятость выпускников, измеряемая количеством выпускников университета, которые занимали высшие руководящие должности в крупнейших мировых компаниях по отношению к размеру университета.



**Рисунок 3. Зависимость рейтинга университета от общего балла**

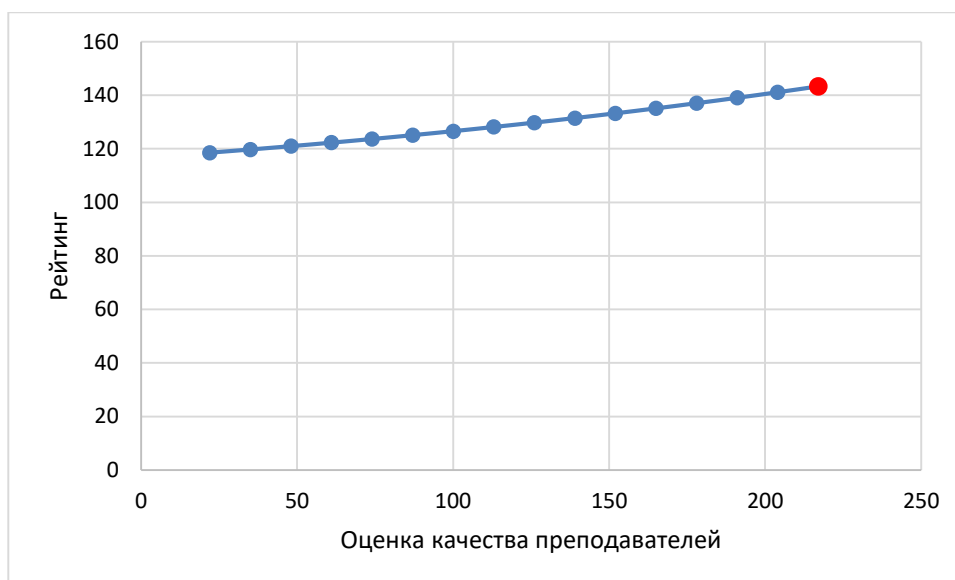


**Рисунок 4. Зависимость рейтинга университета от ранга публикаций**



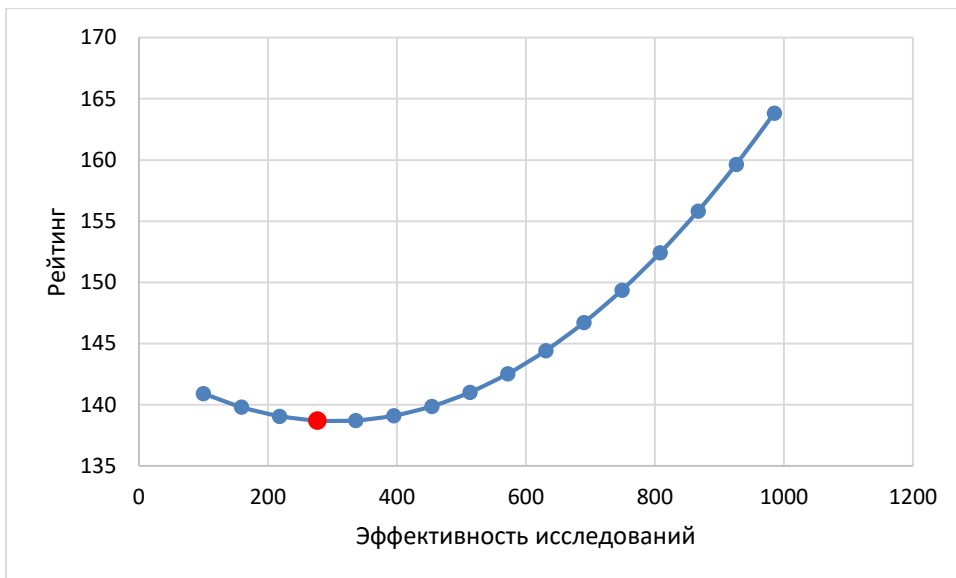
**Рисунок 5. Зависимость рейтинга от трудоустройства выпускников**

На рисунке 6 продемонстрирована зависимость рейтинга от преподавательского состава. Как видно из графика, с увеличением рейтинга преподавательского состава увеличивается и рейтинг самого университета. Качество профессорско-преподавательского состава измеряется количеством преподавателей, получивших крупные академические награды. Фактическая оценка качества преподавателей равна 210. При увеличении оценки будет увеличиваться рейтинг.



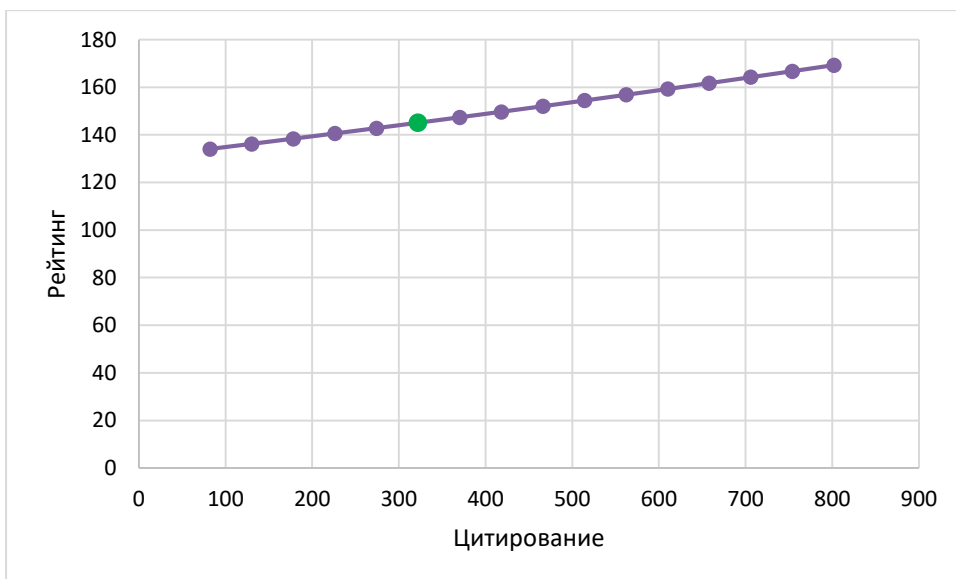
**Рисунок 6. Зависимость рейтинга от качества профессорско-преподавательского состава**

На рисунке 7 продемонстрирована зависимость рейтинга от уровня влияния университета. Влияние измеряется количеством научных статей, опубликованных во влиятельных журналах. Как видно из графика, фактическая эффективность исследований в университете равна 245.



**Рисунок 7. Зависимость рейтинга от эффективности исследований университета**

На рисунке 8 продемонстрирована зависимость рейтинга университета от цитирования. Цитирование измеряется количеством высокоцитируемых научных статей. Фактическое цитирование университета находится на уровне 310. С увеличением уровня цитирования публикаций университета увеличивается его рейтинг.



**Рисунок 8. Зависимость рейтинга от индекса цитирования**

Полученные результаты исследований не противоречат реальности, что подтверждает, что спроектированную нейронную сеть можно считать пригодной для прогнозирования рейтинга лучших университетов мира.

**Работа в Keras.** Вторая часть работы заключается в проведении аналогичного исследования на другом фреймворке. Процессы проектирования, обучения нейросети и тестирования реализуются на языке Python при помощи библиотек для работы с нейронными сетями Keras.

Перед началом работы с нейросетью ее надо обучить. Это производится на заранее подготовленном множестве со входными и выходными данными о показателях различных университетов. Результат обучения модели в Keras оказался хуже, чем в Нейростимуляторе.

При сравнении прогнозного значения и фактического значения результатов выбросов график показывал погрешность до 300 единиц.

Далее выполняю оценку качества модели:

№	Показатель	Результат
1	$R^2$	0,83
2	Max error	374, 77
3	Mean absolute error	68,45
4	Median absolute error	8,34
5	MSE	17015,29
6	RMSE	130,44

После проверки качества модели выполняю повторное построение графиков зависимости между входными параметрами с целью их дальнейшего сравнения. При проверке зависимости рейтинга университета от его общего балла получилась диаграмма аналогичная той, что была в Нейростимуляторе. В том случае, когда общий балл университета стремится к отметке в 100, нейросеть прогнозирует значительное увеличение шанса стать выше в рейтинге.

При проверке зависимости рейтинга университета от ранга публикаций получилась диаграмма, противоположная той, что получилась в Нейростимуляторе. Можно заметить, значительную зависимость между ними входными данными. Чем выше качество публикаций, измеряемых количеством научных статей, опубликованных в ведущих журналах, тем выше рейтинг университета.

При сравнении двух программ «Нейростимулятор 5» и «Keras» получены результаты, которые в некоторой степени отличаются друг от друга. Результаты, полученные в Keras менее точны, а усилий, приложенных для работы в этой программе, потребовалось значительно больше, так как возникали проблемы с файлами, ошибки в коде и др. Нейростимулятор значительно более удобен для использования и работы, а результаты более наглядны.

**Заключение.** Построена система нейросетевого прогнозирования лучшего университета. Спроектированная нейросетевая модель учитывает 11 параметров: страна каждого университета (закодировано последовательно), ранг университета в каждой отдельной стране, уровень качества образования, трудоустройство выпускников, качество профессорско-преподавательского состава, публикации, эффективность исследований, количество студентов в университете, ранг по патентам, общий балл, используемый для определения мирового рейтинга, год рейтинга. Методом сценарного прогнозирования построены графики зависимостей рейтинга от изменения входных параметров. Применение такого набора параметров в модели позволяет с вы-

сокой точностью прогнозировать рейтинг. Данный набор параметров может быть изменен для иных прогнозирований.

### Библиографический список

1. Черепанов Ф.М., Ясницкий Л.Н. Нейросимулятор 4.0. Свидетельство о регистрации программы для ЭВМ RUS 2014612546. Заявка № 2014610341 от 15.01.2014.
2. Ясницкий Л.Н. Введение в искусственный интеллект. М.: Издательский центр «Академия», 2005. – 176 с.
3. World University Rankings, Kaggle.com, [Электронный ресурс]. Режим доступа: <https://www.kaggle.com/datasets/mylesoneill/world-university-rankings>
4. College and university rankings, [Электронный ресурс]. Режим доступа: [https://en.wikipedia.org/wiki/College\\_and\\_university\\_rankings#Center\\_for\\_World\\_University\\_Rankings](https://en.wikipedia.org/wiki/College_and_university_rankings#Center_for_World_University_Rankings)
5. Higher Education Evaluation and Accreditation Council of Taiwan, Performance Ranking of Scientific Papers for World Universities. [Электронный ресурс]. Режим доступа: <https://archive.ph/20120720172842/http://ranking.heeact.edu.tw/en-us/2010%20by%20Field/Page/Methodology#selection-4191.17-4191.80>
6. Methodology. [электронный ресурс]. Режим доступа: <https://cwur.org/methodology/world-university-rankings.php>
7. Global 2000 list by the center for world university rankings. [электронный ресурс]. Режим доступа: <https://cwur.org/2021-22.php>

## NEURAL NETWORK SYSTEM FOR PREDICTING THE WORLDS BEST UNIVERSITY

*Shilova Elena A.*

National Research University “Higher School of Economics”  
Russia, Perm, 614070, Studencheskaya st., 38, eashilova\_1@edu.hse.ru

The article presents a description of the development of a neural network system for predicting the ranking of a university in the world. The system allows to predict the ranking of a certain university with great accuracy based on the data collected about other universities. With the help of the developed intellectual system, a study of the subject area was carried out, patterns of practical importance were identified.

**Keywords:** neural network technologies, forecasting, university ranking, quality of education, employment.

# **ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ В ДЕМОГРАФИИ И ЭКОЛОГИИ**



# ЭФФЕКТИВНЫЙ МОНИТОРИНГ НЕСАНКЦИОНИРОВАННЫХ МЕСТ РАЗМЕЩЕНИЯ ОТХОДОВ ПО СПУТНИКОВЫМ СНИМКАМ С ИСПОЛЬЗОВАНИЕМ МЕТОДОВ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

*Данелян Вадим Самвелович, Ключев Андрей Владимирович*

Пермский национальный исследовательский политехнический университет,  
614990, Россия, г. Пермь, Комсомольский пр., 29, vadydan@icloud.com,  
kav@gelicon.biz

Формулируются постановки задач по выявлению и локализации мест размещения отходов на спутниковых снимках земной поверхности. Собрано обучающее множество для реализации процесса обучения моделей. Предложено несколько моделей нейронных сетей для решения задачи классификации снимков, реализовано их обучение. Проведен анализ полученных результатов, сделаны выводы по задаче. В качестве итоговой модели классификатора был выбран ансамбль из нескольких сверточных нейронных сетей (Xception, ResNet50, VGG19) с жестким голосованием. Для решения задачи детектирования применялась нейронная сеть YOLOv5. Показатель AP для класса свалок при пороге 0.5 (IoU) равняется 0.898, что является хорошим результатом.

**Ключевые слова:** искусственный интеллект, глубокое обучение, сверточные нейронные сети, несанкционированные свалки, классификация, детектирование, техника переноса обучения.

**Введение.** В России и во многих других странах мусор является острой экологической и социальной проблемой. Это напрямую связано с отсутствием полноценной системы переработки и утилизации отходов, а также со слабым контролем со стороны государственных надзорных органов. Конечно, это приводит к появлению большого количества несанкционированных мест размещения отходов, что наносит огромный вред экологии. По оценкам Министерства природных ресурсов и экологии в России ежегодно образуется около 70 млн. тонн твердых отходов, которые, в свою очередь, превращаются в незаконные свалки, площадь которых увеличивается на 500 тыс. гектаров в год [1]. Тенденция вызвана тем, что из всех этих отходов, по данным Гринпис [2], сжигается менее 2%, а перерабатывается около 4%. Остальной мусор отправляется либо на специализированные полигоны, либо на несанкционированные свалки, точное количество которых неизвестно, что не позволяет делать выводы о масштабах последствий. Отметим, что оперативно-розыскная деятельность по таким делам фактически не проводится [3].

В связи с вышеизложенным, очевидна актуальность задачи распознавания и мониторинга несанкционированных мест размещения отходов. На дан-

ный момент существуют решения, заключающиеся в привлечении общественности: созданы специальные интерактивные Интернет-карты, на которых любой пользователь может оставить сообщение и обозначить несанкционированную свалку. Благодаря различным спутниковым сервисам существует возможность поиска незаконных мест размещения отходов с помощью ручного просмотра карт земной поверхности. Оба этих подхода являются довольно трудоемкими и малоэффективными. Поэтому в данной работе предлагается применять методы искусственного интеллекта, а именно – глубокого обучения, что более перспективно в решении обсуждаемой проблемы. Подход обладает явными преимуществами. Во-первых, он дает возможность оперативно выявлять новые места незаконного размещения отходов. Во-вторых, можно относительно недорого осуществлять мониторинг ранее обнаруженных мест, отслеживать изменения площади объекта, активность свалки с момента ее зарождения, фиксировать момент рекультивации и т.д. Таким образом, можно вести важную статистику по свалкам в той или иной местности.

В данной работе ставятся задачи по разработке Web-приложения с картой предполагаемых несанкционированных мест размещения отходов, по реализации и обучению моделей искусственных нейронных сетей (ИНС) для решения связанных задач: бинарной классификации спутниковых снимков земной поверхности и детектирования свалок. А главной целью является повышение эффективности выявления случаев несанкционированного размещения мусора и улучшение ситуации в сфере обращения с отходами за счет применения современных цифровых методов и технологий.

**1. Обучающее множество.** Правильное формирование обучающего множества имеет существенное значение в задачах машинного обучения. В текущей работе датасет представляет собой два набора изображений: содержащих и не содержащих свалки. Известно, что снимков со свалками крайне ограниченное количество, а их поиск может вызвать трудности.

Было принято решение загружать снимки с Яндекс Карт. Изображения представлены в цветовом пространстве RGB. План составления датасета можно описать следующим образом:

1. Поиск географических координат местонахождения свалок.
2. Перевод географических координат (широта/долгота) в координаты тайлов (квадраты размером 256×256 пикселей).
3. Загрузка снимков со спутникового сервиса.
4. Разметка снимков (классификация снимков людьми).

В результате было сформировано обучающее множество объемом более 137700 изображений для задачи классификации: 8372 со свалками и 129361 – без. Набор делился на обучающую (70%), валидационную (15%) и контрольную (15%) выборки. Стоит подметить несбалансированность датасета. Для задачи детектирования – 500 изображений: 400 для обучения и 100 для валидации.

**2. Анализ результатов исследования.** Распространенным подходом к глубокому обучению при маленьком наборе тренировочных данных является использование техники переноса обучения с помощью предварительно обу-

ченной нейронной сети. Для решения задачи бинарной классификации предлагаем использовать несколько популярных архитектур сверточных нейронных сетей: Xception [4], VGG19 [5], InceptionV3 [6], ResNet50 [7], Inception-ResNet-V2 [8]. Каждая из них была предварительно обучена на датасете ImageNet. К сверточной основе нейронных сетей применялась техника Global Average Pooling и Dropout со значением прореживания 0.2, а в конце – полносвязный выходной слой, состоящий из одного нейрона.

В качестве функции потерь выбрана бинарная кросс-энтропия, а для оценки работы нейронных сетей использовались: точность (accuracy), F<sub>1</sub>-мера и матрица неточностей (confusion matrix). Каждая ИНС перебирала данные мини-пакетами по 16 изображений и обучалась несколько эпох (были испробованы разные варианты, в основном от 3 до 10) с полностью замороженной сверточной основой (feature extraction), а после этого еще 50 эпох с размороженными верхними сверточными слоями (fine tuning). В качестве оптимизатора был выбран Adam. Из-за несбалансированности датасета использовались веса для классов в функции потерь, которые были прямо пропорциональны отношению общего числа изображений к количеству изображений в классах. Лучшие результаты приведены на рисунке 1. Уточним, что класс свалок в данной работе – положительный (positive).

№	Сверточная основа	F <sub>1</sub> -мера	Матрица неточностей	
			[TP FN]	[FP TN]
1	Xception (86)	0.978	[1221 35 ]	[ 20 19384]
2	ResNet50 (99)	0.977	[1216 40 ]	[ 17 19387]
3	VGG19 (17)	0.968	[1224 32 ]	[ 50 19354]
4	Inception-ResNet-V2 (516)	0.963	[1234 22 ]	[ 73 19331]
5	InceptionV3 (249)	0.957	[1208 48 ]	[ 60 19344]

**Рисунок 1. Результаты ИНС на валидационной выборке: в скобках указан номер слоя, начиная с которого происходило дообучение модели**

ResNet50 показала наименьшее количество ложных срабатываний и наибольшее число TN. Менее успешной ИНС по этим двум показателям стала модель Inception-ResNet-V2. Однако у нее же наибольшее количество верно распознанных свалок и наименьшее число FN. Если в задаче является приоритетным выявить как можно больше свалок, то модель №4 должна справляться лучше остальных. Если требуется наименьшее количество ложных срабатываний, то стоит посмотреть на модели 1 и 2. В зависимости от целей итогового продукта и расставленных приоритетов в задаче можно выбирать подходящие конфигурации.

Значения функции потерь на этапах обучения и валидации для лучшей модели по мере  $F_1$  (Xception) представлены на рисунке 2. По графику видно, что на тренировочных данных «ошибка» продолжает падать, однако на валидационных – немного иная картина. С 7-10 эпох значения функции потерь начинают расти. Можно наблюдать эффект переобучения модели.



**Рисунок 2. Значение функции потерь модели Xception (№1)**

Отметим, что в текущей задаче приоритетным является распознать наибольшее количество свалок и при этом получить наименьшее количество ложных срабатываний. Для выполнения этих целей и для повышения качества распознавания было предложено объединить три первые модели (Xception, ResNet50, VGG19) в ансамбль нейронных сетей с жестким голосованием – если все три модели предсказывают положительный класс, то только в этом случае ансамбль прогнозирует свалку, иначе – отрицательный. Матрица неточностей на контрольном множестве для данной модели представлена на рисунке 3. Можем заметить, что нам удалось значительно уменьшить число ложных срабатываний и при этом сохранить количество верных распознаваний на очень высоком уровне. При этом значение  $F_1$ -меры довольно высоко – 0.976.

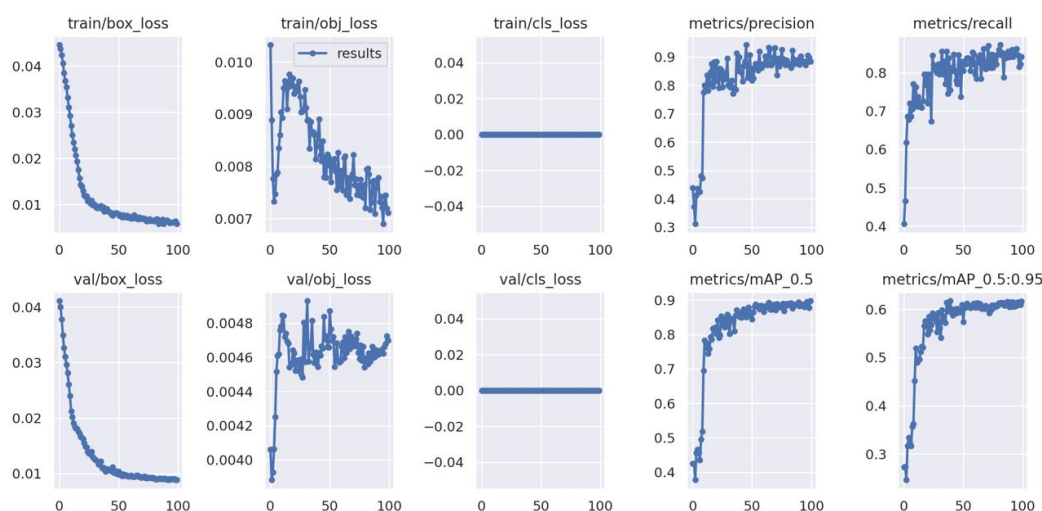
		Predict	
		1	0
Actual	1	1200	56
	0	3	19402

**Рисунок 3. Матрица неточностей для ансамбля моделей**

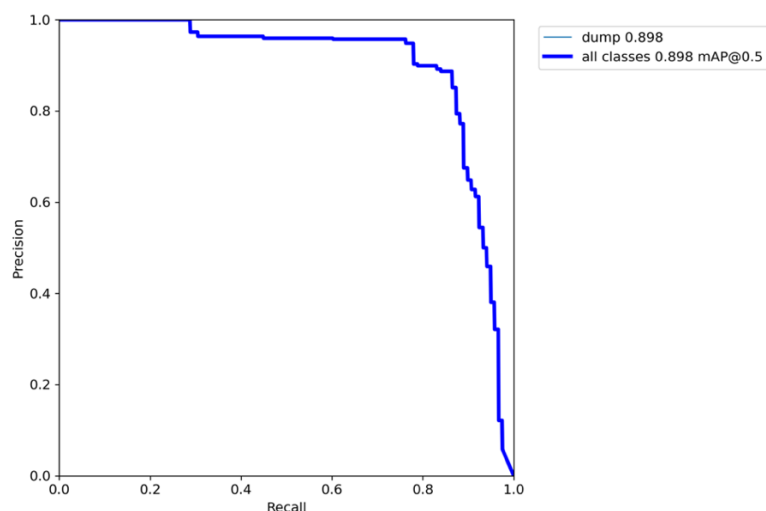
Для решения задачи детектирования свалок предложено использовать популярный алгоритм для обнаружения объектов YOLO [9], а именно – YOLOv5. Поскольку обучающих данных крайне мало, предлагается использовать технику переноса обучения. Модель предлагается взять предварительно обученной на популярном датасете Microsoft COCO.

На этапе выделения признаков было заморожено 12 первых слоев модели. Обучение проходило 150 эпох. Следующим шагом является этап до-

обучения, который состоит из размораживания всей модели и ее повторного обучения на наших данных с очень низкой скоростью обучения. Это может привести к значительным улучшениям за счет адаптации предварительно выделенных признаков к новым данным. Результаты 100 эпох обучения можно наблюдать на рисунке 4. Как мы видим по графикам, на этапе дообучения метрики и потери все еще улучшаются. Для оценки качества модели на валидационных данных предлагаем посмотреть на кривую Precision – Recall, приведенную на рисунке 5.



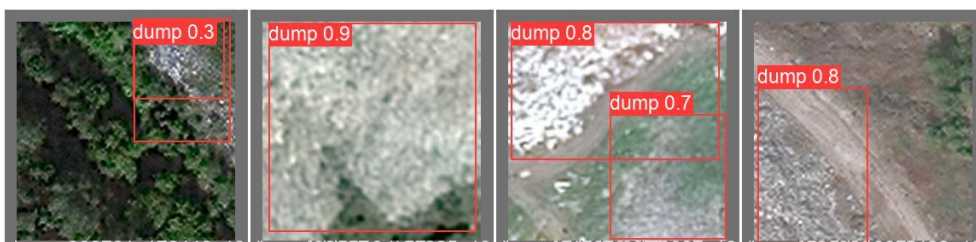
**Рисунок 4. Результаты этапа дообучения модели YOLOv5**



**Рисунок 5. Кривая Precision – Recall для итоговой детектирующей модели YOLOv5 на валидационном наборе данных**

Показатель AP для класса свалок при пороге 0.5 (IoU) равняется 0.898, что является хорошим результатом. Однако важно учитывать, что для получения полной картины о показателях модели необходимо значительно расширить датасет для детектирования. Приведем прогнозы нейронной сети на некоторых картинках из валидационного набора (см. рис. 6).

Можем отметить, что детектирующая модель вполне успешно справляется на многих объектах из валидационного набора данных.



**Рисунок 6. Прогнозы дообученной YOLOv5 на валидационном наборе**

**Заключение.** Сформулированы постановки двух смежных задач – классификации и детектирования. Затронут вопрос подготовки данных к обучению и тестированию моделей, который является немаловажным в процессе решения задач машинного обучения. Приведен алгоритм составления датасета мест размещения отходов. Рассмотрены техники построения моделей ИНС при нехватке обучающих данных. Реализовано обучение нескольких моделей ИНС. Каждая модель показала высокие результаты по мере  $F_1$ . Для борьбы с большим количеством ложных распознаваний в качестве итоговой модели классификатора был выбран ансамбль из нескольких сверточных нейронных сетей с жестким голосованием. Для решения задачи детектирования применялась нейронная сеть YOLOv5. Показатель AP для класса свалок при пороге 0.5 (IoU) равняется 0.898, что является хорошим результатом.

### **Библиографический список**

1. Кобылкин Д. Н. «Отходы – в доходы?» / беседовала Т. Богданова // Аргументы и факты. 24 апреля 2019. URL: [https://aif.ru/society/ecology/othody\\_v\\_dohody\\_glava\\_minprirody\\_o\\_tom\\_zachem\\_nuzhny\\_musornye\\_peremeny](https://aif.ru/society/ecology/othody_v_dohody_glava_minprirody_o_tom_zachem_nuzhny_musornye_peremeny) (дата обращения: 15.06.2019).
2. Chto delat' s musorom v Rossii? [Elektronnyj resurs] // Moskva: Greenpeace, 2019 // URL: <https://greenpeace.ru/wp-content/uploads/2019/10/report-RUSSIA-GARBAGE.pdf> // (data obrashcheniya 15.11.19).
3. Егоров Иван. Погрязли: аналитика: [арх. 25.02.2019] // Российская Газета : сайт. — 2018. — 28 июня.
4. Chollet F. Xception: Deep learning with depthwise separable convolutions // Proceedings of the IEEE conference on computer vision and pattern recognition. – 2017. – С. 1251-1258.
5. Simonyan K., Zisserman A. Very Deep Convolutional Networks for Large-Scale Image Recognition // arXiv 1409.1556. – 2015.
6. Szegedy C. et al. Rethinking the inception architecture for computer vision // Proceedings of the IEEE conference on computer vision and pattern recognition. – 2016. – С. 2818-2826.
7. He K. et al. Deep residual learning for image recognition // Proceedings of the IEEE conference on computer vision and pattern recognition. – 2016. – С. 770-778.
8. Szegedy C. et al. Inception-v4, inception-resnet and the impact of residual connections on learning (2016) // arXiv preprint arXiv:1602.07261. – 2016.

9. Redmon J. et al. You only look once: Unified, real-time object detection // Proceedings of the IEEE conference on computer vision and pattern recognition. – 2016. – С. 779-788.

## **EFFECTIVE MONITORING OF UNAUTHORIZED WASTE DISPOSAL SITES USING SATELLITE IMAGES USING ARTIFICIAL INTELLIGENCE METHODS**

*Danelyan Vadim S., Klyuev Andrey V.*

Perm National Research Polytechnic University,  
614990, Russia, Perm, Komsomolsky pr., 29, vadydan@icloud.com,  
kav@gelicon.biz

Statements of tasks for identifying and localizing waste disposal sites on satellite images of the earth's surface are formulated. A training set was assembled to implement the model training process. Several models of neural networks are proposed for solving the problem of image classification, their training is implemented. The analysis of the received results is carried out, conclusions on a task are drawn. An ensemble of several convolutional neural networks (Xception, ResNet50, VGG19) with hard voting was chosen as the final classifier model. The YOLOv5 neural network was used to solve the detection problem. The AP for the landfill class at the threshold of 0.5 (IoU) is 0.898, which is a good result.

**Keywords:** artificial intelligence, deep learning, convolutional neural networks, unauthorized dumps, classification, detection, transfer learning technique.

УДК 004.032.26

## **НЕЙРОСЕТЕВАЯ СИСТЕМА ПРОГНОЗИРОВАНИЯ ОЖИДАЕМОЙ ПРОДОЛЖИТЕЛЬНОСТИ ЖИЗНИ ЛЮДЕЙ ДЛЯ СТРАН**

*Шавкунов Павел Александрович*

Пермский государственный национальный исследовательский университет,  
БАС. 614990, Россия, г. Пермь, ул. Букирева, 15, pavel.shavkunov@ya.ru

В статье представлено описание разработки нейросетевой системы для прогнозирования ожидаемой продолжительности жизни в стране. Система позволяет с большой точностью предсказать среднюю ожидаемую продолжительность жизни граждан страны на основании нескольких показателей страны.

**Ключевые слова:** искусственный интеллект, нейросетевые технологии, прогнозирование, продолжительность жизни.

**Введение.** В данной работе проводилось исследование продолжительности жизни людей различных стран.

Ожидаемая продолжительность жизни зависит не только от расположения страны, пола человека. Поэтому ее прогноз является достаточно трудной задачей. Как показал опыт Пермской научной школы искусственного интеллекта, например, [1-3], правильно натренированная нейросетевая система может добиться успешности большей части прогнозов. Знание продолжительности жизни может помочь сформировать более правильную законодательную политику в отношении пенсионеров, улучшить систему здравоохранения.

При анализе литературных источников выяснилось, что работ на тему прогнозирования продолжительности жизни было проведено много. Но в данной работе использовался новый параметр «форма правления в стране».

Основная цель настоящей работы заключается в сборе множества публичных данных о различных странах, а также создание и обучение нейросетевой модели на этих данных. Конечным результатом должна быть нейросетевая система, способная прогнозировать среднюю продолжительность жизни населения страны, отклоняясь не более чем на 15%.

Для создания нейросетевой системы были выбраны следующие параметры: X1 – форма правления в государстве, X2 – площадь страны, X3 – количество граждан страны, X4 – ВВП (ППС), X5 – ВВП (номинал), X6 – Индекс человеческого развития (ИЧР). Выходной параметр – ожидаемая продолжительность жизни в стране.

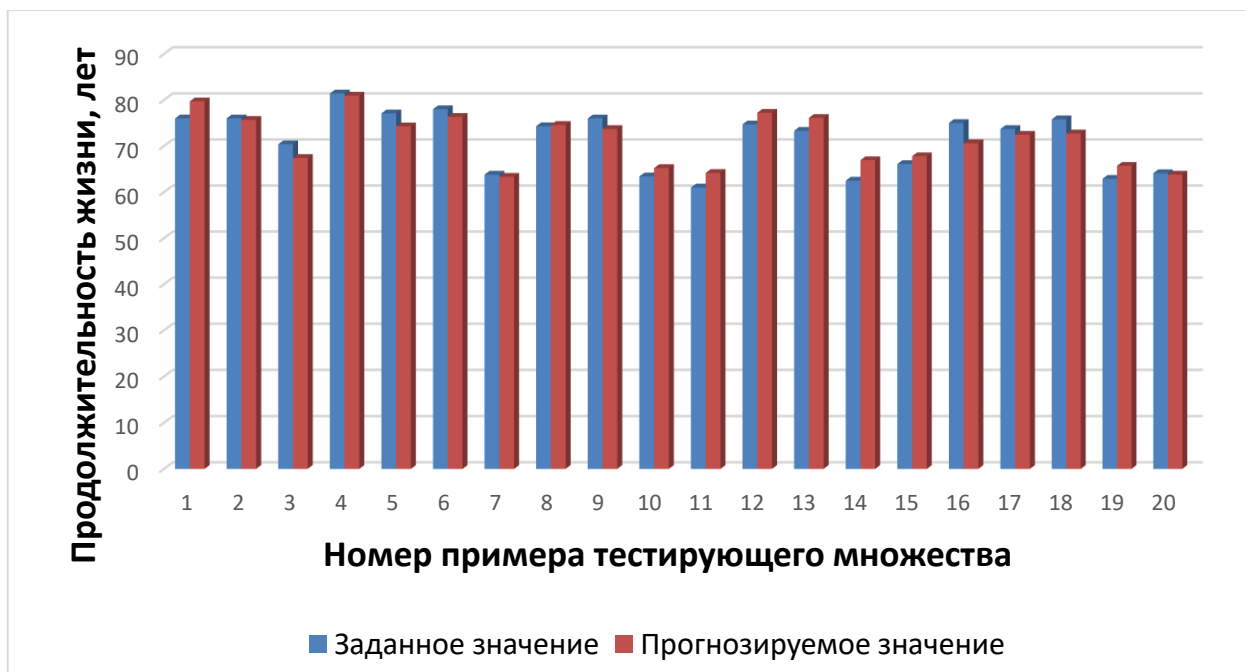
Обучающее множество было собрано вручную [4, 5]. Перед переходом к проектированию и обучению нейросетевой модели, была выполнена очистка исходного множества от противоречивых примеров, выбросов [8], дубликатов. Например, некорректными примерами считались те, где не хватало информации о стране, либо информация, собранная из разных источников, сильно отличалась. Таким образом, объем итогового множества включает в себя 175 примеров. Данное множество было разделено на обучающее, тестовое, подтверждающее в соотношении 78%, 11%, 11% соответственно. Большая часть данных была собрана с интернет-ресурсов [4, 5].

Проектирование, обучение, тестирование нейронной сети выполнялись с помощью программы «Нейросимулятор 5» [6] по методике [7]. В результате оптимизации спроектированная нейронная сеть представляла собой персептрон, который имеет шесть входных нейронов, один выходной и один скрытый слой с одним нейроном.

Для оценки корректной работы спроектированной нейронной сети использовалось тестирующее множество, состоящее из 20 примеров. Средняя относительная ошибка тестирования составила 10.6%, что можно считать приемлемым результатом. На рисунке 1 представлена гистограмма, демонстрирующая разницу между фактической и прогнозируемой нейросетью продолжительности жизни.

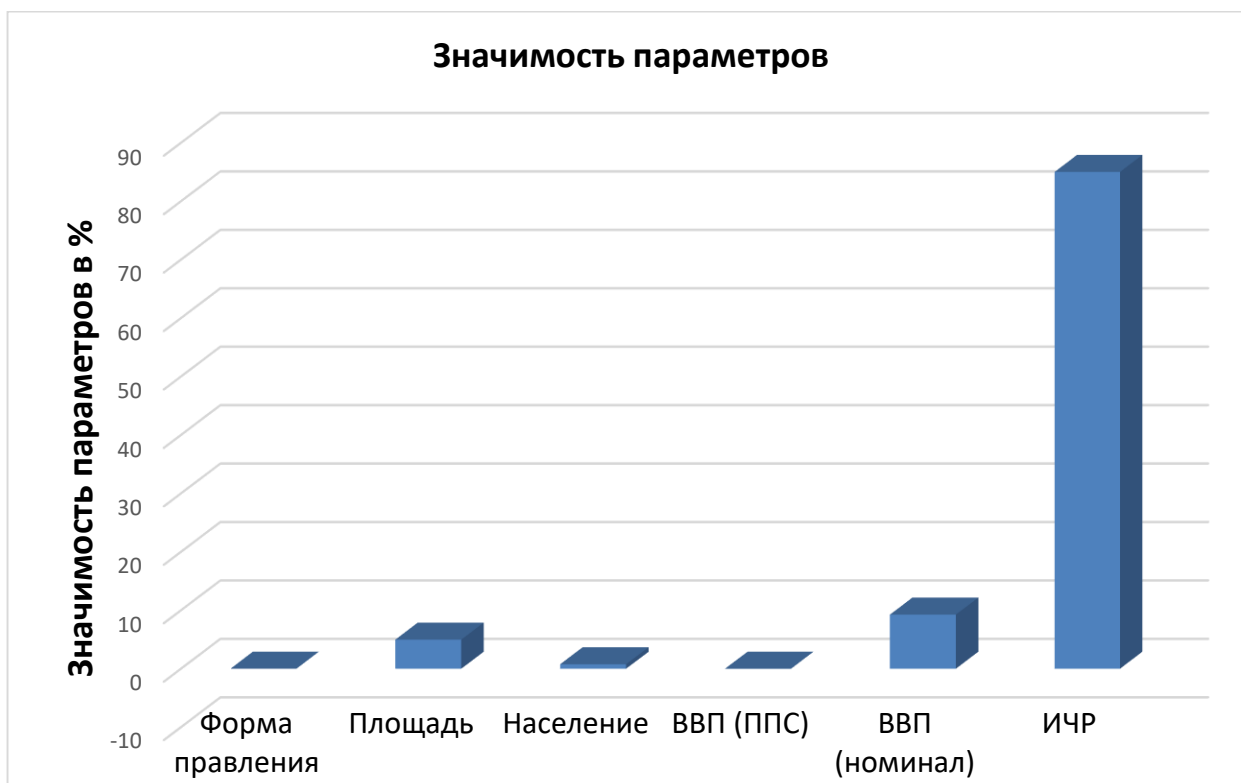
Из результатов, изображенных на рисунке 1, можно сделать вывод об адекватной работе спроектированной нейронной сети.





**Рисунок 1. Результат тестирования нейронной сети**

Оценка значимости входных параметров была выполнена с помощью программы «Нейросимулятор 5», результат отображен на рисунке 2.



**Рисунок 2. Значимость входных параметров нейросетевой модели**

Как видно из рисунка 2, наиболее значимыми являются ИЧР, ВВП (номинал), площадь страны. Наиболее влиятельным параметром оказался ИЧР. А такие параметры, как форма правления и ВВП (ППС) вообще снижали эффективность обучения нейронной сети.

Далее было проведено исследование полученных зависимостей между входными параметрами и ожидаемой продолжительности жизни. Исследование производилось с помощью метода «замораживания» [1-3, 7], суть которого заключается в варьировании значения одного параметра и фиксации значений всех других параметров.

На рисунке 3 показан график зависимости продолжительности жизни от площади страны. При увеличении площади страны, ожидаемая продолжительность жизни уменьшается, но очень незначительно.



Рисунок 3. Зависимость продолжительности жизни от площади страны

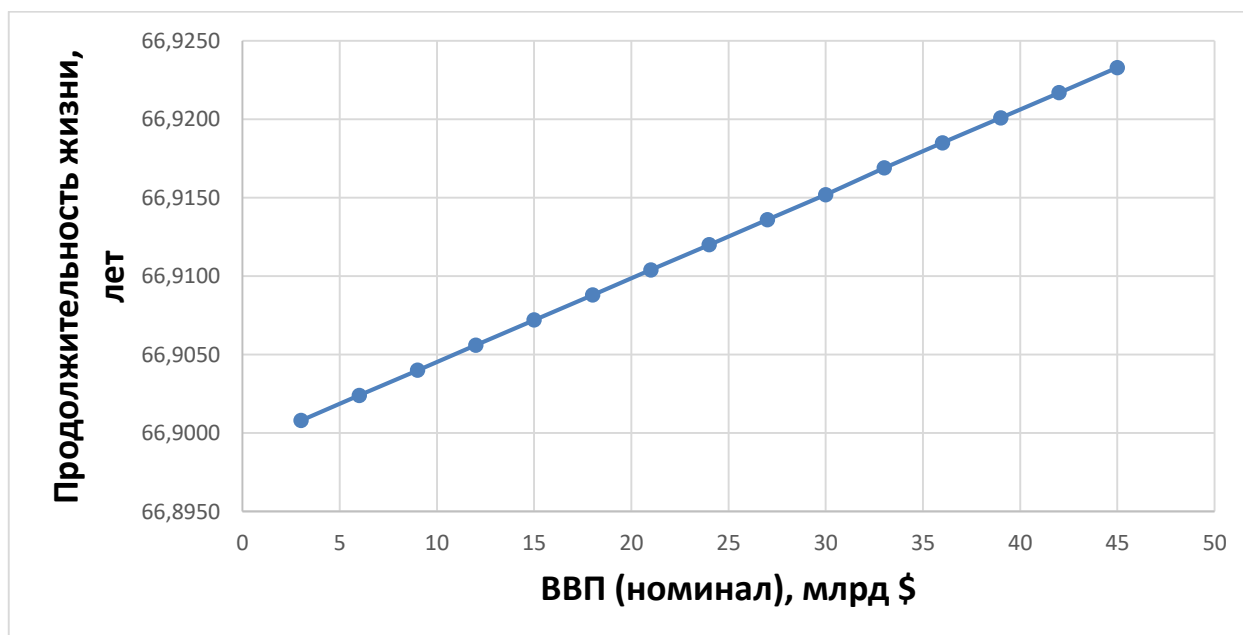
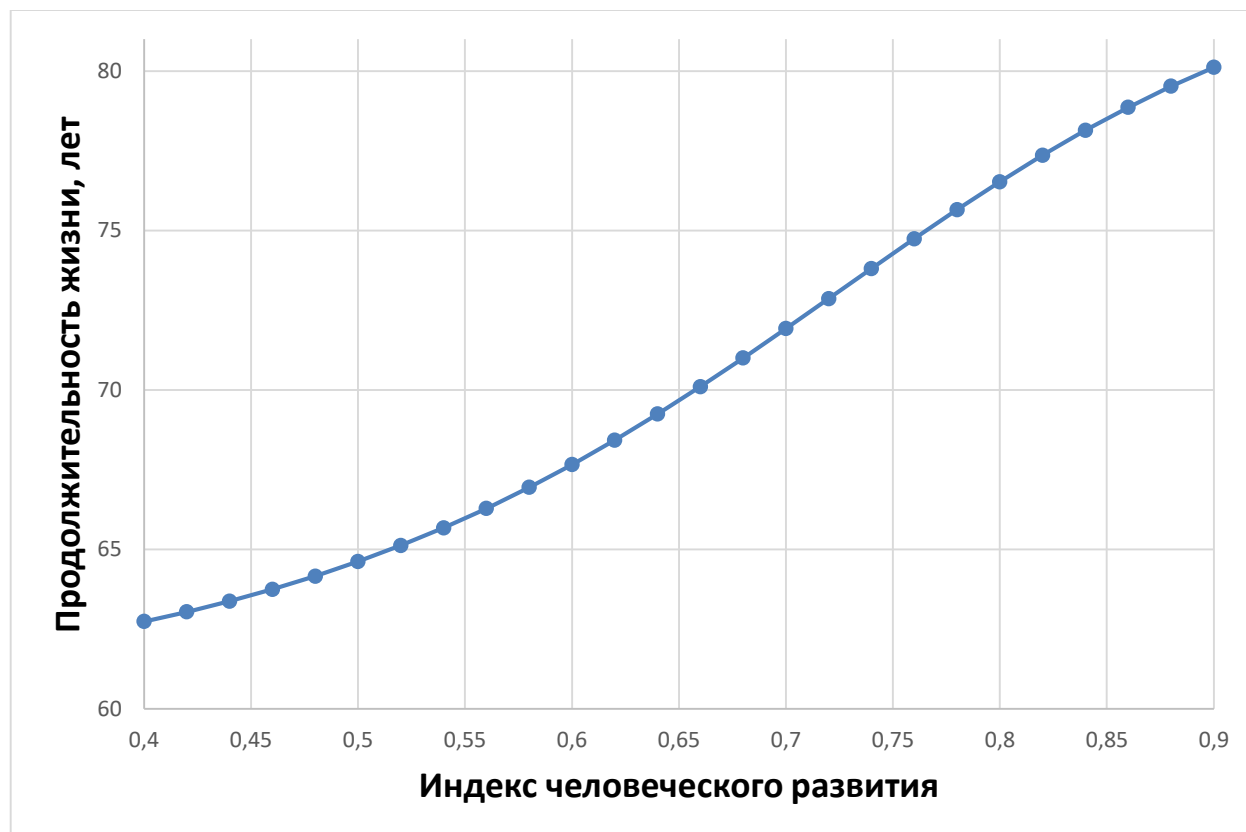


Рисунок 4. Зависимость ожидаемой продолжительности жизни от ВВП (номинал)

На рисунке 4 продемонстрирована зависимость продолжительности жизни от ВВП (номинал). Можно заметить, что при увеличении ВВП продолжительности жизни увеличивается, но незначительно.

На рисунке 5 изображен график зависимости продолжительности жизни от ИЧР. Как видно из графика, при увеличении ИЧР продолжительность жизни в стране очень быстро растет.



**Рисунок 5. Зависимость ожидаемой продолжительности жизни от ИЧР**

Полученные результаты исследований не противоречат реальности, что подтверждает, что спроектированную нейронную сеть можно считать пригодной для прогнозирования ожидаемой продолжительности жизни в различных странах.

**Заключение.** Построена система нейросетевого прогнозирования ожидаемой продолжительности жизни в стране. Спроектированная нейросетевая модель учитывает 6 параметров: форму правления в стране, площадь страны, население, ВВП (ППС), ВВП (номинал), ИЧР. Методом сценарного прогнозирования построены графики зависимостей ожидаемой продолжительности жизни от изменения входных параметров. Применение такого набора параметров в модели позволяет с высокой точностью прогнозировать продолжительность жизни людей в стране.

### **Библиографический список**

1. Ясницкий Л.Н., Зайцева Н.В., Гусев А.Л., Шур П.З. Нейросетевая модель региона для выбора управляющих воздействий в области обеспечения

гигиенической безопасности // Информатика и системы управления. 2011. № 3 (29). С. 51-59.

2. Ясницкий Л.Н., Грацилёв В.И., Куляшова Ю.С., Черепанов Ф.М. Возможности моделирования предрасположенности к наркозависимости методами искусственного интеллекта // Вестник Пермского университета. Философия. Психология. Социология. 2015. № 1 (21). С. 61-73.

3. Ясницкий Л. Н., Порошина А. М., Тавафиев А. Ф. Цвет глаз предпринимателя и успешность его бизнеса. Нейросетевые технологии как инструмент для прогнозирования успешности предпринимательской деятельности // Российское предпринимательство. 2010. № 4-2. С. 8-13.

4. wikipedia.org – Список стран по ожидаемой продолжительности жизни. [Электронный ресурс]. Режим доступа: [https://ru.wikipedia.org/wiki/%D0%A1%D0%BF%D0%B8%D1%81%D0%BE%D0%BA\\_%D1%81%D1%82%D1%80%D0%B0%D0%BD\\_%D0%BF%D0%BE\\_%D0%BE%D0%B6%D0%B8%D0%B4%D0%B0%D0%B5%D0%BC%D0%BE%D0%B9\\_%D0%BF%D1%80%D0%BE%D0%B4%D0%BE%D0%BB%D0%B6%D0%B8%D1%82%D0%B5%D0%BB%D1%8C%D0%BD%D0%BE%D1%81%D1%82%D0%B8\\_%D0%B6%D0%B8%D0%B7%D0%BD%D0%B8](https://ru.wikipedia.org/wiki/%D0%A1%D0%BF%D0%B8%D1%81%D0%BE%D0%BA_%D1%81%D1%82%D1%80%D0%B0%D0%BD_%D0%BF%D0%BE_%D0%BE%D0%B6%D0%B8%D0%B4%D0%B0%D0%B5%D0%BC%D0%BE%D0%B9_%D0%BF%D1%80%D0%BE%D0%B4%D0%BE%D0%BB%D0%B6%D0%B8%D1%82%D0%B5%D0%BB%D1%8C%D0%BD%D0%BE%D1%81%D1%82%D0%B8_%D0%B6%D0%B8%D0%B7%D0%BD%D0%B8)

5. gtmarket.ru – Рейтинг стран мира по уровню продолжительности жизни. [Электронный ресурс]. Режим доступа: <https://gtmarket.ru/ratings/life-expectancy-index>

6. Черепанов Ф.М., Ясницкий Л.Н. Нейросимулятор 4.0. Свидетельство о регистрации программы для ЭВМ RUS 2014612546. Заявка № 2014610341 от 15.01.2014.

7. Ясницкий Л.Н. Интеллектуальные системы: учебник. М.: Лаборатория знаний, 2016. 221 с.

8. Yasnitsky L.N. Algorithm for searching and analyzing abnormal observations of statistical information based on the Arnold – Kolmogorov – Hecht-Nielsen theorem // International Journal of Advanced Trends in Computer Science and Engineering. 2020. 9(2). Pp. 1814-1819. DOI: 10.30534/ijatcse/2020/139922020

## **NEURAL NETWORK SYSTEM FOR PREDICTING PEOPLE'S LIFE EXPECTANCY FOR COUNTRIES**

*Shavkunov Pavel A.*

Perm State University

Str. Bukireva, 15, Perm, Russia, 614990, [pavel.shavkunov@ya.ru](mailto:pavel.shavkunov@ya.ru)

The article describes the development of a neural network system for predicting life expectancy in the country. The system allows you to predict with great accuracy the average life expectancy of citizens of the country based on several indicators of the country.

**Keywords:** artificial intelligence, neural network technologies, forecasting, life expectancy.

# **ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ В ЛИНГВИСТИКЕ**

## РАЗРАБОТКА СИСТЕМЫ ДЛЯ АНАЛИЗА СИНТАКСИСА НАУЧНЫХ ТЕКСТОВ НА АНГЛИЙСКОМ ЯЗЫКЕ

*Гайнетдинова Вероника Андреевна*

Национальный исследовательский университет Высшая школа экономики,  
614070, г. Пермь, ул. Студенческая, д. 38, gaynetdinova.v.a@yandex.ru

*Ланин Вячеслав Владимирович*

Пермский государственный национальный исследовательский университет,  
614990, Россия, г. Пермь, ул. Букирева, 15, vlanin@live.com

*Стринюк Светлана Александровна*

Государственный университет морского  
и речного флота имени адмирала С.О. Макарова,  
198035, Россия, г. Санкт-Петербург, ул. Двинская, 5/7, strinyuksa@gumr.ru

Статья посвящена разработке системы стилистического анализа синтаксиса научных текстов на английском языке. На основе собранного корпуса текстов с помощью интеллектуального анализа текстов (text mining) разработаны алгоритмы реализации, проектирования базы данных и архитектуры системы, которая позволяет осуществлять синтаксический анализ научных текстов на английском языке. В статье описана программная реализация системы и результаты ее тестирования.

**Ключевые слова:** интеллектуальный анализ текстов, научный текст, корпусные исследования научной речи, стилистический анализ синтаксиса

**Введение.** Функциональный стиль речи, принятый в академической среде, отличается определенными признаками на многих уровнях языка: как лингвистических – формальность, обилие терминологии, номинализация (более частое, по сравнению с другими стилями речи использование существительных), наличие средств логической связи, и т.д., так и экстралингвистических – строгая логичность и однозначность (непротиворечивость) высказываний [1-3]. В письменной речи научный стиль существует в разнообразных жанрах, один из наиболее распространенных – научная статья, что обусловило отбор текстов для проведения исследования.

Анализ стилистики синтаксических конструкций научных текстов может найти самое широкое применение, в том числе, для проверки научных статей на соответствие правилам оформления статей издательств, в педагогических целях – для определения «уровня» научного текста, выполненного студентом на английском языке, с точки зрения используемого синтаксиса. Учитывая большие объемы научных текстов, которые должны регулярно проверяться и отсеиваться (в частности, издательствами), проблема быстрого и качественного анализа стиля синтаксических конструкций является актуальной.

Целью работы является разработка системы автоматизированного анализа стиля синтаксических конструкций научных текстов на английском

языке, способной сформулировать рекомендации по улучшению стиля анализируемого текста.

В ходе исследования были решены практические задачи: выделение синтаксических конструкций предложения (определение основных характеристик входящего предложения); сбор маркеров и конкретных значений по ним (метрик), являющиеся индикаторами синтаксического стиля. Автоматизация расчета метрик; выдача рекомендаций по стилю текста для улучшения качества проанализированной работы с точки зрения синтаксиса; расчет пользовательских метрик, а также проведение статического синтаксического анализа по заданным правилам; интерфейсная часть приложения – пользовательский интерфейс для работы с системой (веб-приложения), позволили сформулировать выделить основные модули разрабатываемой системы: «Синтаксический анализ», «Сбор метрик», «Стилевые рекомендации», «Пользовательские правила» и «Веб-приложение» и сформулировать функциональные требования, предъявляемые к каждому из модулей.

**Предлагаемый подход.** Реализация системы требует применения ряда средств, предназначенных для автоматизированного анализа текста – библиотеки и алгоритмы для парсинга документов, построения синтаксических деревьев и выявления частей речи в предложениях. Для синтаксического анализа текстов нами был выбран парсер SyntaxNet; выбор обусловлен скоростью обработки и предоставлении подробной информации о результате обработки.

Для получения целевых показателей и метрик был составлен корпус документов на английском языке. Полученные характеристики будут применены для проведения сравнительного анализа эталонных документов с проверяемым текстом. Следует заметить, что под набором эталонным документов понимаются не только «образцовые» с точки зрения научного стиля статьи, выполненные профессионалами, компетентными авторами, но и тексты, «имеющие отклонения» – студенческие работы (ВКР студентов разных направлений подготовки НИУ ВШЭ-Пермь на английском языке).

Оба корпуса собраны научно-учебной группой НИУ ВШЭ (Пермь) «Разработка программного обеспечения для проведения корпусных исследований английского языка» [4]. Научно-учебные корпуса разделены по направлениям подготовки студентов, а именно: менеджмент, экономика, политология, история, право, бизнес-информатика и программная инженерия. Описанные шесть корпусов содержат работы жанра Research Proposal, написанные студентами НИУ ВШЭ в рамках дисциплины «Академическое письмо на английском языке». Документы представлены в формате txt.

Корпусы профессиональных текстов представлены в виде набора статей, написанных компетентными авторами и опубликованных в ведущих зарубежных журналах по указанным областям знания. Корпусы профессиональных работ будут использованы для выявления синтаксических маркеров «образцового» научного стиля на английском языке. Для валидирования найденных маркетов будут проводиться сравнения с научно-учебными рабо-

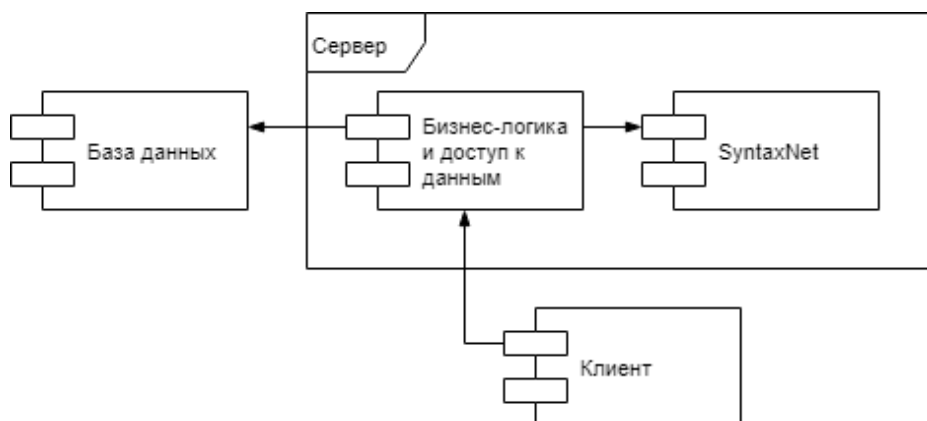
тами студентов, завершающих изучение английского языка дисциплиной Академическое письмо на английском языке.

Описанные корпуса статей разделены на два набора: обучающая выборка и тестовая. Обучающая выборка будет применена для получения показателей и метрик, являющихся индикаторами «образцового» синтаксиса научного текста. Тестовый набор будет использован для оценки того, насколько правильные результаты выдает система при проверке работы.

В метрики для проведения анализа на уровне предложения и текста включены количество подлежащих в предложении – среднее значение, максимальное и минимальное значения, медиана, мода; «relative connectors» (who, in which, at which, when, into which, who, that, where, whose, whom); деепричастные обороты (и дополнения) – отдельный подсчет по типу оборота (описывающие время, место и т.д.); conjunctions (co-coordinate and subordinate); пассивный залог – частота на единицу текста, местоположение вхождений (введение, основная часть, заключение); наличие базовой структуры научного текста – введение, основная часть, заключение; вопросительные предложения, условные предложения; анафоры -связки\ссылки на предыдущие предложения (this, that, these, those) [5, 6].

Для каждой из выборок можно рассчитать среднеквадратичное отклонение от среднего эталонного значения (посчитанного для выборки работ, написанных профессиональными авторами) [7]. Кроме того, можно учитывать глубину синтаксического дерева каждого из предложений и собирать общую статистику по проверяемому тексту – среднее значение, минимальное и максимальное значения, медиану, моду. Необходимо также учесть среднеквадратичное отклонение – показывает абсолютное отклонение измеренных значений от среднего арифметического. Для определения интервалов допустимых значений можно воспользоваться квантилем и коэффициент вариации, который характеризует меру отклонения измеренных значений от среднего. Чем меньше значение коэффициента, тем относительно меньший разброс и большая выровненность исследуемых значений. В парсере SyntaxNet используются стандартные аббревиатуры (тэги) для описания назначения токена в предложении [8].

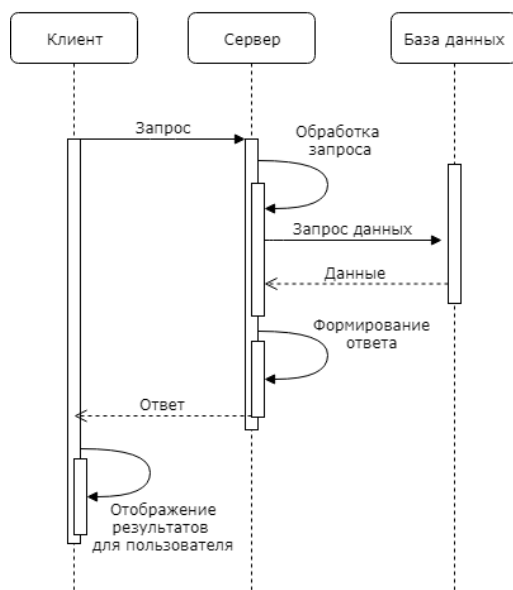
**Архитектура системы.** Взаимодействие между компонентами в виде диаграммы компонентов представлено на рисунке 1.



**Рисунок 1. Архитектура разрабатываемой системы**



Алгоритм работы разрабатываемой системы проиллюстрирован диаграммой последовательности (рисунок 2).



**Рисунок 2. Диаграмма последовательности**

Клиентское приложение – тонкий клиент, реализующий интерфейс пользователя; позволяет вводить пользовательские данные и отображать результаты работы в виде, понятном для пользователя.

Серверное приложение – приложение, реализующее бизнес-логику системы; реализует взаимодействие с базой данных; производит синтаксический анализ пользовательских документов.

База данных – хранилище данных системы; взаимодействие с сервером осуществляется посредством запросов.

Проверка стиля синтаксических конструкций загруженных текстов по определенной тематике производится следующим образом:

- С помощью SyntaxNet строится синтаксическое дерево каждого из предложений проверяемого текста.

- Производится подсчет статистики использования заданных синтаксических конструкций.

- Полученные значения по каждому маркеру сравниваются с эталонными диапазонами, рассчитанными по корпусу профессиональных работ (по каждой тематике отдельно).

- Формируется текстовый отчет о соответствии работы эталонным значениям.

В ходе анализа производится подсчет частотности употребления различных синтаксических конструкций. Возможные варианты конструкций формировались с помощью дерева: считался узел и его предок. Производился подсчет синтаксических конструкций, состоящих из одного, двух и более синтаксических единиц текста. Всего была подсчитана частота употребления более чем 900 различных синтаксических конструкций. Также к этим метри-

кам были добавлены подсчет количества вопросительных предложений и определение синтаксической сложности предложений на основе подсчета глубины (высоты) синтаксического дерева.

Кроме описанных выше маркеров, длина предложений (выступает как один из индикаторов синтаксической сложности предложений), порядок слов в предложении, использование атрибутов перед существительными, подлежащее в начале предложения и др.

В ходе анализа подтвердились не все гипотезы. Для первоначального отсеивания неподходящих гипотез использовался коэффициент вариации, показывающий насколько выборка выровнена, иначе насколько велик разброс значений случайной величины. И таким образом, в процессе исследования данный коэффициент позволял отсеивать самые неудачные гипотезы.

Набор значимых метрик определялся по каждой области научного знания. Как оказалось, даже работы, принадлежащие к одной области научных знаний (например, к социальным наукам), но к разным дисциплинам (например, экономика и политология), имеют разные не только значения эталонных диапазонов по собираемым метрикам, но в них отличается и сам набор метрик.

Для определения наиболее вероятных значений параметров по каждой из выбранных статистических характеристик с помощью двустороннего квантиля уровня  $\alpha$  можно задать интервал, в который анализируемая случайная величина попадает с заданной вероятностью. Далее необходимо построить функцию распределения. Диаграмма накопленных частот является аналогом (эмпирическим) интегральной функции распределения. Полученные результаты означают, что 5% выборки имеют среднюю синтаксическую сложность предложений менее 2,4002 и 5% выборки со значением средней сложности более 2,7758. Таким образом, на основе собранных эмпирических данных были определены допустимые диапазоны по каждому маркеру.

**Заключение.** Таким образом, была разработана система, которая позволяет производить анализ стиля синтаксических конструкций научных текстов на английском языке. Планируется использовать данную систему для обучения студентов хорошему синтаксическому стилю научных текстов на английском языке. Также планируется произвести работы по расширению функциональности разработанного веб-приложения. В частности, планируется добавить визуализацию результатов проверки работы в графическом виде для всех пользователей (сейчас такая функциональность доступна только администратору), а также улучшить систему стилевых рекомендаций.

### Библиографический список

1. Кожина М.Н. О соотношении стилей языка и стилей речи с позиции языка как функционирующей системы // Принципы функционирования языка в его речевых разновидностях / Под ред. Э.И. Матвеевой. – Пермь: Пермский государственный университет, 1984. – С. 3-18.
2. Кожина М.Н. Стилистика русского языка. – М., 1993. – 224 с. 15.

3. Кожина М.Н. Пути развития стилистики русского языка во 2-ой половине XX века // Речеведение и функциональная стилистика: вопросы теории / Избранные труды под ред. Кожиной М.Н. – Пермь, 2002. – С. 63-100.
4. Корпусы – научно-исследовательская группа «Разработка программного обеспечения для проведения корпусных исследований английского языка» // НИУ ВШЭ – Пермь. URL: <https://perm.hse.ru/bi/sfcr/corpora> (дата обращения: 07.05.2018).
5. Lynch T., Anderson K. Grammar for Academic Writing. English Language Teaching Centre University of Edinburgh. 2013. 90 с.
6. Siepmann D., Gallagher J.D., Hannay M., Mackenzie J.L. Writing in English: A Guide for Advanced Learners. 2011. 479 с.
7. Ивченко Г., Медведев Ю. Математическая статистика. Либроком. 2014. 353 с.
8. Marneffe M-C., Manning C.D. Stanford typed dependencies manual. 2008. 20 с.

## **A SYSTEM FOR SYNTACTIC ANALYSIS OF ACADEMIC TEXTS IN ENGLISH**

*Gainetdinova Veronica A.*

National Research University “Higher School of Economics”  
Studencheskaya str., 38, Perm, Russia 614070, [gainetdinova.v.a@yandex.ru](mailto:gainetdinova.v.a@yandex.ru)

*Lanin Viacheslav V.*

Perm State University

Bukireva Str., 15, Perm, Russia, 614990, [vlanin@live.com](mailto:vlanin@live.com)

*Strinyuk Svetlana A.*

Admiral Makarov State University of Maritime and Inland Shipping  
Dvinskaya Str., 5/7, St. Petersburg, Russia, 198035, [strinyuksa@gumrf.ru](mailto:strinyuksa@gumrf.ru)

The article is aimed at the development of application for analysis of academic text syntax in English. Based on the academic text corpus of competent and novice writers using text mining the authors have worked out implementation algorithms and database design and architecture of the system which allows carrying out syntactic analysis of scientific texts in English. The paper describes the software implementation of the application and the results of its testing.

**Keywords:** text mining, data mining, academic text, corpus research on academic writing, stylistic analysis of syntax.

# **ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ В КИНЕМАТОГРАФИИ**

## НЕЙРОСЕТЕВАЯ СИСТЕМА ПРОГНОЗИРОВАНИЯ РЕНТАБЕЛЬНОСТИ КИНОБИЗНЕСА

*Чепокоев Елизар Сергеевич*

Национальный исследовательский университет «Высшая школа экономики»  
614060, Россия, г. Пермь, ул. Бульвар Гагарина, 37а, eschepokov@mail.ru

В статье представлено описание разработки нейросетевой системы для прогнозирования рентабельности кинобизнеса, а также описывается анализ предыдущих исследований в этой области и существующих решений. Система позволяет предсказать кассовые сборы фильма и шанс окупаемости. С помощью разработанной интеллектуальной системы проведено исследование предметной области, выявлены закономерности, имеющие практического значения.

**Ключевые слова:** искусственный интеллект, нейросетевые технологии, прогнозирование, кинобизнес, фильмы.

**Введение.** Производство кинофильмов, было и остается одним из наиболее рискованных видов предпринимательства, особенно во времена ограничений на посещение общественных мест. Затрачивая огромные бюджеты в несколько сотен миллионов долларов на производство фильма кинокомпания ожидают соответствующие доходы от показа фильма в кинотеатрах и продаж на стриминговых сервисах, но инвесторы отмечают, что предсказать финансовый успех фильма крайне сложно даже для опытных участников рынка. Система прогнозирования способна снизить риски неудачных инвестиций в кинопроизводство на ранних этапах создания кинокартины, а также спрогнозировать целесообразность выпуска кинокартины на экраны кинотеатров с сопутствующими затратами на рекламу и аренду кинозалов.

Одними из первых исследователей, применивших в кинобизнесе метод экономико-математического моделирования, были J. Prag и J. Casavant, которые в 1994 опубликовали статью с сообщением о создании регрессионной модели на основе выборки из 625 американских фильмов [3]. Рассматриваемый ими набор входных переменных включал производственный бюджет, критические обзоры, наличие звезд, наличие франшизы, наличие премий, жанр и рейтинг. Аппарат нейронных сетей для прогнозирования кассовых сборов фильмов был впервые применен в 2002 году американскими учеными R. Sharda и D. Delen. В 2006 году эти же авторы построили модели на основе логистической регрессии, дискриминантного анализа, классификационного и регрессионного дерева, а также нейронной сети, показавшей наилучший результат [5]. В 2010 году они повторили прогнозирование и смогли достичь лучшего результата включив деревья решений и более полную выборку

фильмов [6]. Эффективность использования искусственного интеллекта при оценке кассовых сборов фильмов так же была обоснована в работах российских и зарубежных исследователей [1, 7, 8, 11, 12].

Данное исследование является расширением вышеупомянутых работ по созданию динамической системы оценки успеха фильмов. Основная цель настоящей работы заключается в создании множества на основе фильмов, вышедших в прокат, а также создании и обучении нейросети на основе собранных данных. Конечным результатом должна быть нейросетевая система, способная прогнозировать кассовые сборы фильмов с погрешностью не более 30 %.

При построении нейросетевой модели в качестве входных данных было решено выбрать параметры, представленные в таблице 1.

Таблица 1

Входные параметры нейронной сети

Номер	Название	Расшифровка, комментарии
X1	Возрастное ограничение фильма	Целое число (закодированы в числа от 1 до 5: 0+: 1; 6+: 2; 12+: 3; 16+: 4; 18+: 5)
X2	Длительность фильма в минутах	Целое число
X3	Сезон выхода фильма	Целое число (Закодированы в числа от 1 до 4: 1-весна, 2-лето, 3-осень, 1-зима)
X4	Выход фильма в период высокой посещаемости кинотеатров	Шкала 0/1, описывает вышел ли фильм в период летних или зимних каникул или иных длинных праздников
X5	Наличие у режиссеров престижных наград	Шкала 0/1, описывает наличие у режиссера наличие таких наград как «Оскар», «Золотой глобус», SAAG и др.
X6	Наличие у сценаристов престижных наград	Шкала 0/1, описывает наличие у сценаристов наличие таких наград как «Оскар», «BAFTA» и др.
X7	Наличие у 3-х звезды на главных ролях престижных наград	Шкала 0/1, описывает наличие у каждого из 3х звезд на главных ролях наличие таких наград как «Оскар», «Золотой глобус», SAAG и др.
X8	Количество оскаров у съемочной группы	Целое число, суммарное количество полученных премий «Оскар» у звезд, режиссера и сценаристов
X9	Основной жанр фильма	Целое число (Закодированы в числа от 1 до 25: 1-Action, 2-Adventure, 3-Drama и т.д.)
X10	Является ли фильм частью франшизы	Шкала 0/1, описывает является ли фильм продолжением франшизы
X11	Бюджет фильма	Положительное число (в долларах США)

Выходной параметр (Y1) – Кассовые сборы фильма (сгруппированные по множествам \$1-2,5 млн, \$2,5-5 млн и т. д.).

Обучающее множество было собрано с помощью программы-парсера с интернет-ресурса IMDb [2]. Всего были собраны данные о 1000 фильмах из рейтинга топ-500 и низ-500, после чего была произведена очистка данных от выбросов, пустых и нестандартных данных, таким образом в итоговое мно-

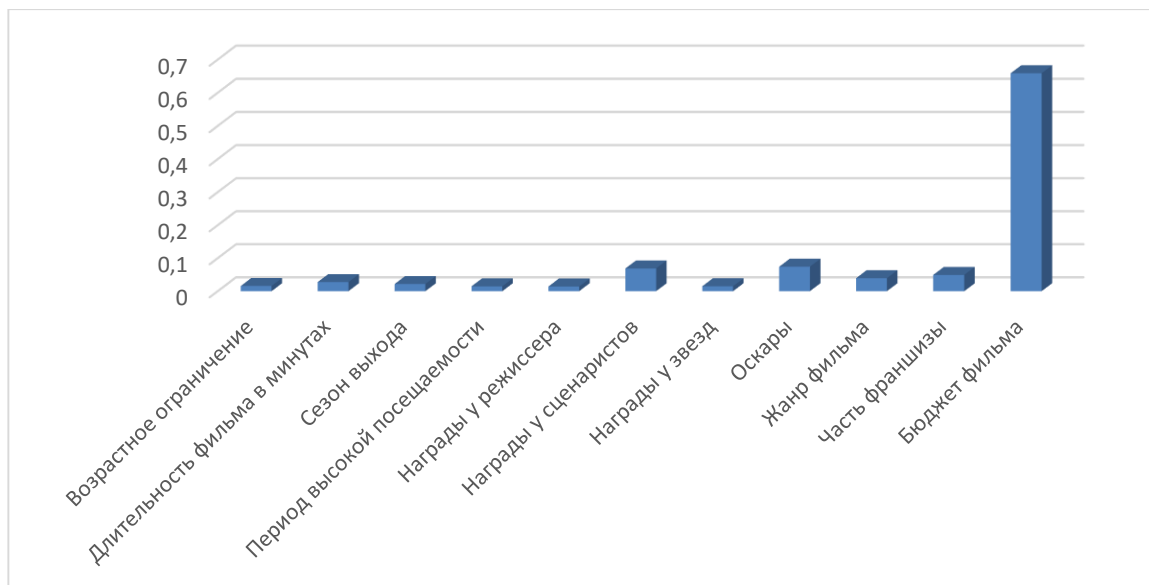
жество вошли 632 примера. Данное множество было разделено на обучающее и тестирующее в соотношении 80% к 20%.

Проектирование, обучение, тестирование нейронной сети выполнялись с помощью программы «Нейросимулятор 5» [9] по методике [10]. После оптимизации, спроектированная нейронная сеть представляла собой персептрон, который имеет одиннадцать входных нейронов, один выходной и один скрытый слой с тремя нейронами.

Средняя ошибка тестирования составила 21.7%, что можно считать приемлемым результатом. На рисунке 1 представлена гистограмма, демонстрирующая разницу между фактическими и прогнозируемыми нейросетью кассовыми сборами, среди случайно отобранных 35 фильмов из тестирующего множества.



**Рисунок 1. Результат тестирования нейронной сети (Нейросимулятор)**



**Рисунок 2. Значимость входных параметров нейросетевой модели**

Оценка значимости входных параметров была выполнена с помощью программы «Нейросимулятор 5», результат отображен на рисунке 2.

Как видно из рисунка 2, наиболее значимым являются бюджет фильма. На втором и третьем месте по значимости являются количество оscarов у съемочной группы и Наличие наград у сценаристов.

Далее был разработан нейросетевой симулятор с применением языка программирования «Python», с использованием библиотеки «Keras». Средняя ошибка тестирования с помощью разработанной нейросети составила 24.2%. На рисунке 3 представлена гистограмма, демонстрирующая разницу между фактическими и прогнозируемыми нейросетью кассовыми сборами.

Далее было проведено исследование кинорынка. Исследование производилось с помощью метода «замораживания» [12], суть которого заключается в варьировании значения одного параметра и фиксации значений всех других параметров. Данный метод позволяет выявить влияние исследуемого параметра на выходной. Для этого были выбраны 3 случайных фильма. Результаты представлены в виде графика на рисунке 4.



Рисунок 3. Результат тестирования нейронной сети (Keras)

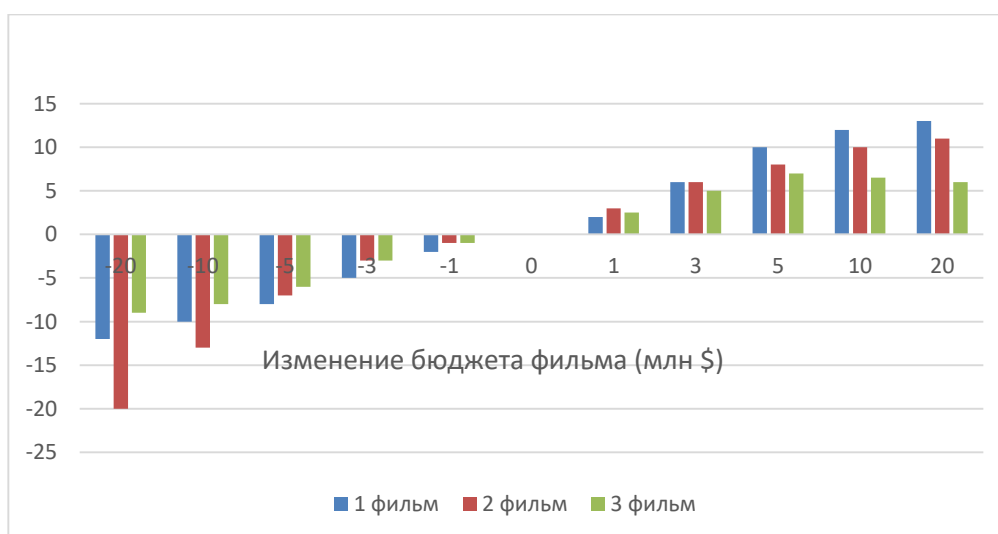


Рисунок 4. Зависимость кассовых сборов (млн \$) от бюджета фильма



Опираясь на данные результаты, можно с уверенностью сказать, что бюджет влияет на итог кассовых сборов, но данное влияние просматривается по-разному для каждого фильма, в зависимости от совокупности побочных входных данных. Полученные результаты исследования не противоречат действительности, что подтверждает, что спроектированную нейронную сеть можно считать пригодной для прогнозирования кассовых сборов фильмов.

**Заключение.** В ходе проделанного исследования выявлены наиболее значимые критерии, влияющие на успех фильма. Так же продемонстрированы способы использования созданной модели для получения различных статистических данных способствующих повышению качества фильма.

Практическая ценность исследования основывается на необходимости решения важных задач, таких как, научное обоснование и создание эффективно функционирующего механизма прогнозирования коммерческого потенциала кинопроекта, принятия рациональных управленческих решений о целесообразности его реализации, определение направлений и принципов эффективного управления кинематографическим бизнес-процессом, в частности, на ранних стадиях создания кинофильмов.

### Библиографический список

1. Eliashberg J., Sawhney M.S. Modeling Goes to Hollywood: Predicting Individual Differences in Movie Enjoyment. *Management Science*, 1994, vol. 40, iss. 9, pp. 1151–1173.
2. Internet Movie Database [Электронный ресурс], 2022, URL: <https://www.imdb.com/> (дата обращения: 20.03.2022).
3. Prag J., Casavant J. An empirical study of the determinants of revenues and marketing expenditures in the motion picture industry // *Journal of Cultural Economics*, 1994, vol. 18(3), pp. 217–235.
4. Litman B.R. Predicting Success of Theatrical Movies: An Empirical Study. *Journal of Popular Culture*, 1983, vol. 16, no. 9, pp. 159–175. doi: 10.1111/j.0022-3840.1983.1604\_159.x
5. Sharda R., Delen D. Predicting box-office success of motion pictures with neural networks // *Expert Systems with Applications*, 2006, vol. 30, pp. 243–254.
6. Sharda R., Delen D. Predicting the financial success of Hollywood movies using an information fusion approach // *Industrial Engineering Journal*, 2010, vol. 21, pp. 30–38.
7. Yasnitskii L.N., Beloborodova N.O., Medvedeva E.Yu. The method for forecasting box-office grosses of movies with neural network // *Digest Finance*. 2017. Т. 22. № 3 (243). С. 298-309.
8. Yasnitsky, L.N., Mitrofanov, I.A., Immis, M.V. Intelligent System for Prediction Box Office of the Film // *Lecture Notes in Networks and Systems*. 2020. Vol. 78. Pp. 18-25. [https://doi.org/10.1007/978-3-030-22493-6\\_3](https://doi.org/10.1007/978-3-030-22493-6_3).
9. Черепанов Ф.М., Ясницкий Л.Н. Нейросимулятор 4.0. Свидетельство о регистрации программы для ЭВМ RUS 2014612546. Заявка № 2014610341 от 15.01.2014.

10. Ясницкий Л.Н. Введение в искусственный интеллект. М.: Издательский центр «Академия», 2005. – 176 с

11. Ясницкий Л.Н., Белобородова Н. О., Медведева Е.Ю. Методика нейросетевого прогнозирования кассовых сборов кинофильмов // Финансовая аналитика: проблемы и решения. 2017. Т. 10. № 4 (334). С. 449-463.

12. Ясницкий Л.Н., Плотников Д.И. Экономико-математическая нейросетевая модель для оптимизации финансовых затрат в кинобизнесе // Фундаментальные исследования. 2016. № 11-2. С. 339-342.

## **DEVELOPMENT OF MVP MACHINE LEARNING SYSTEM FOR PREDICTING THE PROFITABILITY OF THE FILM BUSINESS**

*Chepokov Elizar S.*

National Research University “Higher School of Economics”  
Str. Gagarin Boulevard, 37a, Perm, Russia, 614060, eschepokov@mail.ru

The article describes the development of a neural network system for predicting the profitability of the film business, and also describes the analysis of previous research in this area and existing solutions. The system allows you to predict the box office of the film and the chance of payback. With the help of the developed intellectual system, a study of the subject area was carried out, patterns of practical importance were identified.

**Keywords:** artificial intelligence, neural network technologies, forecasting, film-making industry, films.

УДК 004.032.26

## **НЕЙРОСЕТЕВАЯ СИСТЕМА ПРОГНОЗИРОВАНИЯ ПОЛУЧЕНИЯ ОСКАРА**

*Веверица Карина Евгеньевна*

Национальный исследовательский университет «Высшая школа экономики»,  
614990, Россия, г. Пермь, ул. Студенческая, 38, karinaev99@mail.ru

В статье представлено описание разработки нейросетевой системы для прогнозирования получения Оскара. Система позволяет с большой точностью предсказать победителя Оскар, главной кинематографической премией в США, на основании данных о кинофильмах. С помощью разработанной интеллектуальной системы проведено исследование предметной области, выявлены закономерности, имеющие практическое значение.

**Ключевые слова:** нейронная сеть, искусственный интеллект, прогноз, фильмы, оскар, награда, персептрон, тестовое множество.

**Введение.** Прогнозирование получения Оскара кинофильмом всегда было, есть и будет очень актуально. Особенно это актуально для ценителей кино и тех, кто это кино создает: это актёры, режиссёры, продюсеры, дизайнеры и любые другие профессионалы, задействованные в кинопроизводстве. Какие факторы влияют на получение Оскара? Всегда ли номинация означает победу? Влияет ли жанр кинофильма? Как получить Оскар? Именно этому и посвящено данное исследование. Целью данной работы стало создание системы прогнозирования получения Оскара. Данная цель также подразумевает определение ключевых факторов успеха и их влияние на получение заветной награды. В данной статье представлен анализ существующих работ по этому направлению, выявлены тезисы, которые были применены при проектировании системы. В процессе работы в нейросимуляторе была разработана система, с помощью которой можно выполнять прогнозы получения фильмом Оскара. В основе системы лежит нейронная сеть, обученная на результатах кассовых сборов фильмов США и других стран при участии США. Итогом работы стала система, способная дать прогноз в отношении того, получит ли фильм Оскар, а также показать влияние различных параметров на итоговый результат и выявить самые важные из них. По мимо этого, на примерах различных фильмов разработан ряд рекомендаций, которые помогут увеличить шанс получения Оскара.

**Анализ литературы.** Киноиндустрия — это молодая, но в то же время перспективная и непрерывно растущая отрасль мировой экономики. Возможные параметры успешности фильма, его кассовых сборов, получения потенциальных наград всегда притягивали большое внимание многих исследований. Так И. Г. Князева и Д. М. Иванова в своей статье «Прогнозирование кассовых сборов проката фильма» [1] отмечают, прогнозирование успеха кинофильма, и, следовательно, его кассовых сборов – порой очень трудно в связи с большим количеством причин. Факторы качества рекламы кинофильма, актерский состав, сюжет, конечно, имеют большое значение, но далеко не всегда могут дать гарантию успеха. Однако авторы отмечают, что фактор жанра кинофильма имеет большую значимость, определяющую его потенциальный успех. В частности, жанры «боевик» и «комедия» имеют больший шанс на успех у публики. Помимо этого, авторы отмечают и значимость такого параметра как «страна производства» и «Рейтинг».

Сложность предсказания успеха кинокартин подтверждают и авторы Ясницкий Л. Н., Белобородова Н. О., Медведева Е. Ю., в своем исследовании «Методика нейросетевого прогнозирования кассовых сборов кинофильмов» [2]. Авторы утверждают, что на параметры успеха картины влияют множество факторов. В исследовании также отмечается, что немаловажным фактором помимо вышеперечисленных являются и кассовые сборы. Однако нельзя не отметить тот факт, что зачастую кассовые сборы не коррелируются с оценкой, рейтингом кинофильма в действительности. Помимо этого было отмечено, что сиквелы (книги или фильмы, которые являются продолжением предыдущего произведения) приветствуются публикой лучше, чем фильмы с новым сюжетом.

Магистр НИУ ВШЭ, Е.А. Педяш, в своем исследовании «Эконометрическое прогнозирование кассового успеха кинофильмов» [3] аналогичным образом пытался найти зависимость между параметрами, которые влияют на кассовых успех фильма во время проката. Так, в своем исследовании помимо критериев жанра, рейтинга, страны производства кинофильма, актерского состава, бюджета, он также использовал критерии объема рекламы, влияние праздников, технические параметры фильма и другие. И если технические параметры фильма целиком зависят от его бюджета, то критерий выпадения даты выхода фильма на выходные (или же праздничные дни) является несколько субъективным, т. к., как правило, прокат кинофильма происходит в течение нескольких недель, что в любом случае позволит любому желающему посетить кинокартину, если же он этого захочет.

Таким образом, все исследователи сходятся на том, что прогнозы кассовых сборов, успешности кинофильма как среди обычной аудитории, так и среди именитых критиков зависит от множества факторов, более того эти факторы часто меняются. Однако несмотря на все это можно продолжать делать попытки в поиске наиболее релевантных параметрах для оценки успешности фильма и продвигаться все глубже в изучение данной предметной области.

**Методика прогнозирования.** Для того, чтобы обучить нейронную сеть, были собраны данные о фильмах в количестве 100 с 1975 по 2021 годов производства. Главным источником данных стал сайт с фильмами номинантами и победителями премии Оскар <https://www.kinopoisk.ru/lists/movies/oscar-best-film/> [4], агрегирующий информацию о кинофильмах, актерах, процессе съемок и др. Перед началом работы был сформулирован список входных параметров, однако стоит отметить, что параметры «Страна производства» и «Жанр» были разбиты на множество входных параметров по причине того, что очень часто фильмы снимаются сразу же в нескольких странах. По такому же принципу распределены параметры жанра. В качестве примера можно привести фильм «Крестный отец», жанром которого является и драма, и криминал просто в разной степени. Входные параметры представлены ниже:

$x_1$  – Год производства.

$x_2$  – США.

$x_3$  – Европа.

$x_4$  – Азия.

$x_5$  – Другие страны.

$x_6$  – Детектив.

$x_7$  – Драма.

$x_8$  – Фэнтези.

$x_9$  – Боевик.

$x_{10}$  – Триллер.

$x_{11}$  – Комедия.

$x_{12}$  – Криминал.

$x_{13}$  – Биография.

$x_{14}$  – Приключения.

$x_{15}$  – Фантастика.

$x_{16}$  – Рейтинг IMDb.

$x_{17}$  – Номинация на Оскар.

Выходным параметром является получение награды Оскар.

$d_1 - 0$  – не получит,  $100$  – получит.

Примеры были разбиты на три множества: обучающее, валидирующее и тестирующее. Для обучения нейронной сети были отобраны 73 примера, валидирующее множество состоит из 13 примеров, для проведения проверки также было создано 13 примеров. Тестирующее множество позволяет проверить корректность работы нейросети. На рисунке 1 представлена часть обучающего множества, а на рисунке 2 – тестирующее множество.

	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12	X13	X14	X15	X16	X17	D1	
Название фильма	Год прои: США	США	Европа	Азия	Другие страны	Детектив	Драма	Фэнтези	Боевик	Триллер	Комедия	Криминал	Биография	Приключен	Фантастика	Рейтинг IMDb	Номинация на Оскар	Оскар	
Тренер Картер	2005	90	60	0	0	0	90	0	0	0	0	0	0	70	0	0	70	0	0
Рестлер	2008	90	80	0	0	0	100	0	0	0	0	0	0	0	0	0	60	0	0
Побег из Шоушенка	1994	100	0	0	0	0	100	0	0	0	0	0	0	0	0	0	80	100	0
Нелюбовь	2017	0	100	0	0	0	100	0	0	0	0	0	0	0	0	0	50	100	0
Головой о стену	2004	0	80	80	0	0	100	0	0	0	0	0	0	0	0	0	40	0	0
Джокер	2019	100	0	0	0	0	90	0	0	90	0	0	0	0	0	0	90	100	100
Пятый элемент	1997	0	100	0	0	0	0	0	80	0	70	0	0	0	80	100	70	0	0
Паразиты	2019	0	0	100	0	0	80	0	0	50	0	0	0	0	0	0	90	100	100
И грянул гром	2002	100	0	0	0	0	0	0	0	0	0	0	0	0	100	0	40	0	0
Все любят китов	2012	100	0	0	0	0	100	0	0	0	0	0	0	0	0	0	60	0	0
Призрачная красота	2016	100	0	0	0	0	100	0	0	0	0	0	0	0	0	0	50	0	0
В диких условиях	2007	100	0	0	0	0	90	0	0	0	0	0	0	60	80	0	70	100	0
Исцеляющая Любовь	2007	0	100	0	0	0	100	0	0	0	0	0	100	0	0	0	50	0	0
Привидение	1990	100	0	0	0	0	100	0	0	80	0	0	0	0	0	0	90	100	100
Реквием по мечте	2000	100	0	0	0	0	90	0	0	20	0	0	0	0	0	0	90	100	100
Утомленные солнцем	1995	0	100	0	0	0	100	0	0	0	0	0	0	0	0	0	90	100	100
Знакомьтесь, Джо Блек	1998	100	0	0	0	0	100	90	0	0	0	0	0	0	0	0	70	0	0
Миллионер из Трущоб	2008	80	70	0	0	0	100	0	0	0	0	0	0	0	0	0	80	100	100
Амели	2001	0	100	0	0	0	100	0	0	0	80	0	0	0	0	0	70	100	0
Человек дождя	1988	100	0	0	0	0	100	0	0	0	0	0	0	0	0	0	90	100	100
Эффект Бабочки	2004	100	0	0	0	0	90	0	0	80	0	0	0	0	0	80	70	0	0
День, когда Земля остановилась	2008	100	0	0	0	0	80	0	0	80	0	0	0	0	80	0	50	0	0
Александр	2004	90	80	20	0	0	90	0	80	0	0	0	0	80	0	0	60	0	0
Лара Крофт 1	2001	100	0	0	0	0	0	60	80	60	0	0	0	0	0	0	50	0	0

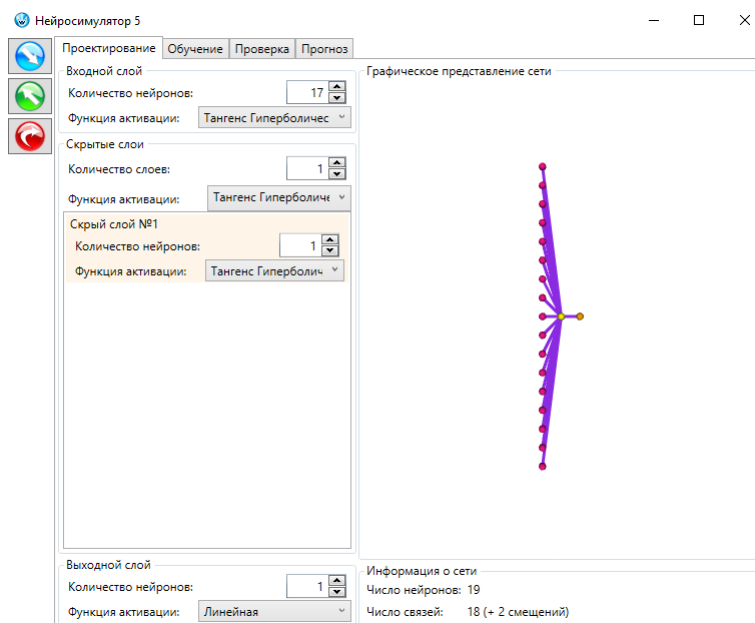
Рисунок 1. Часть обучающего множества

X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12	X13	X14	X15	X16	X17	
Год прои: США	США	Европа	Азия	Другие страны	Детектив	Драма	Фэнтези	Боевик	Триллер	Комедия	Криминал	Биография	Приключен	Фантастика	Рейтинг IMDb	Номинация на Оскар	
1999	100	0	0	0	0	100	0	0	0	0	0	80	0	0	0	90	100
2015	90	0	70	0	0	90	0	70	0	0	0	80	80	0	0	90	100
2018	100	0	0	0	0	0	0	0	0	40	0	80	0	0	0	85	100
2010	100	0	0	0	0	100	0	0	0	0	0	0	0	0	0	60	0
2013	100	0	0	0	90	0	0	70	0	0	0	0	0	0	0	90	100
2000	90	80	0	0	0	90	0	0	0	0	0	0	80	0	0	90	100
2005	0	0	100	0	0	100	0	0	0	0	0	0	0	0	0	60	0
2009	100	0	0	0	80	90	0	0	80	0	0	0	0	0	0	80	0
2012	100	0	0	0	0	100	0	0	0	0	0	0	0	0	0	60	0
2003	90	0	0	80	0	70	80	0	0	0	0	0	0	90	0	90	100
2001	100	0	0	0	0	90	0	0	0	0	0	80	0	0	0	90	100
1997	0	100	0	0	0	100	0	0	0	80	0	0	0	0	0	80	0

Рисунок 2. Тестирующее множества

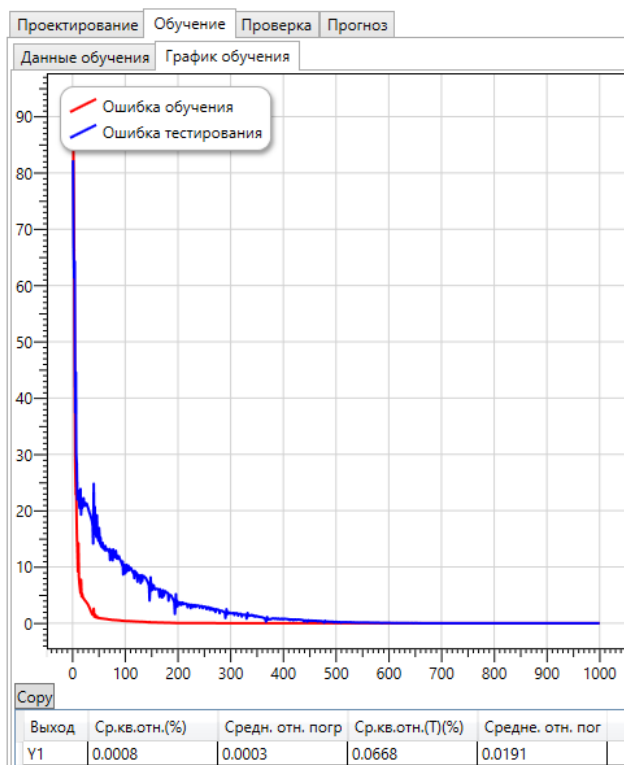
Обучающее множество было собрано вручную. Перед переходом к проектированию и обучению нейросетевой модели, была выполнена очистка исходного множества от противоречивых примеров, выбросов и дубликатов.

Проектирование, обучение, тестирование нейронной сети выполнялись с помощью программы «Нейросимулятор 5» [5] по методике [6]. В результате оптимизации спроектированная нейронная сеть представляла собой персептрон, который имеет семнадцать входных нейронов, один выходной и один скрытый слой. Персептрон представлен на рисунке 3.



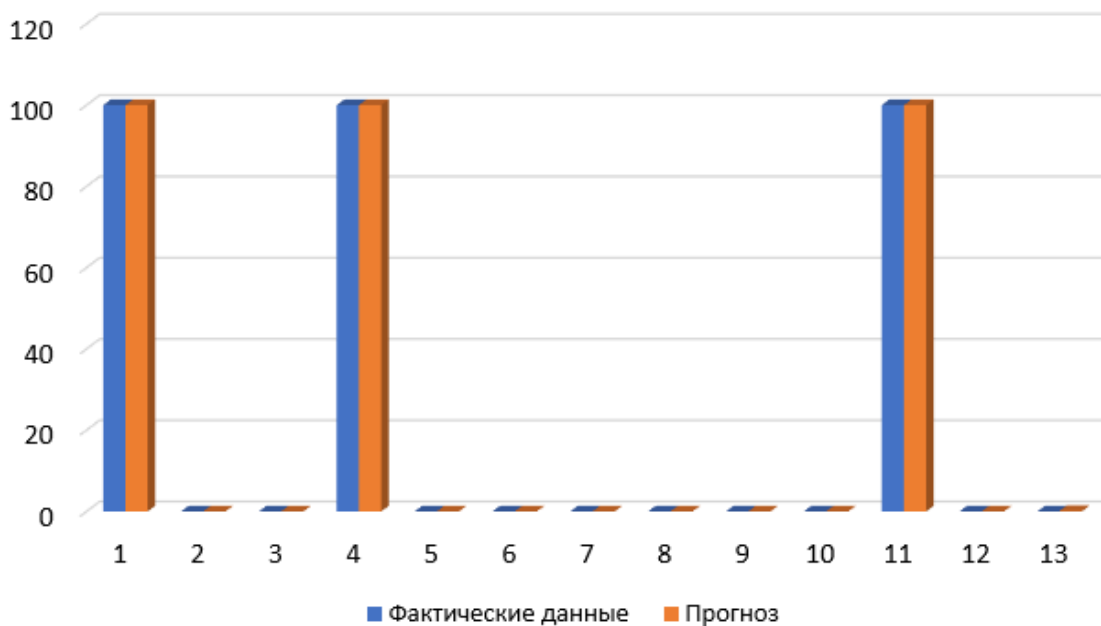
**Рисунок 3. Перцептрон**

После обучения прогнозирующие свойства нейронной сети проверялись на примерах тестирующего множества. Среднеквадратичная ошибка тестирования составила 0,06%:



**Рисунок 4. График обучения нейронной сети**

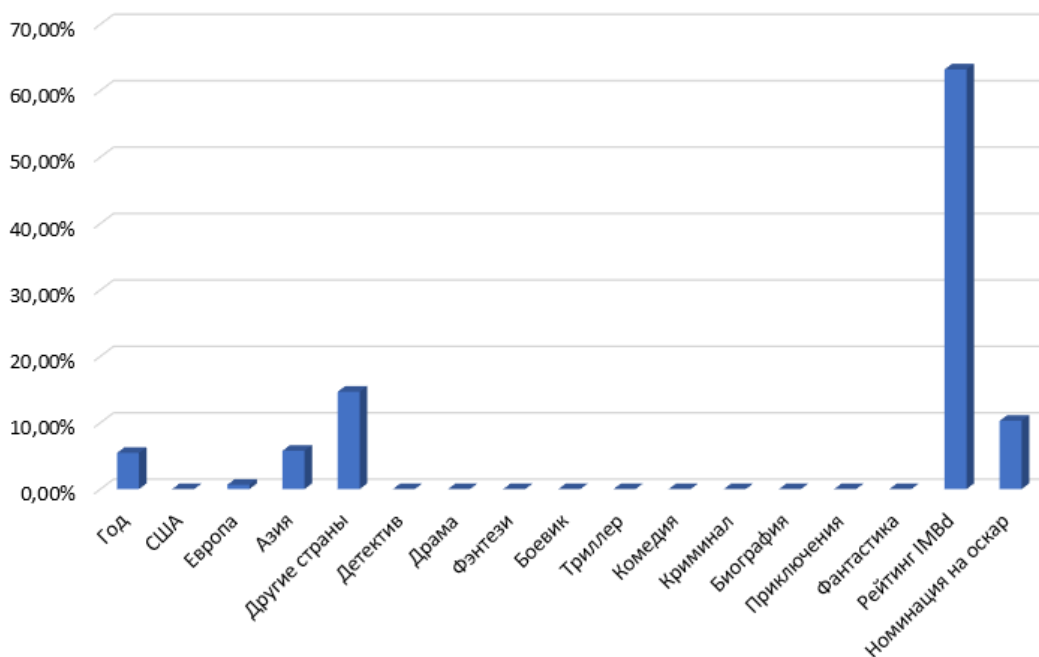
Для оценки корректной работы спроектированной нейронной сети использовалось тестирующее множество, состоящее из 13 примеров. На рисунке 5 представлена гистограмма, демонстрирующая разницу между фактическим и прогнозируемым нейросетью победителями кинопремии Оскар.



**Рисунок 5. Результат тестирования нейронной сети**

Из результатов, изображенных на рисунке 5, можно сделать вывод об адекватной работе спроектированной нейронной сети.

Оценка значимости входных параметров была выполнена с помощью программы «Нейросимулятор 5», результат отображен на рисунке 6



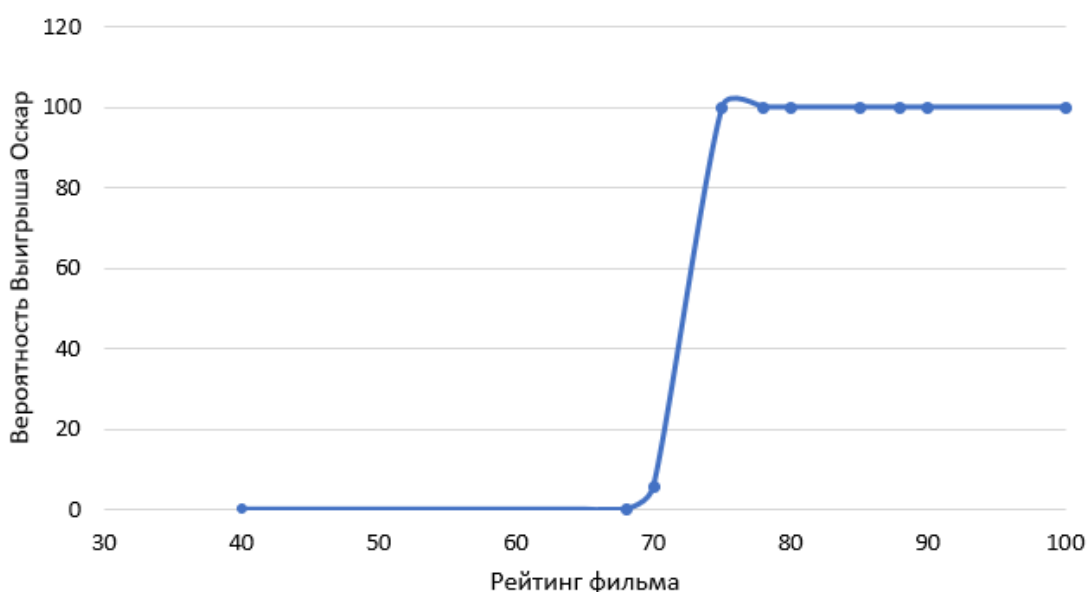
**Рисунок 6. Значимость входных параметров нейронной сети**

Как видно из рисунка 6, самым значимым параметром является процент рейтинга IMDb, также значимыми являются факторы номинации на Оскар, страна производства фильма и год выпуска фильма.

Применение методов нейросетевого моделирования позволяет не только выполнять прогнозы, но и ставить над моделями эксперименты, пытаться

изменять эти прогнозы в наиболее благоприятную для успеха кинобизнеса сторону. Так, варьируя входные параметры обученной нейронной сети и производя вычисления, можно сформировать конкретный список рекомендаций, способствующих увеличению шансов выиграть Оскар.

В качестве примера возьмем фильм «Паразиты». Это азиатский фильм 2019 года производства, который покорила миллионы зрителей, имеет высокий рейтинг и соответственно статуэтку кинопремии Оскар. На примере данного фильма рассмотрим зависимость между рейтингом кинофильма и выигрышем Оскара. Для того, чтобы исследовать данную зависимость все входные параметры кроме параметра «Рейтинг» были заморожены. После этого было произведено изменение самого параметра «Рейтинг». Его показатели были изменены от наименьшего к наибольшему, как и представлено на рисунке 7.



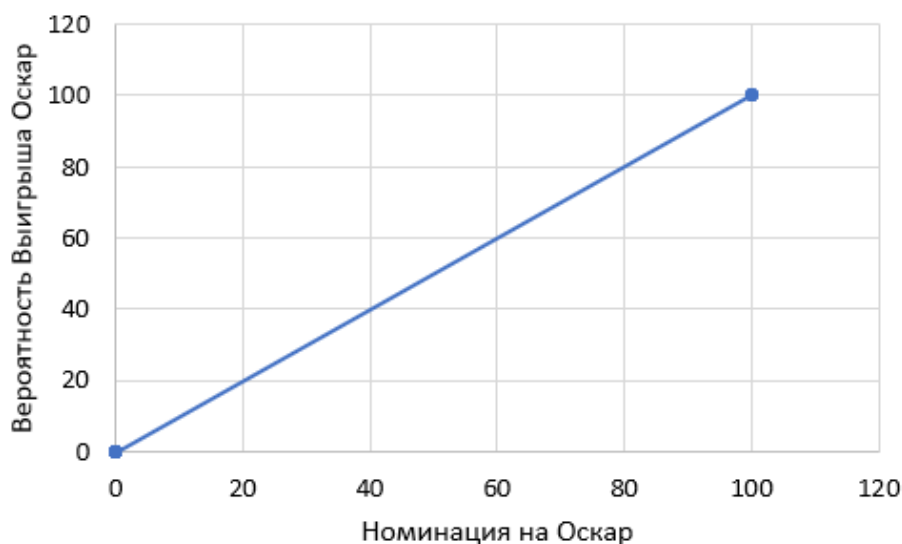
**Рисунок 7. Зависимость рейтинга фильма от вероятности выигрыша Оскар**

Исходя из рисунка 7 можно сделать вывод, что связь между рейтингом кинофильма и его возможностью выиграть Оскар прямая. В целом это подтверждается и самой моделью, построенной в нейросимуляторе, где параметр рейтинга был самым важным среди других входных.

Другим примером является фильм «Земля Кочевников». Фильм созданный совместно США и Германией 2021 года производства. Имеет относительно высокий рейтинг, что является немаловажным фактором, т. к. как правило, Оскар получают картины именно с высоким рейтингом. Несмотря на это данный фильм получил Оскар. Аналогичным образом, заморозив все входные параметры кроме параметра «Номинация на Оскар» исследуем зависимость между изменением данных по параметру «Номинация на Оскар» и вероятностью выиграть сам Оскар. Результаты представлены на рисунке ниже.

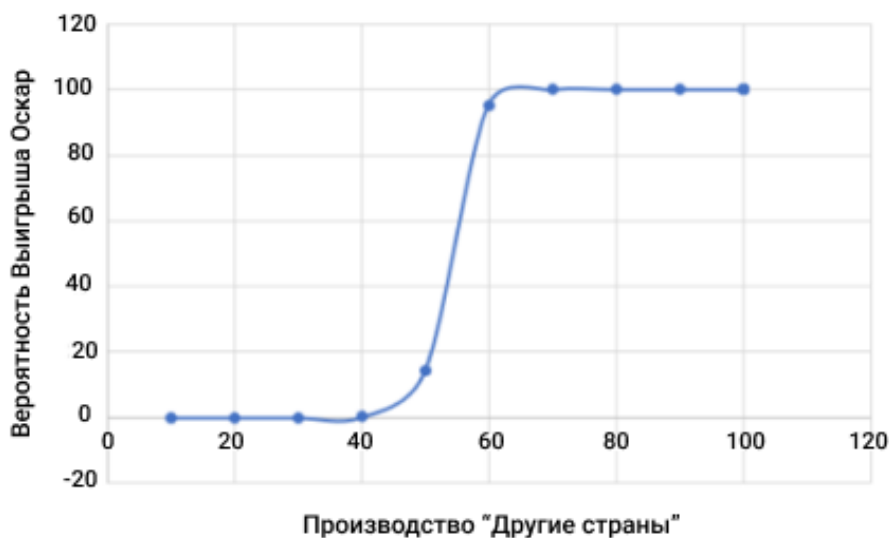
Исходя из рисунка видна прямая зависимость между двумя данными параметрами. Однако также стоит отметить, что иногда бывают случаи, когда фильм при наличии номинации на премию не выигрывает саму награду. Случается это в силу других его параметров, например, не очень высокий рейтинг.





**Рисунок 8. Зависимость параметра «Номинация на Оскар» и «Вероятность выигрыша Оскара»**

Еще одним примером является фильм «Несломленный». Это фильм в жанрах «Военный», «Драма», «История», 2014 года производства. В целом был тепло встречен критиками, и даже был номинирован на Оскар. Однако саму премию не получил. Одной из причин как раз может быть вышеупомянутый рейтинг, ведь у данного фильма он действительно не слишком высок. С помощью данного фильма попробуем найти зависимость между потенциальным выигрышем Оскара и страной производства. В силу того, что параметр «Другие страны» имеет относительно высокое значение попробуем проверить получил бы «Несломленный» Оскар если бы в его создании участвовали не только США. Заморозив все входные параметры кроме параметра «Другие страны» исследуем зависимость между изменением данных по параметру «Другие страны» и вероятностью выиграть сам Оскар. Результаты представлены на рисунке ниже.



**Рисунок 9. Зависимость параметра «Производство «Другие страны»» и «Вероятность выигрыша Оскар»**

Исходя из рисунка 9 можно сделать вывод, что связь между страной производства кинофильма и его возможностью выиграть Оскар имеется.

Полученные результаты исследований не противоречат реальности, что подтверждает, что спроектированную нейронную сеть можно считать пригодной для прогнозирования выигрыша кинопремии Оскар.

**Заключение.** Таким образом, в ходе данной работы была разработана нейросетевая модель, предназначенная для прогнозирования получения премии Оскар. Модель может быть использована в кинобизнесе для поддержки принятия решений на этапе планирования будущих фильмов с целью получения наибольшего успеха от их проката.

### Библиографический список

1. Кзянева Т.Г., Иванова Д.М. Прогнозирование кассовых сборов проката фильма // Омский государственный университет им Ф. М. Достоевского. 2020. №2 (10). URL: <https://cyberleninka.ru/article/n/prognozirovanie-kassovyh-sborov-prokata-filma/viewer>

2. Ясницкий Л. Н., Белобородова Н. О., Медведева Е. Ю. Методика нейросетевого прогнозирования кассовых сборов кинофильмов // Финансовая аналитика: проблемы и решения. 2017. т. 10. Вып. 4. С. 449–463.

3. Педяш Е.А. Эконометрическое прогнозирование кассовых сборов кинофильмов/ НИУ ВШЭ-Пермь. – Москва, 2013. – 50 с. – Библиогр.: с. 42–50.

4. Портал об индустрии кино. [Электронный ресурс] URL: <https://www.kinopoisk.ru/lists/movies/oscar-best-film/>

5. Черепанов Ф. М., Ясницкий Л.Н. Нейросимулятор 4.0. Свидетельство о регистрации программы для ЭВМ RUS 2014612546. Заявка № 2014610341 от 15.01.2014.

6. Ясницкий Л.Н. Введение в искусственный интеллект. М.: Издательский центр «Академия», 2005. – 176 с.

## NEURAL NETWORK SYSTEM FOR OSCAR PREDICTION

*Veveritsa Karina E.*

National Research University Higher School of Economics,  
614990, Russia, Perm, Str. Studencheskaya, 38, infoperm@hse.ru

The article presents a description of the development of a neural network system for predicting the obtainment of an Oscar. The system makes it possible to predict with great confidence the winner of the Oscar, the premier film award in the United States, based on movie data. With the help of a study of the developed intellectual system, carried out within the subject area, observations of practical significance were identified.

**Key words:** neural network, artificial intelligence, forecast, movies, Oscar, award, perceptron, test task.

**ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ  
В АСТРОНОМИИ, МЕТЕОРОЛОГИИ,  
КАТАСТРОФАХ**

## НЕЙРОСЕТЕВАЯ СИСТЕМА ПРОГНОЗИРОВАНИЯ НЕОБХОДИМОСТИ ТУШЕНИЯ ЛЕСНЫХ ПОЖАРОВ

*Соскин Андрей Иванович*

Пермский государственный национальный исследовательский университет,  
КМБ. 614990, Россия, г. Пермь, ул. Букирева, 15, soskin\_andrey@mail.ru

В статье представлено описание разработки нейросетевой системы для прогнозирования необходимости тушения лесных пожаров. Система позволяет определить, какие пожары будут распространяться и приносить большой вред природной среде, а какие нет.

**Ключевые слова:** искусственный интеллект, нейросетевые технологии, прогнозирование, лесные пожары, природа, экология.

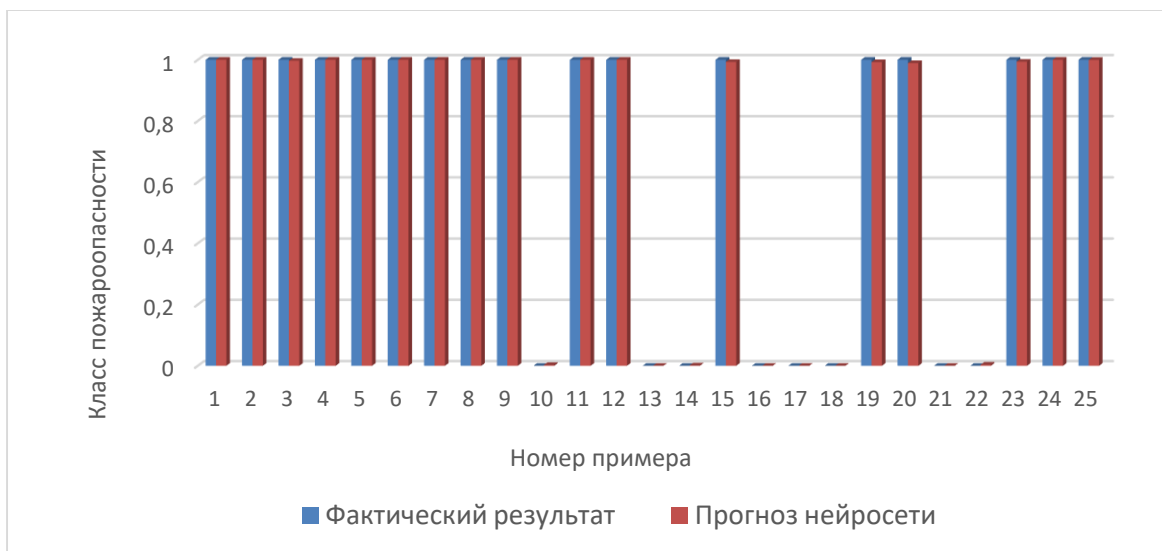
**Введение.** Лесные пожары, безусловно, одно из худших бедствий именно для природной среды. В некоторых регионах планеты, где возгорания происходят регулярно, возникает достаточно большая необходимость прогнозировать то, какие пожары тушить стоит, а какие нет, ведь ресурсы и специалисты в этой области ограничены. Данная нейросеть, как и знание о том, какие пожары подлежат тушению, будут полезны пожарным службам тех областей планеты, в которых возгорания происходят достаточно часто.

Для создания нейросетевой системы были выбраны следующие параметры: X1 – день возгорания, X2 – месяц, X3 – год, X4 – температура воздуха в градусах Цельсия, X5 – оценка опасности от уровня относительной влажности, X6 – скорость ветра, X7 – количество осадков за весь день, X8 – FFMС индекс из FWI системы, X9 – DMC индекс из FWI системы, X10 – DC индекс из FWI системы, X11 – ISI индекс из FWI системы, X12 – BUI индекс из FWI системы и X13 – FWI индекс. В качестве выходного параметра D1 получается значение – класс пожароопасности fire, или not fire – необходимо тушить пожар, или же нет.

Обучающее множество было собрано с сайта UCI Machine Learning Repository [3]. Перед переходом к проектированию и обучению нейросетевой модели, была выполнена очистка исходного множества от некорректных примеров. Такими примерами считались те, где были пропущены некоторые данные, столбцы таблицы. Объем итогового множества включает в себя 189 примеров. Данное множество было разделено на обучающее и тестирующее в соотношении 86% и 14% соответственно. Собранные данные охватывают период с июня по сентябрь 2012 года.

Проектирование, обучение, тестирование нейронной сети выполнялись с помощью программы «Нейросимулятор 5» [1] по методике [2]. В результате оптимизации спроектированная нейронная сеть представляла собой персептрон, который имеет 13 входных нейронов и один выходной.

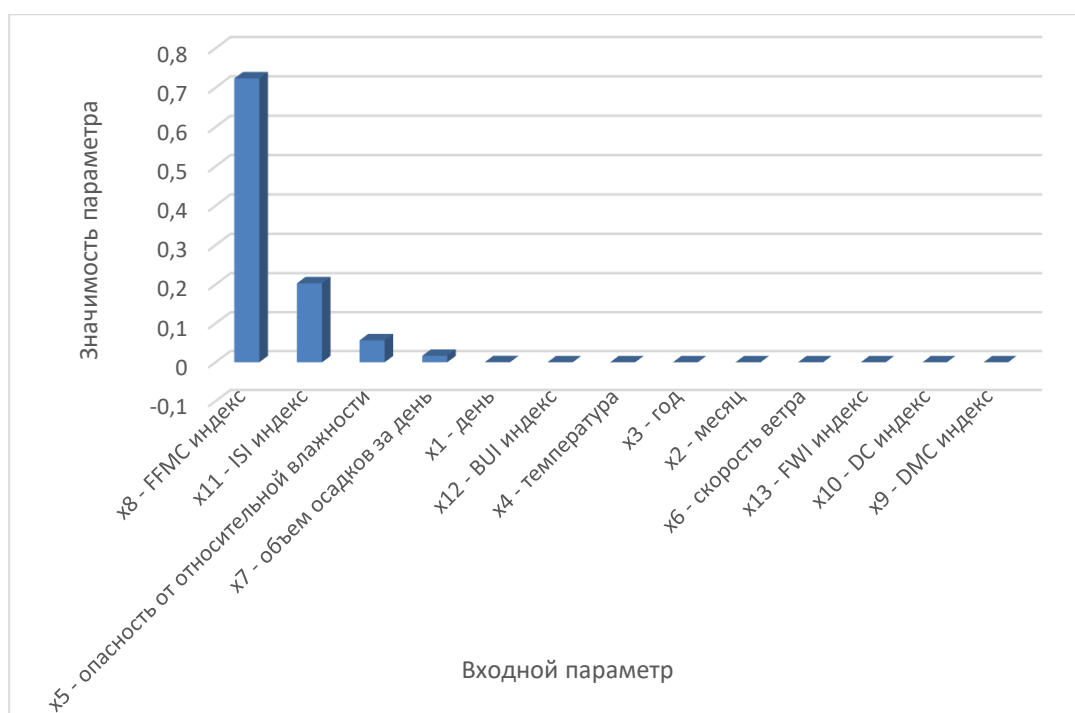
Для оценки корректной работы спроектированной нейронной сети использовалось тестирующее множество, состоящее из 26 примеров. Средняя относительная ошибка тестирования составила 3,3%, что можно считать очень неплохим результатом. На рисунке 1 представлена гистограмма, демонстрирующая разницу между фактической и прогнозируемой нейросетью необходимостью тушения 25 пожаров из тестирующего множества.



**Рисунок 1. Результат тестирования нейронной сети**

Из результатов, изображенных на рисунке 1, можно сделать вывод об адекватной работе спроектированной нейронной сети.

Оценка значимости входных параметров была выполнена с помощью программы «Нейросимулятор 5», результат отображен на рисунке 2.



**Рисунок 2. Значимость входных параметров нейросетевой модели**

Как видно из рисунка 2, наиболее значимыми параметрами являются FFMC индекс, ISI индекс и оценка опасности от уровня относительной влажности.

Далее было проведено исследование влияния FFMC индекса. Исследование производилось с помощью метода «замораживания» [2], суть которого заключается в варьировании значения одного параметра и фиксировании значений всех других.

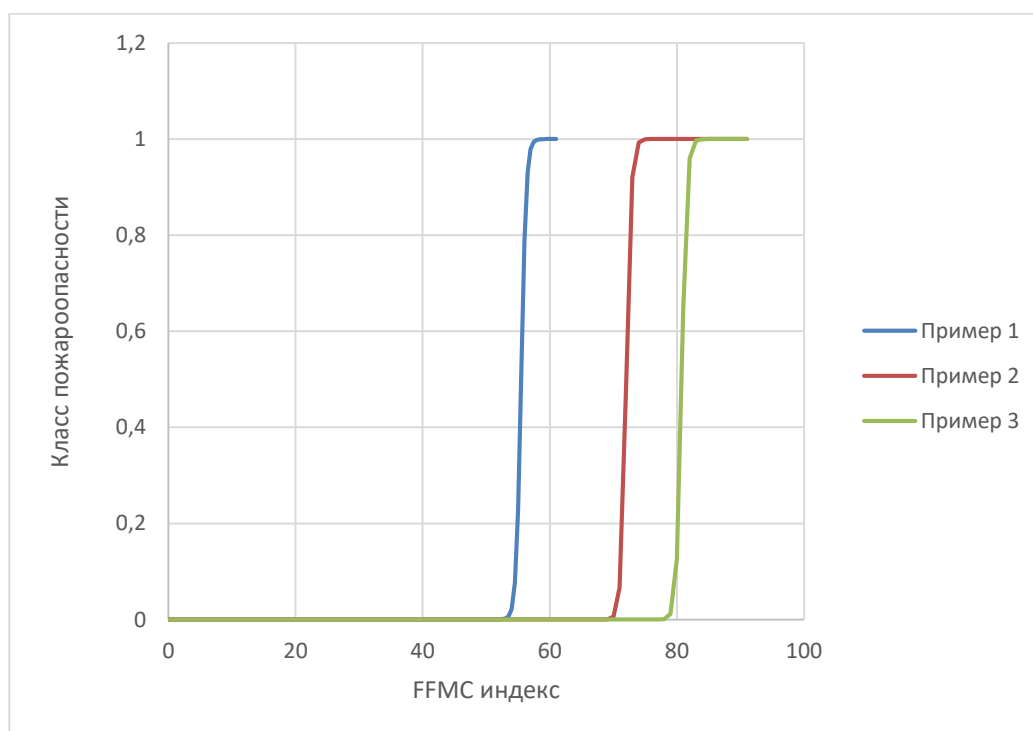
Для этой цели были отобраны 3 примера, информация о параметрах которых указана в таблице 1.

**Таблица 1**

**Характеристики пожаров, выбранных для исследования**

X1	X2	X3	X4	X5	X6	X7	X9	X10	X11	X12	X13
18	8	2012	37	37	14	0	35.9	86.8	16	35.9	26.3
21	8	2012	36	71	15	0	36.9	117.1	5.1	41.3	12.2
28	8	2012	35	56	14	0.4	37	166	2.1	30.6	6.1

На рисунке 3 показан график зависимости необходимости тушения лесного пожара от величины FFMC индекса. Исходя из графика можно сказать, что увеличение FFMC индекса напрямую определяет класс пожароопасности.



**Рисунок 3. Зависимость необходимости тушения пожара от величины FFMC индекса**

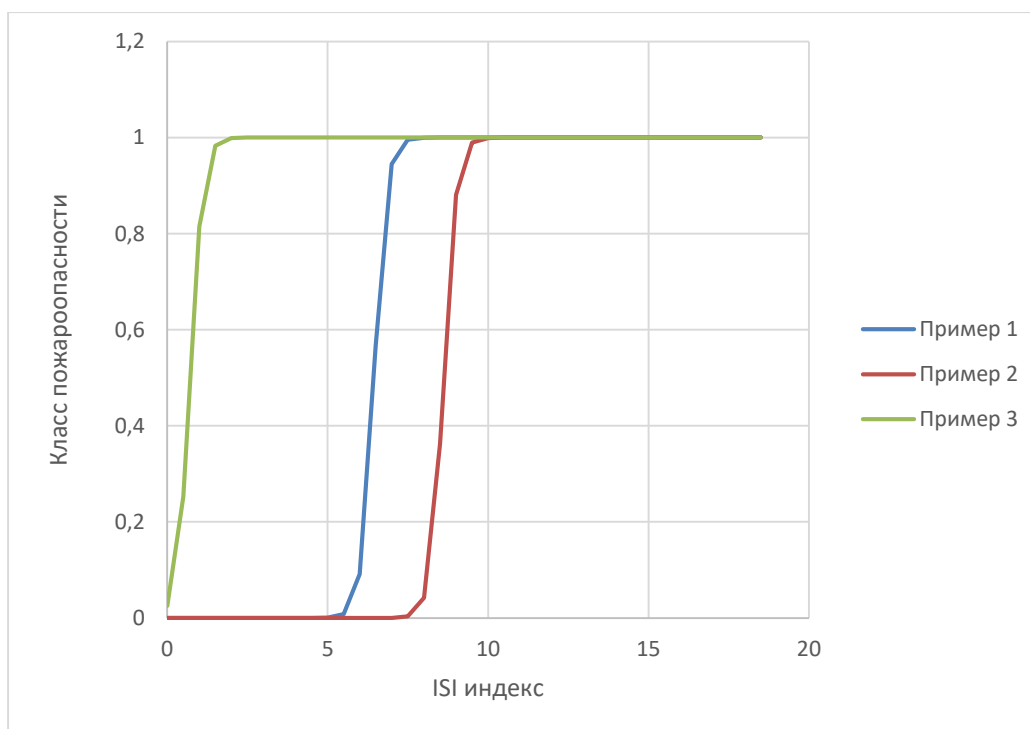
Далее было проведено исследование влияния ISI индекса на необходимость тушения лесного пожара. Для этого были отобраны 3 примера, информация о параметрах которых указана в таблице 2.

**Таблица 2**

**Характеристики пожаров, выбранных для исследования**

X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X12	X134
4	9	2012	30	66	15	0.2	73.5	4.1	26.6	6	0.7
6	9	2012	34	71	14	6.5	64.5	3.3	9.1	6.5	0.4
12	9	2012	31	72	14	0	84.2	8.3	25.2	9.1	3.9

На рисунке 4 продемонстрирована зависимость необходимости тушения лесного пожара от величины ISI индекса. Из графика следует, что величина ISI индекса увеличивает вероятность того, что пожар нужно будет тушить.



**Рисунок 4. Зависимость класса пожароопасности от величины ISI индекса**

Далее было проведено исследование влияния оценки опасности по уровню относительной влажности воздуха на необходимость тушить лесной пожар. Для этого были отобраны 3 примера, информация о параметрах которых показана в таблице 3.

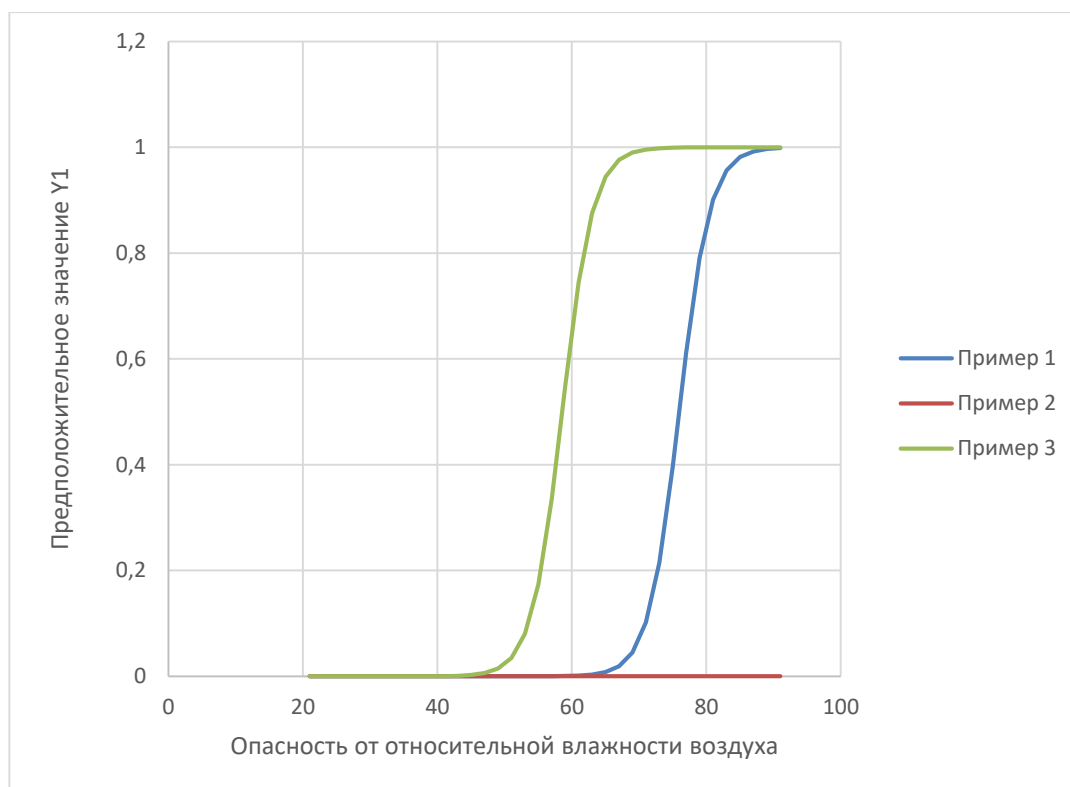
**Таблица 3**

**Характеристики пожаров, выбранных для исследования**

X1	X2	X3	X4	X6	X7	X8	X9	X10	X11	X12	X13
17	8	2012	42	9	0	50	30.3	76.4	15.7	30.4	24
4	9	2012	30	15	0.2	73.5	4.1	26.6	1.5	6	0.7
7	9	2012	31	15	0	80	5.8	17.7	3.8	6.4	3.2

На рисунке 5 показан графики зависимости класса пожароопасности от величины оценки опасности по уровню относительной влажности воздуха. Как следует из графика, этот параметр увеличивает вероятность того, что

класс пожароопасности будет “fire”, но не для всех примеров. Объясняется это тем, что значимость этого параметра не так велика, как у предыдущих.



**Рисунок 5. Зависимость класса пожароопасности от величины оценки опасности по уровню относительной влажности воздуха**

Полученные результаты исследований не противоречат реальности, что подтверждает, что спроектированную нейронную сеть можно считать пригодной для прогнозирования необходимости тушения лесных пожаров.

**Заключение.** Построена система нейросетевого прогнозирования определения класса пожароопасности лесных пожаров и необходимости их тушения.

Спроектированная нейросетевая модель учитывает 13 параметров: день возгорания, месяц, год, температура воздуха в градусах Цельсия, индекс опасности от уровня относительной влажности, скорость ветра, количество осадков за весь день, FFMC индекс из FWI системы, DMC индекс из FWI системы, DC индекс из FWI системы, ISI индекс из FWI системы, BUI индекс из FWI системы и FWI индекс.

### **Библиографический список**

1. Черепанов Ф.М., Ясницкий Л.Н. Нейросимулятор 4.0. Свидетельство о регистрации программы для ЭВМ RUS 2014612546. Заявка № 2014610341 от 15.01.2014.

2. Ясницкий Л.Н. Интеллектуальные системы: учебник. М.: Лаборатория знаний, 2016. 221 с.



3. UCI Machine Learning Repository – открытый электронный ресурс с базами данных. Режим доступа: <https://archive.ics.uci.edu/ml/datasets/Algerian+Forest+Fires+Dataset++#>

## NEURAL NETWORK SYSTEM FOR FORECASTING THE NEED FOR EXTINGUISHING FOREST FIRES

*Soskin Andrey I.*

Perm State University

Str. Bukireva, 15, Perm, Russia, 614990, soskin\_andrey@mail.ru

The article describes the development of a neural network system for predicting the need to extinguish forest fires. The system allows you to determine which fires will spread and cause great harm to the natural environment, and which will not.

**Keywords:** artificial intelligence, neural network technologies, forecasting, forest fires, nature, ecology.

УДК 004.032.26

## НЕЙРОСЕТЕВАЯ СИСТЕМА ПРОГНОЗИРОВАНИЯ ПОКАЗАТЕЛЕЙ СОЛНЕЧНОЙ ОСВЕЩЕННОСТИ

*Карибова Айнура Сулеймановна*

Национальный исследовательский университет «Высшая школа экономики»,  
614060, Россия, г. Пермь, ул. Бульвар Гагарина, 37,  
askaribova@edu.hse.ru

В статье представлено описание разработки нейросетевой системы прогноза значений прямой нормальной солнечной радиации (DNI) или лучевого излучения, рассеянной горизонтальной солнечной радиации (DHI) или диффузного излучения неба и суммарной горизонтальной солнечной радиации (GHI). Система позволяет прогнозировать интенсивность солнечного излучения в течение нескольких лет. Был проведен ряд исследований, результаты которых приводятся в данной статье.

**Ключевые слова:** солнечное излучение, типы измеряемой солнечной освещенности, прогнозирование выработки солнечной энергии, нейронная сеть, экология, устойчивое развитие, солнечное воздействие, солнечная инсоляция, DHI, DNI, GHI.

**Введение.** Популяризация экологичных источников энергии и растущее производство солнечной энергии создают проблемы для работы электрических сетей [1], [2], которые необходимо сбалансировать в режиме реального

времени [3]. Интенсивность солнечного излучения у земной поверхности сильно варьируется из-за атмосферных процессов, времени суток, года, местоположения и погодных условий. Изменение интенсивности излучения у земной поверхности влияет на стабильность выработки солнечной энергии [4], что, в свою очередь, ставит под угрозу стабильность и затраты на интеграцию для сетей с высоким проникновением солнечной энергии [2]. Точные прогнозы солнечной радиации и мощности позволяют использовать технологии, которые потенциально могут снизить неопределенность в производстве солнечной энергии и оптимизировать спрос и решения для хранения [1], [2], [3].

Важно понимать разницу между этими видами излучения. Прямая солнечная радиация – это количество солнечного излучения, полученного на единицу площади поверхностью, перпендикулярно (или нормально) расположенной к лучам, идущим по прямой линии от направления солнца. Рассеянная солнечная радиация (DNI) — это количество излучения, полученного на единицу площади поверхностью (без теней), которое идет не напрямую от солнца, а рассеивается молекулами и частицами в атмосфере и приходит одинаково со всех сторон. Суммарная солнечная радиация (GHI) — это общее количество коротковолнового излучения, полученного поверхностью, расположенной горизонтально по отношению к земле. Это значение представляет особый интерес для фотоэлектрических установок и включает в себя как DNI, так и DHI [1].

Эффективные методы прогнозирования солнечной активности были разработаны для различных временных горизонтов от нескольких минут до нескольких дней. Обычно используемые методы включают регрессивные или стохастические модели обучения [4] и физические модели, основанные на методах дистанционного или локального зондирования [3]. Для прогнозов в последние годы были разработаны усовершенствованные гибридные модели, которые объединяют методы стохастического обучения и локального зондирования [2]. При оценке в режиме реального времени гибридные модели достигают навыков прогнозирования в диапазоне от 6% до 32% по сравнению с эталонными моделями постоянства [1].

Целью данного исследования было построение прогностической модели, предсказывающей значения прямой нормальной солнечной радиации (DNI) или лучевого излучения, рассеянной горизонтальной солнечной радиации (DHI) или диффузного излучения неба и суммарной горизонтальной солнечной радиации (GHI). Для этого необходимо было собрать множество данных о факторах, влияющих на эти виды излучений, а также создать и обучить нейросетевую модель на этих данных. Конечным результатом должна быть нейросетевая система, способная прогнозировать значения DHI, DNI, GHI на основании входных параметров.

В результате были выделены следующие параметры:

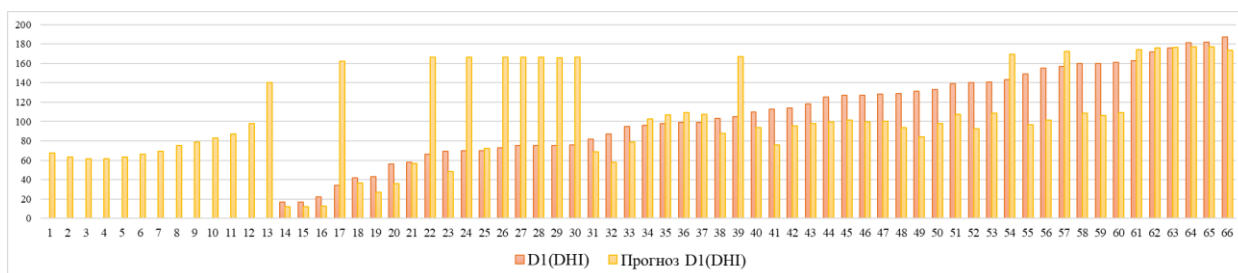
- X1 – Год,
- X2 – Месяц,
- X3 – День,
- X4 – Час,

- X5 – Минута,
- X6 – Температура, °С,
- X7 – Тип облака (0 Чистое, 1 Вероятно ясно, 2 Туман, 3 Водяное, 4 Холодное водяное, 5 Смешанное, 6 Ледяное, 7 Перистое, 8 Кучевое, 9 Пролетающее облако, 10 Неизвестно, 11 Пылевое, 12 Облако дыма, 15 Нет данных),
- X8 – Точка росы (температура), °С,
- X9 – Относительная влажность, %,
- X10 – Солнечный зенитный угол, градусы для расчета  $\cos(\theta)$ ,
- X11 – Давление, мбар,
- X12 – Осадки, см,
- X13 – Направление ветра, градусы,
- X14 – Скорость ветра, м/с.

Итоговое множество для нейросети содержит разнообразные данные. Перед переходом к проектированию и обучению нейросетевой модели, была выполнена очистка исходного множества от противоречивых примеров, выбросов, дубликатов. Таким образом, объем итогового множества включает в себя 1961 примеров. Данное множество было разделено на обучающее и тестирующее. Данные были предоставлены Wipro, которые совместно с MachineHack разработали задачу прогнозирования для оптимизации выработки солнечной энергии с использованием моделей машинного обучения. Компания по производству солнечной энергии хочет оптимизировать производство солнечной энергии и нуждается в модели прогнозирования «Clearsky DNI», «Clearsky DNI», «Clearsky GHI» [5].

Проектирование, обучение, тестирование нейронной сети выполнялись с помощью программы «Нейросимулятор 5» [6] по методике [7]. В результате оптимизации спроектированная нейронная сеть представляла собой персептрон, который имеет 14 входных нейронов, один выходной и один скрытый слой с семью нейронами.

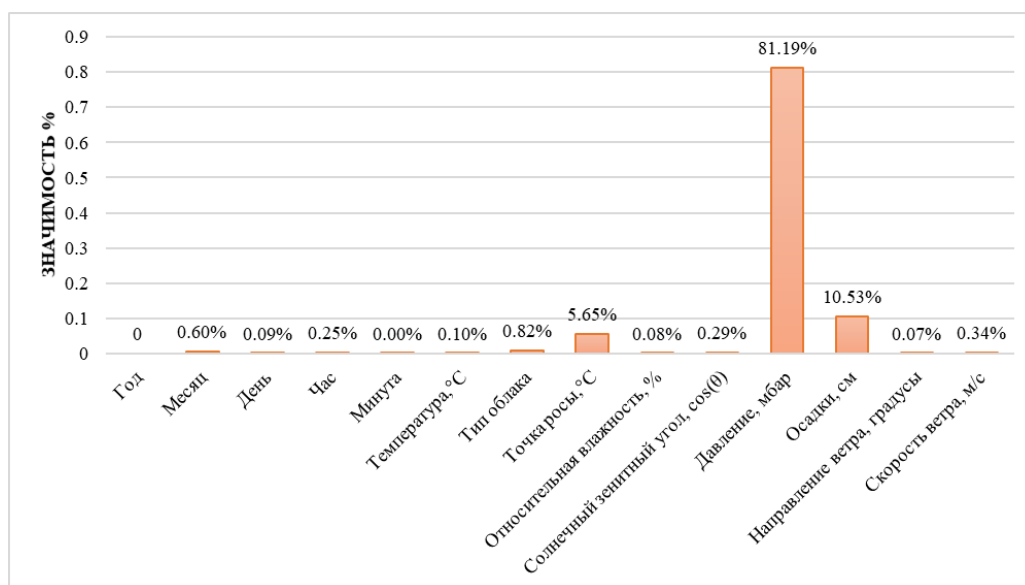
Для оценки корректной работы спроектированной нейронной сети использовалось тестирующее множество, состоящее из 104 примеров. Средние относительные ошибки тестирования – 4.26%, 2.44%, 1.51% для D1, D2, D3 соответственно, что можно считать хорошим результатом. На рисунке 1 представлена гистограмма, демонстрирующая разницу между фактическим и прогнозируемым нейросетью результатом значений прямого диффузного излучения неба (DNI).



**Рисунок 1. Результат тестирования нейросети**

Из результатов, изображенных на рисунке 1, можно сделать вывод об адекватной работе спроектированной нейронной сети. Тестирование работы сети на прогнозирование остальных выходных параметров, а именно DNI, GHI показали результаты еще лучше.

Оценка значимости входных параметров была выполнена с помощью программы «Нейросимулятор 5». На рисунке 2 представлена гистограмма, показывающая распределение значимости параметров.



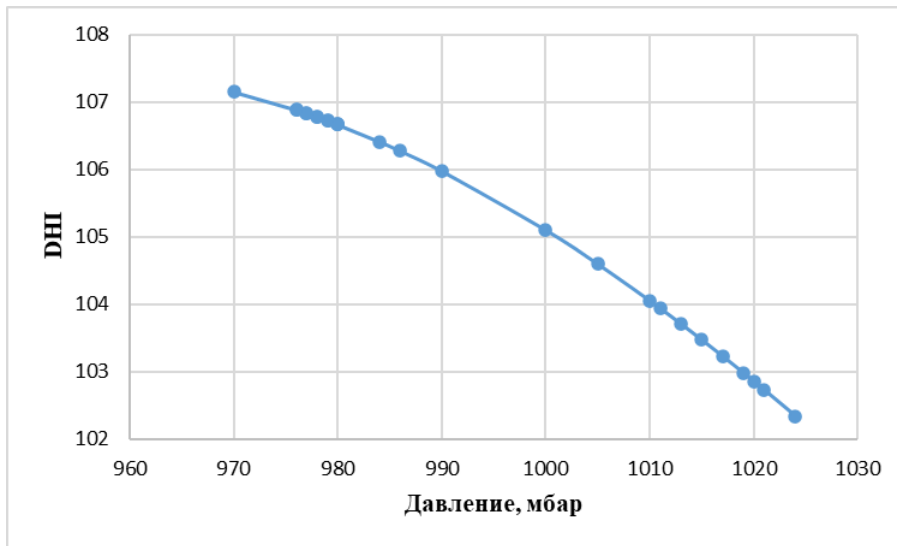
**Рисунок 2. Значимость входных параметров нейросетевой модели**

Как видно из рисунка 2, самыми значимыми входными параметрами оказались соответственно: атмосферное давление (уровень значимости – 81.19%), уровень осадков (уровень значимости – 10.53%), точка росы (уровень значимости – 5.65%). Полностью неэффективные и незначимые параметры: год, месяц, день, час, минута, относительная влажность, температура, тип облака, направление и скорость ветра.

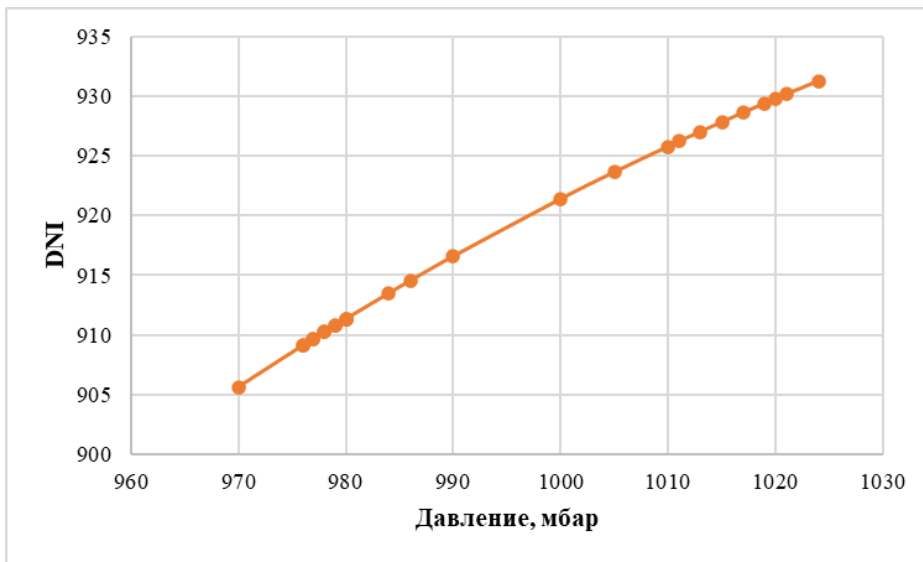
Далее было проведено исследование полученных зависимостей между входными параметрами и значениями DNI, DNI, GHI. Исследование производилось с помощью метода «замораживания», суть которого заключается в варьировании значения одного параметра и фиксировании значений всех других параметров. Для этой цели был отобран «нейтральный пример».

На рисунках 3, 4, 5 показаны графики зависимости DNI, DNI, GHI от уровня атмосферного давления. В том случае, когда давление растет, значение DNI (прямой солнечной радиации) падает, а значения DNI (рассеянное солнечной радиации), GHI (суммарной солнечной радиации) возрастают [8].

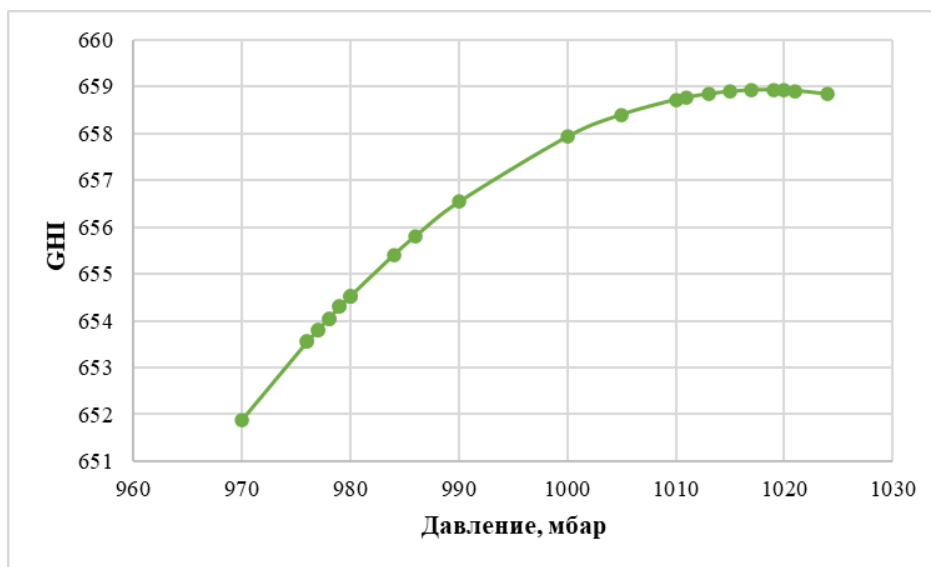
На рисунке 6 продемонстрирован график зависимости DNI, DNI, GHI от количества осадков. Выводы на основании графика, следующие: чем выше осадки, тем меньше значения показателей рассеянного излучения (DNI) и суммарного излучения (GHI). Прямая солнечная радиация (DNI), наоборот, увеличивается.



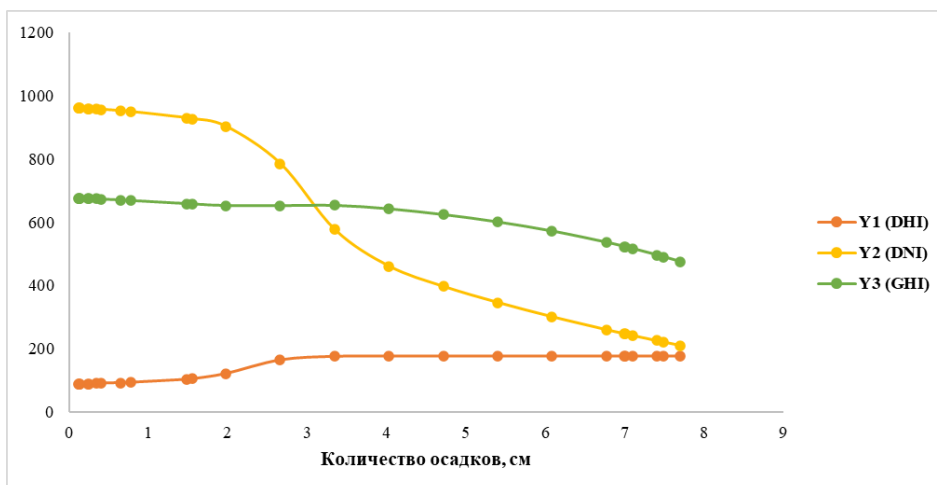
**Рисунок 3. Зависимость DNI от уровня атмосферного давления**



**Рисунок 4. Зависимость DNI от уровня атмосферного давления**

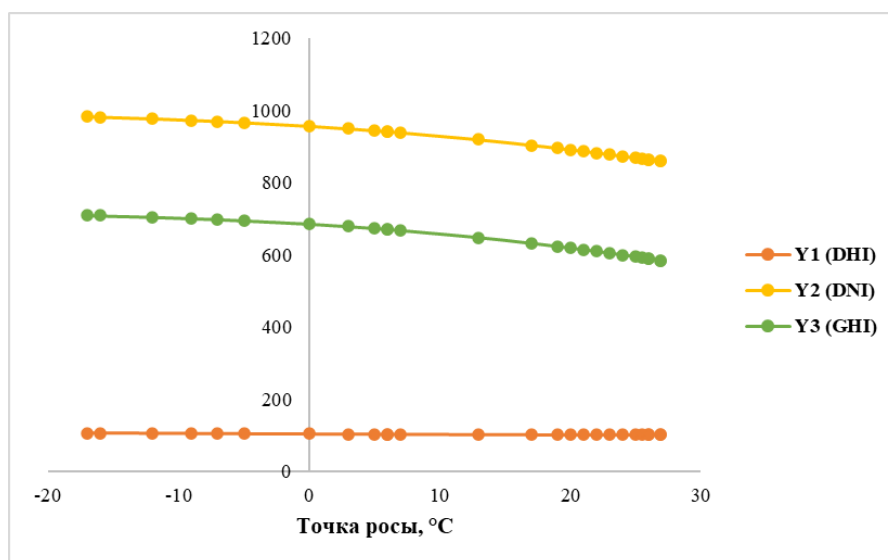


**Рисунок 5. Зависимость GHI от уровня атмосферного давления**



**Рисунок 6. Зависимость DHI, DNI, GHI от количества осадков**

На рисунке 7 изображен график зависимости DHI, DNI, GHI от значения точки росы. Как можно заметить показатели незначительно меняются, наблюдается тренд: при повышении значения точки росы, GHI, DNI становятся меньше, а DHI остается примерно на одном и том же уровне.



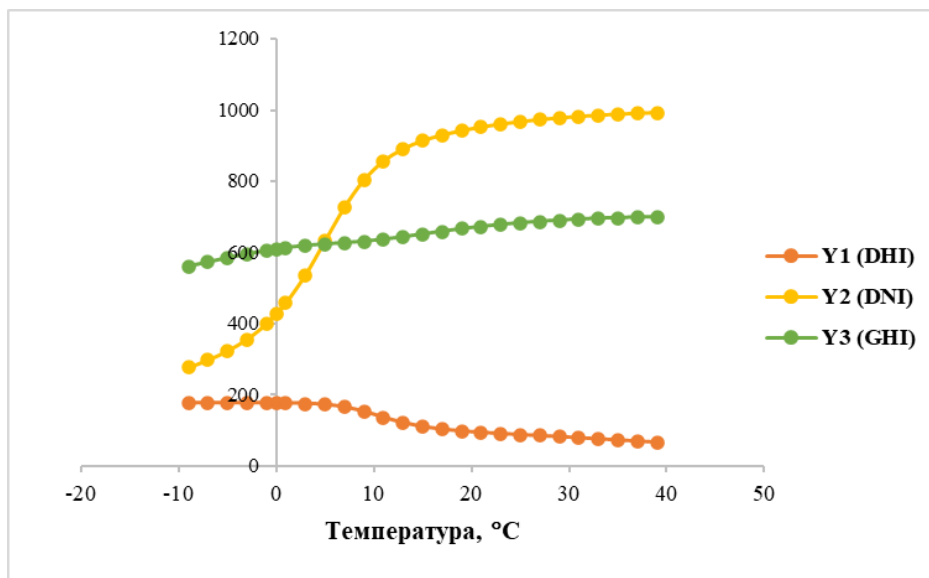
**Рисунок 7. Влияние параметра точка росы на DHI, DNI, GHI**

На рисунке 8 продемонстрирована зависимость DHI, DNI, GHI от температуры. График показывает тенденцию понижения значения DNI при повышении температуры. DHI, наоборот, растет, а GHI незначительно меняется, с небольшим повышением.

Полученные результаты исследований не противоречат реальности, что подтверждает, что спроектированную нейронную сеть можно считать пригодной для прогнозирования значений DHI, DNI, GHI.

Далее построение нейросетевой модели происходило в интерактивном облачном сервисе блокноте Google Colab на языке программирования Python с помощью библиотеки Keras, однако рассматривался только 1 выходной параметр (DHI). Были получены похожие результаты исследования для первого

выходного параметра, что и в нейросимуляторе. Однако, прогнозирование на нейросимуляторе имеет преимущества в виде некоторых встроенных функций, простого и понятного интерфейса (на русском языке), быстроте работы. Для построения прогнозов не требуются дополнительные знания других языков программирования, типа Python.



**Рисунок 8. Влияние параметра температура на DHI, DNI, GHI**

**Заключение.** Построена система нейросетевого прогнозирования значений прямой нормальной освещенности (DNI), рассеянного горизонтального излучения (DHI) или диффузного излучения неба и глобальной горизонтальной освещенности (GHI). Спроектированная нейросетевая модель учитывает 14 параметров, которые могут влиять на уровни излучения. Методом сценарного прогнозирования построены графики зависимостей выходных данных от изменения входных параметров. Применение такого набора параметров в модели позволяет с высокой точностью прогнозировать значения типов излучения.

### Библиографический список

1. Perez-Astudillo D. DNI, GHI and DHI Ground Measurements in Doha, Qatar / D. Perez-Astudillo, D. Bachour // *Energy Procedia*. – 2014. – Vol. 49. – P. 2398-2404.
2. Use of physics to improve solar forecast: Physics-informed persistence models for simultaneously forecasting GHI, DNI, and DHI / W. Liu, Y. Liu, X. Zhou [et al.] // *Solar Energy*. – 2021. – Vol. 215. – Use of physics to improve solar forecast. – P. 252-265.
3. Chu Y. Sun-tracking imaging system for intra-hour DNI forecasts / Y. Chu, M. Li, C. F. M. Coimbra // *Renewable Energy*. – 2016. – Vol. 96. – P. 792-799.
4. Increasing the Accuracy of Hourly Multi-Output Solar Power Forecast with Physics-Informed Machine Learning / D. V. Pombo, H. W. Bindner, S. V. Spataru [et al.] // *Sensors*. – 2022. – Vol. 22. – № 3. – P. 749.

5. Verma K. Wipro's-Sustainability-Machine-Learning-Challenge / K. Verma. – 2022. – URL: <https://github.com/skaran786/Wipro-s-Sustainability-Machine-Learning-Challenge> (date accessed: 25.03.2022). – Text : electronic.

6. Черепанов Ф. М., Ясницкий Л.Н. Нейросимулятор 5.0 // Свидетельство о государственной регистрации программы для ЭВМ № 2014618208. Заявка Роспатент № 2014614649. Зарегистрировано в Реестре программ для ЭВМ 12 августа 2014г.

7. Ясницкий Л.Н. Введение в искусственный интеллект. М.: Издательский центр «Академия», 2005. – 176 с.

## NEURAL NETWORK SYSTEM FOR FORECASTING THE EMPLOYEE'S LEAVING THE COMPANY WITHIN TWO YEARS

*Karibova Ainura S.*

National Research University “Higher School of Economics”  
St. Gagarin Boulevard, 37, Perm, Russia, 614060, askaribova@edu.hse.ru

The article presents a description of the development of a neural network system for predicting the values of direct normal illumination (DNI) or beam radiation, diffuse horizontal radiation (DHI) or diffuse sky radiation and global horizontal illumination (GHI). The system predicts the intensity of solar radiation for several years. A few studies have been carried out, the results of which are presented in this article.

**Keywords:** solar radiation, types of measured solar illumination, forecasting of solar energy production, neural network, ecology, sustainable development, solar impact, solar insolation, DHI, DNI, GHI.

УДК 004.032.26

## НЕЙРОСЕТЕВАЯ СИСТЕМА ПРОГНОЗИРОВАНИЯ ТИПА ЗВЕЗДЫ

*Жуйкова Светлана Константиновна*

Национальный исследовательский университет «Высшая школа экономики»,  
614070, Россия, г. Пермь, ул. Студенческая, 38, skzhuykova@edu.hse.ru

В статье представлено описание разработки нейросетевой системы для определения типа звезды по её физическим свойствам (температуре, цвету, светимости и других). С помощью разработанной нейросетевой системы было проведено исследование свойств звезд и их основных типов, а также были выявлены закономерности признаков и их значимость.



**Ключевые слова:** искусственный интеллект, нейросетевые технологии, тип звезды, классификация, диаграмма Герцшпрунга-Рассела.

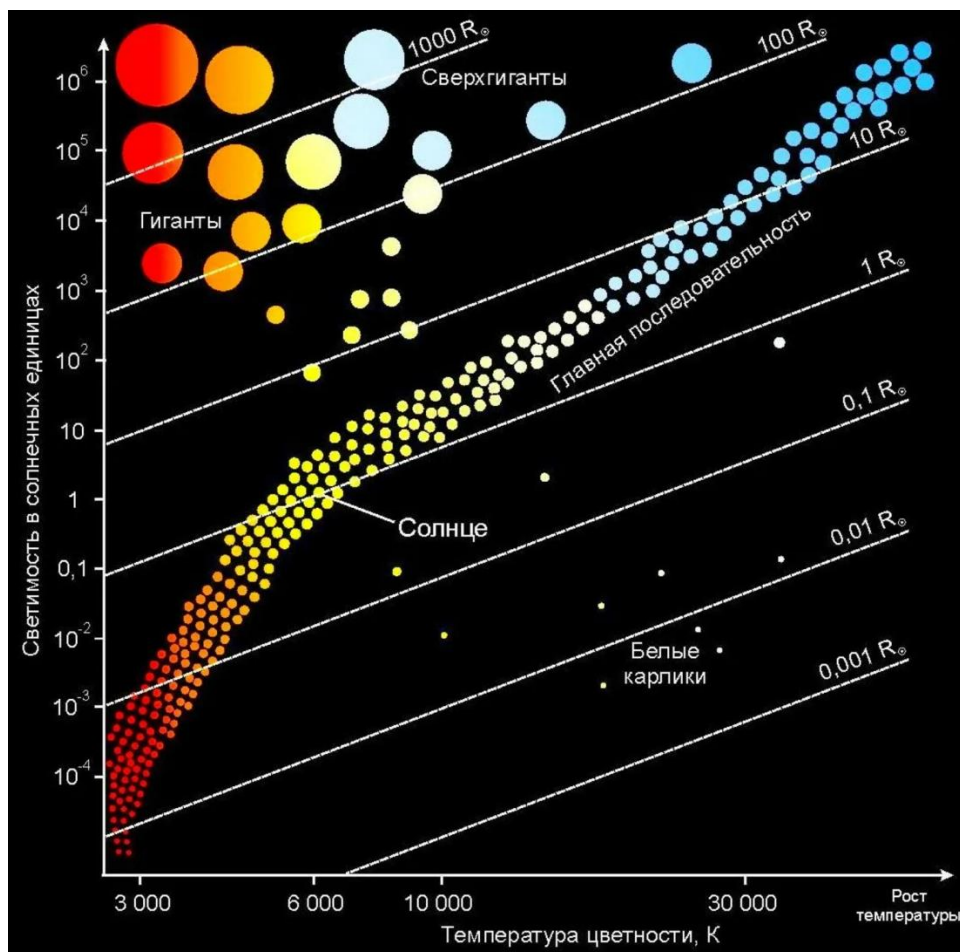
**Введение.** Определение типа звезды на основе её физических свойств сводится к задаче классификации. Такая задача может быть решена путем обучения нейросетевой модели в системе «Нейросимулятор 5». Данная модель поможет в значительной мере сократить время, затрачиваемое учеными на определение типа (размера) звезды для проведения дальнейшего исследования небесных тел. Кроме того, построенная модель позволит подтвердить гипотезу, что звезды следуют определенному графику в небесном пространстве: диаграмме Герцшпрунга-Рассела.

При анализе литературных источников выяснилось, что научных статей на тему автоматизации процесса определения типа звезды на основе её физических свойств – нет в свободном доступе. Единственные наработки, которые удалось найти – это построенные модели пользователями Kaggle. Модели построены на основе данных о цвете, спектральном классе, светимости, радиусе и температуре звезды. Для решения задачи пользователи Kaggle использовали такие методы машинного обучения, как дерево решений [1], логистическую регрессию [2], случайный лес [3]. Данные работы были выполнены на языке программирования Python с использованием дополнительных библиотек. Точность (accuracy) данных моделей составляет около 98-99%.

Цель данной исследовательской работы заключается в прогнозировании типа звезды на основе её физических свойств в более чем 99% случаев. Для решения задачи классификации необходимо собрать и подготовить данные о звездах, обработать датасет, поделить его на обучающую, валидационную и тестовую выборки, создать и обучить нейросетевую модель. Далее необходимо оценить адекватность модели и проанализировать спрогнозированные результаты. Построенная нейросетевая система должна правильно определять тип звезды в 90% случаев.

**Построение нейросетевой модели.** Диаграмма Герцшпрунга–Рассела представляет собой график, на котором по вертикальной оси отсчитывается светимость (интенсивность светового излучения) звезд, а по горизонтальной – наблюдаемая температура их поверхностей (рис. 1). Оба этих количественных показателя поддаются экспериментальному измерению при условии, что известно расстояние от Земли до соответствующей звезды [4]. Данная диаграмма упорядочивает классификацию наблюдаемых во Вселенной звезд. Построенная модель должна подтвердить гипотезу о том, что звезды следуют графику Герцшпрунга–Рассела.

Для создания нейросетевой системы были выбраны такие свойства звезд, как: температура, относительная светимость, относительный радиус, абсолютная величина, цвет и спектральный класс. Относительные величины в данном случае определяются относительно соответствующих характеристик Солнца. Опишем данные свойства более подробно с точки зрения астрономии.



**Рисунок 1. Классификация звезд на диаграмме Герцшпрунга–Рассела**

В астрономии температура и цвет тесно связаны между собой, поэтому они зачастую объясняются вместе.

Многообразие цветов звезд определяется длиной волны видимого излучения, которое они испускают: чем больше длина волны, тем более красный оттенок имеет звезда. В свою очередь, длина волны излучения, зависит от поверхностной (фотосферной) температуры звезды. Фотосфера имеет цвет, который зависит от ее температуры: если она очень горячая, звезда будет казаться белой, если холодная – желто-оранжевой, если еще холоднее – красной [5].

Светимость – это количество электромагнитной энергии, излучаемой звездой в единицу времени (то есть ее мощность). Она измеряется в ваттах, эрг/сек или в солнечной светимости. Светимость звезды зависит как от ее температуры, так и от ее поверхностной площади [6].

Радиус звезд обычно выражается относительно радиуса Солнца. Более крупные радиусы, например радиусы сверхгигантов или гипергигантов или полубольшая ось бинарной системы, часто выражаются в астрономических единицах (А.Е.), что эквивалентно расстоянию между Солнцем и Землей (около 150 млн км) [5].

В астрономии абсолютная величина – это видимая величина, которую имел бы объект, если бы находился на расстоянии 10 парсек (32,6 световых лет) от наблюдателя [7].

Спектральный класс определяется на основе температуры звезды. В настоящее время большинство звезд классифицируется по системе Моргана-Кинана (МК) с помощью букв О, В, А, F, G, К и М – последовательность от самых горячих (тип О) до самых холодных (тип М) [8].

В качестве выходного параметра модели будет один из шести типов звезд:

1. Коричневый карлик – небольшое гибридное небесное тело с малыми температурами и светимостью.

2. Красный карлик – маленькая и относительно холодная звезда. Это самая распространенная звездная типология во Вселенной.

3. Белый карлик – это маленькая звезда с очень низкой светимостью и цветом, стремящимся к белому. Во время формирования белые карлики обладают очень высокой температурой, которая постепенно снижается в соответствии с тепловыми обменами с космическим пространством.

4. Главная последовательность – это часть кривой внутри диаграммы Герцшпрунга-Рассела, расположенная в нижнем центре. Когда звезда находится в любой точке этой кривой, это означает, что она находится в самой длинной и стабильной части своей жизни.

5. Сверхгиганты – это звезды, которые классифицируются как I в спектральной классификации Йеркса. Большая часть сверхгигантов – массивные звезды, которые на последней стадии своей жизни значительно увеличивают свой радиус.

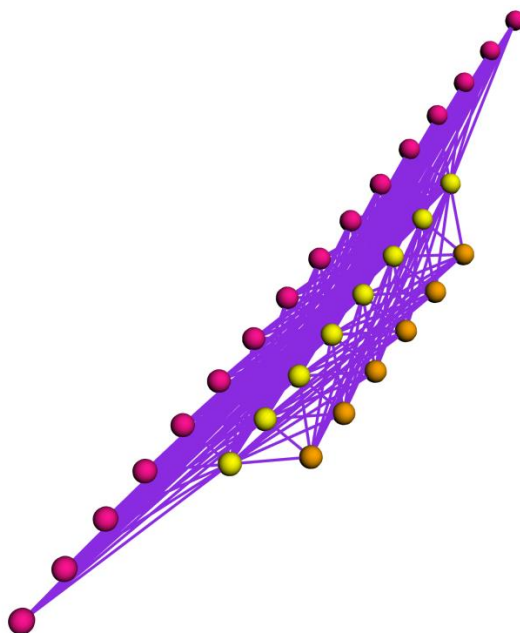
6. Гипергиганты – это массивные звезды, которые считаются самыми яркими, а их температура колеблется между 3500К и 35000К. Продолжительность их жизни оценивается примерно в два миллиона лет, в конце которой они взрываются в ярких сверхновых или даже гиперновых [9].

Следует отметить, что цвет и спектральный класс относятся к категориальным переменным, поэтому их необходимо разбить на отдельные переменные (с соответствующим цветом и спектральным классом). Таким образом, были определены следующие входные параметры: X1 – температура, X2 – относительная светимость, X3 – относительный радиус, X4 – абсолютная величина, X5 – красный цвет, X6 – желтый цвет, X7 – оранжевый цвет, X8 – белый цвет, X9 – голубой цвет, X10 – М спектральный класс, X11 – В спектральный класс, X12 – А спектральный класс, X13 – F спектральный класс, X14 – О спектральный класс, X15 – К спектральный класс, X16 – G спектральный класс. К выходным параметрам были отнесены шесть типов звезд: Y1 – Красный карлик, Y2 – Коричневый карлик, Y3 – Белый карлик, Y4 – Главная последовательность, Y5 – Сверхгигант, Y6 – Гипергигант.

Обучающее множество было взято из источника Kaggle [10]. Данная выборка была преобразована путем разбиения категориальных переменных на отдельные параметры. В исходном множестве были введены данные о 240 звездах. Данная выборка была поделена на обучающую, валидационную и тестовую в соотношении 80%, 15% и 5%.

Проектирование, обучение, тестирование нейронной сети выполнялись с помощью программы «Нейросимулятор 5» [11] по методике, описанной в

учебном материале «Введение в искусственный интеллект» Ясницким Л. Н. [12, с. 142-149]. Разработанная оптимизированная нейронная сеть представляет собой персептрон, который имеет шестнадцать входных нейронов, шесть выходных и один скрытый слой с восемью нейронами (рис. 2).



**Рисунок 2. Оптимизированная нейронная сеть, определяющая тип звезды**

Для оценки корректной работы спроектированной нейронной сети использовалось тестирующее множество, состоящее из 10 примеров. Средняя относительная ошибка тестирования составила 0.002% для типов Коричневый карлик и Белый карлик. В остальных случаях средняя относительная ошибка составила 0. На рисунке 2 (рис. 3) представлены гистограммы, демонстрирующие разницу между фактическим и прогнозируемым нейросетью типом звезды Коричневый карлик, Белый карлик и Сверхгигант.

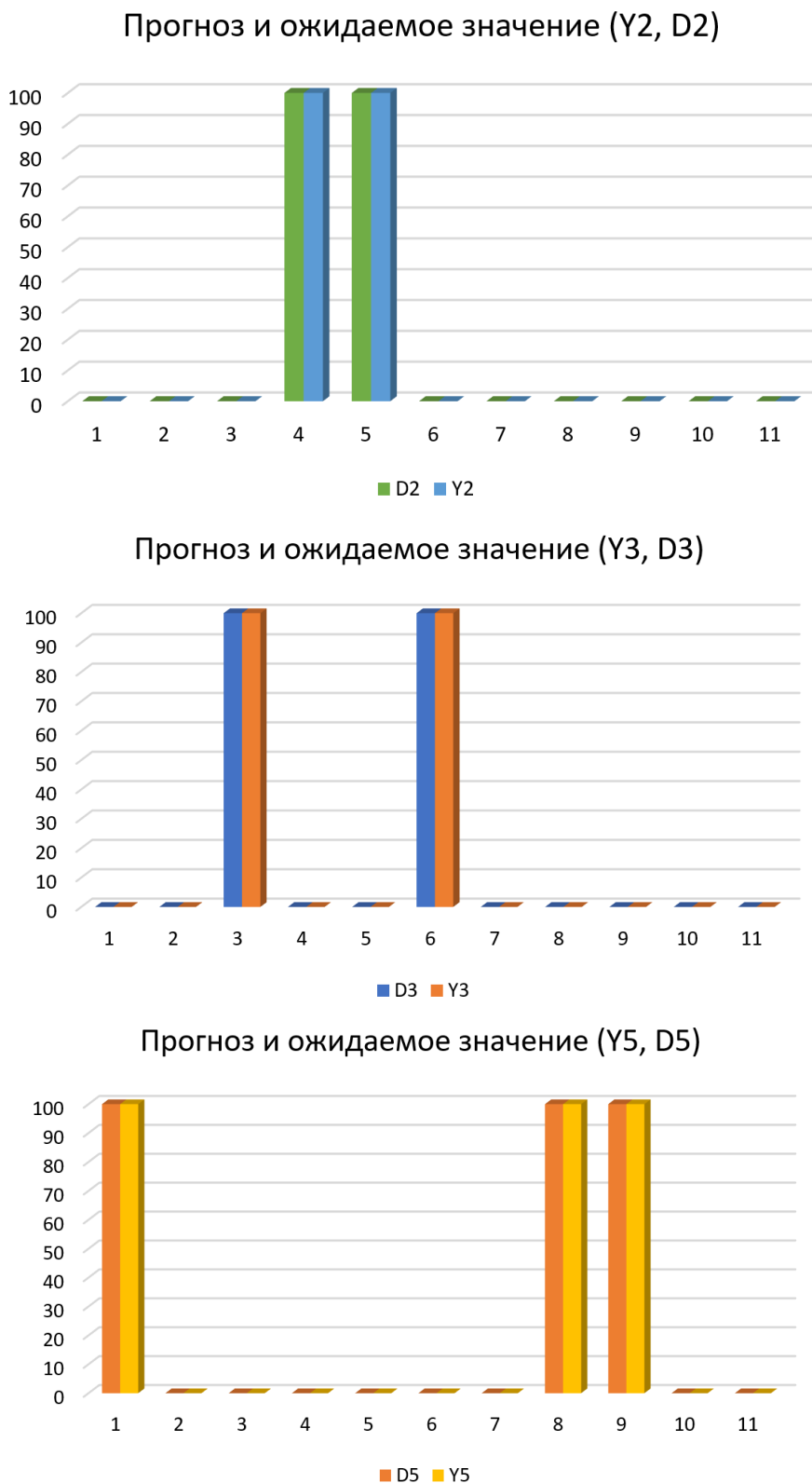
Далее при помощи средств «Нейросимулятор 5» была вычислена значимость каждого входного параметра (рис. 4). Наибольшая значимость параметра зафиксирована у абсолютной величины звезды (около 30%), а параметры «Оранжевый цвет», «К спектральный класс» и «G спектральный класс» была определены как незначимые.

Однако было решено не убирать данные переменные из модели, поскольку они могут быть использованы при дополнительном тестировании модели.

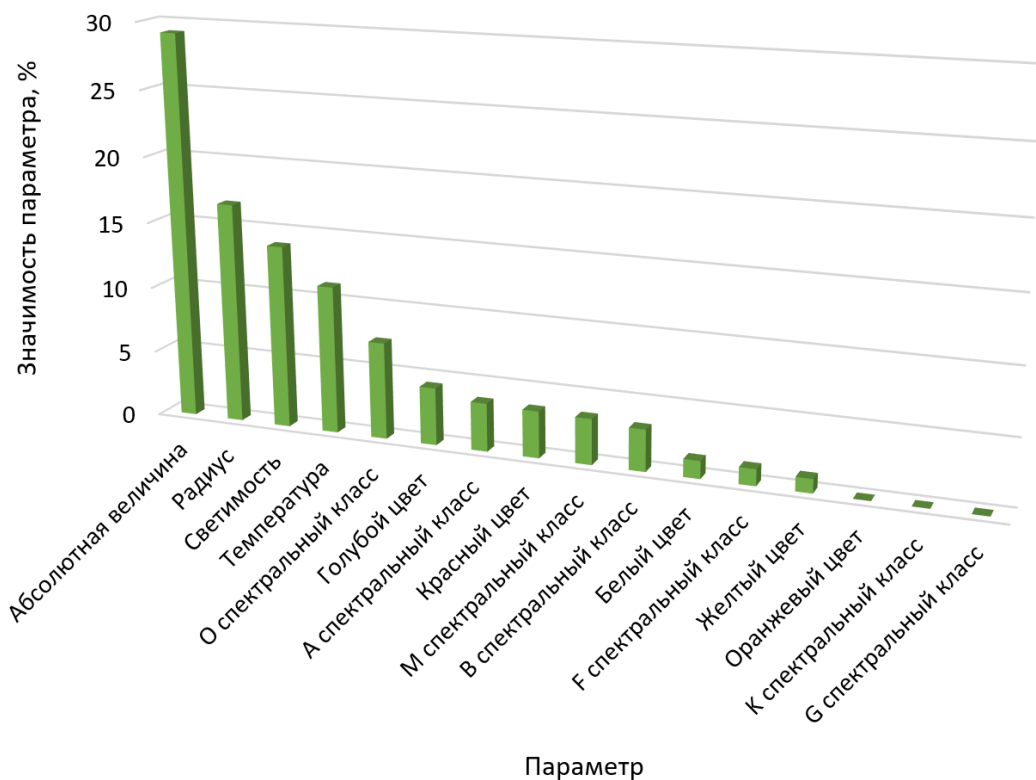
Далее было проведено исследование полученных зависимостей между входными параметрами и определенным типом звезды. Для этого использовался метод «замораживания» [12, с. 59-60], при котором меняется значение одного входного параметра, а остальные остаются неизменными.

Данное исследование было решено провести для четырех наиболее значимых входных параметров (со значимостью >10%): абсолютной величины, радиуса, светимости и температуры. Был отобран пример звезды с типом «Сверхгигант», отличающийся большим радиусом, высокой светимостью и

высокой температурой. Следует отметить, что модель имеет шесть выходных параметров, значение которых ранжируется от 0 до 100 (определяется как степень принадлежности звезды к определенному классу). В примерах ниже показан прогноз

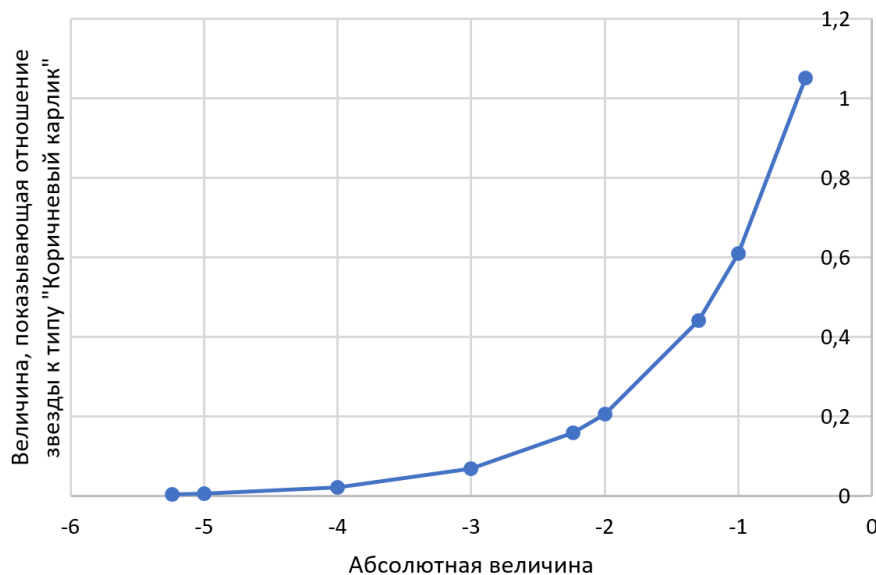


**Рисунок 3. Сравнение прогнозируемого и ожидаемого значений для типов звезд Коричневый карлик, Белый карлик и Сверхгигант**



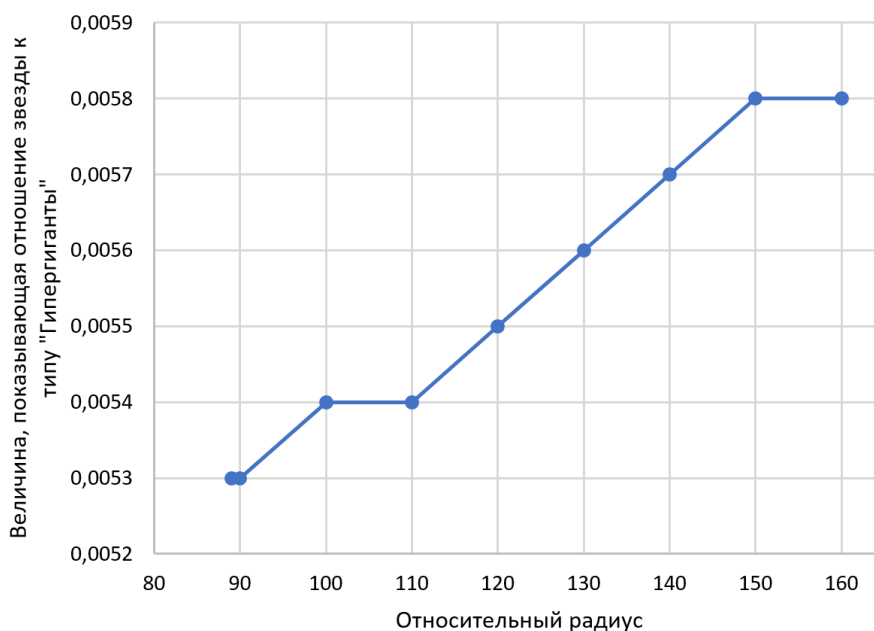
**Рисунок 4. Значимость входных параметров нейросетевой модели**

На рисунке ниже (см. рис. 5) отображен график зависимости прогнозируемого типа звезды от значения абсолютной величины. Как можно заметить, при увеличении абсолютной величины, модель все больше склоняется к типу «Коричневый карлик» (увеличивается значение выходного параметра Y2).



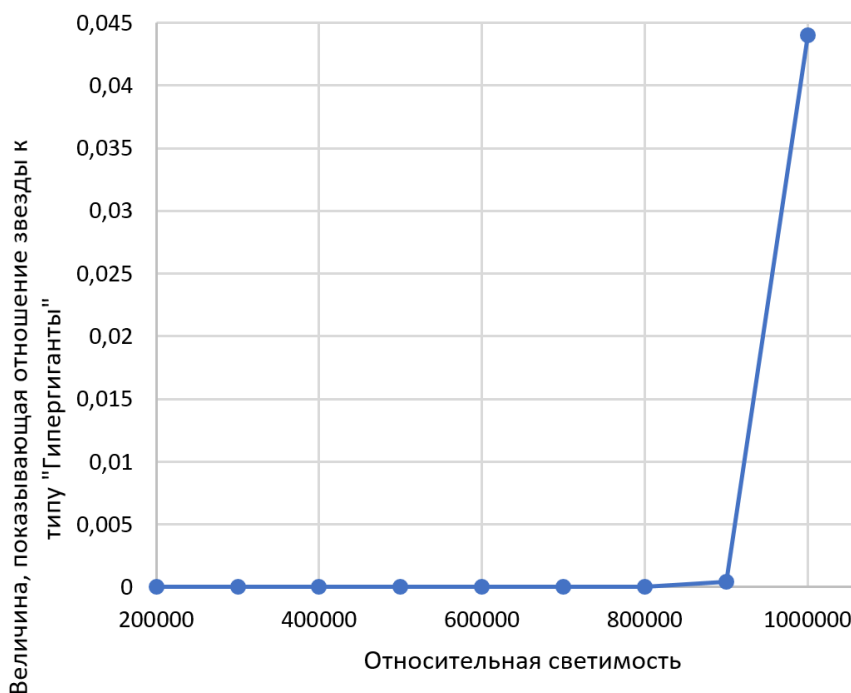
**Рисунок 5. Зависимость типа звезды от абсолютной величины**

При изменении относительного радиуса наблюдалось смещение модели к типу «Гипергигантов» (рис. 6). Для Гипергигантов характерен большой радиус звезды по сравнению со Сверхгигантами, поэтому поведение модели предсказуемо.



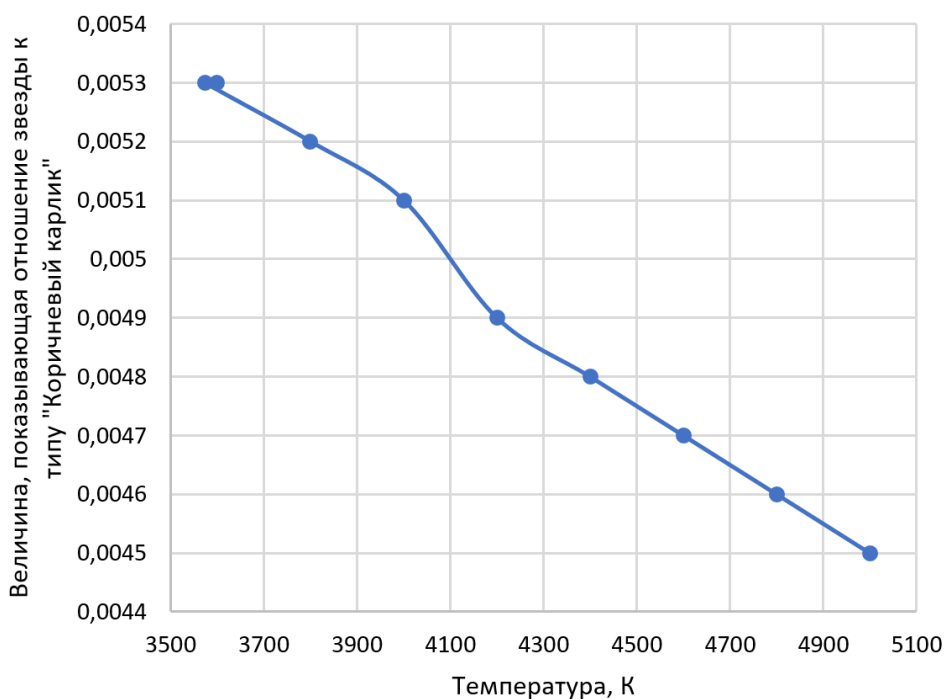
**Рисунок 6. Зависимость типа звезды от относительного радиуса**

Далее изменялся параметр относительной светимости с фиксацией остальных входных параметров (см. рис. 7). После значения 900 000 относительной светимости модель начала в большей мере склоняться к типу Гипергигантов (увеличение выходного параметра  $Y_6$ ).



**Рисунок 7. Зависимость типа звезды от относительной светимости**

На рисунке 7 отображена зависимость типа звезды от температуры (рис. 8). При увеличении температуры модель все в меньшей мере определяет звезду к типу «Коричневый карлик» (уменьшается значение выходного параметра  $Y_2$ ).



**Рисунок 8. Зависимость типа звезды от температуры**

Полученные зависимости наиболее значимых входных параметров не противоречат реальности. Следовательно, можно сделать вывод о том, что спроектированная нейронная сеть является пригодной для определения типа звезд на основе их физических свойств. На тестовом множестве типы звезд были определены корректно в 100% случаев.

**Построение нейросетевой модели с использованием библиотеки keras.** В данном разделе представлено описание построения нейросетевой модели прогнозирования типа звезды при помощи методов библиотеки Keras. После загрузки датасета из csv и деления множества на обучающую и тестовую выборку, стандартизируем входные параметры – приведем их такому виду, чтобы входной параметр имел нулевое среднее выборочное значение и стандартное отклонение, равное 1.

На этапе проектирования модели воспользуемся методами Keras: создадим модель и добавим в неё один скрытый слой, состоящий из 8 нейронов, и слой с выходными параметрами, состоящий из 6 нейронов (рис. 9).

```
# ПРОЕКТИРУЕМ МОДЕЛЬ
model = Sequential()

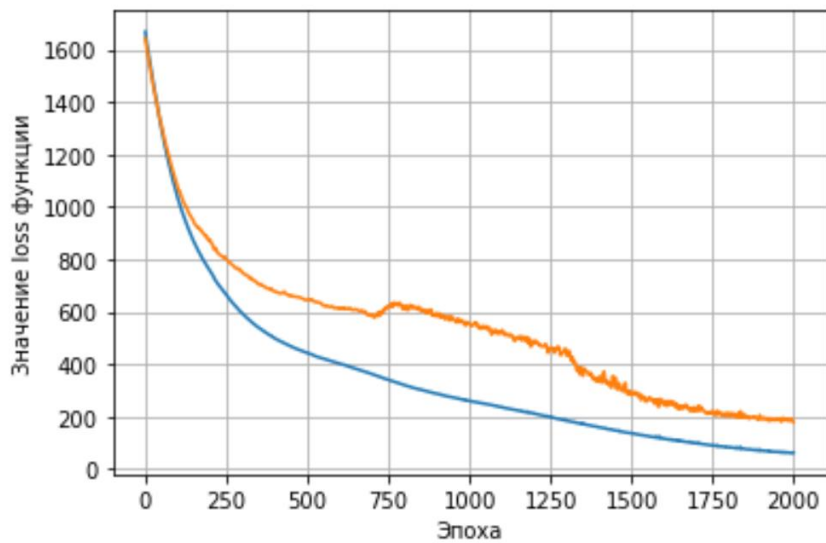
model.add(Dense(8, activation="tanh"))
model.add(Dense(6))
```

**Рисунок 9. Создаем модель нейронной сети прогнозирования типа звезды**

Далее обучаем построенную нейросетевую модель – было выбрано 1500 эпох. В результате была получена абсолютная ошибка mae, равная 4,08.

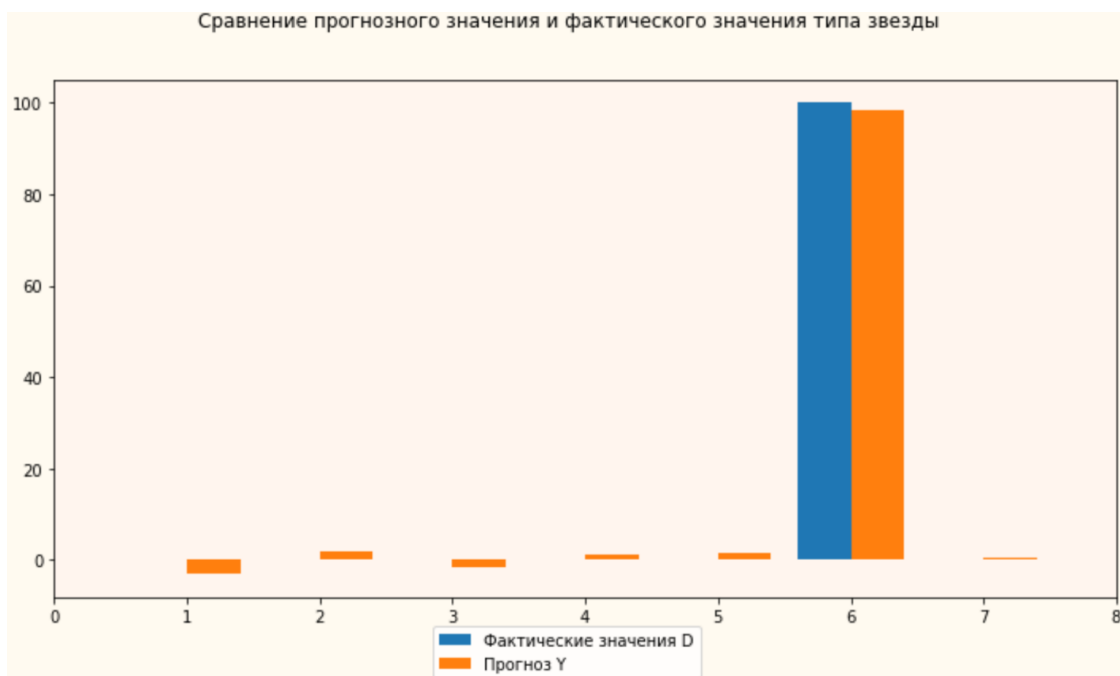


Кроме того, был построен график loss функции (рис. 10). По горизонтальной оси отмечен номер эпохи, а по вертикальной – значение loss функции.



**Рисунок 10. График loss функции при обучении нейросетевой модели**

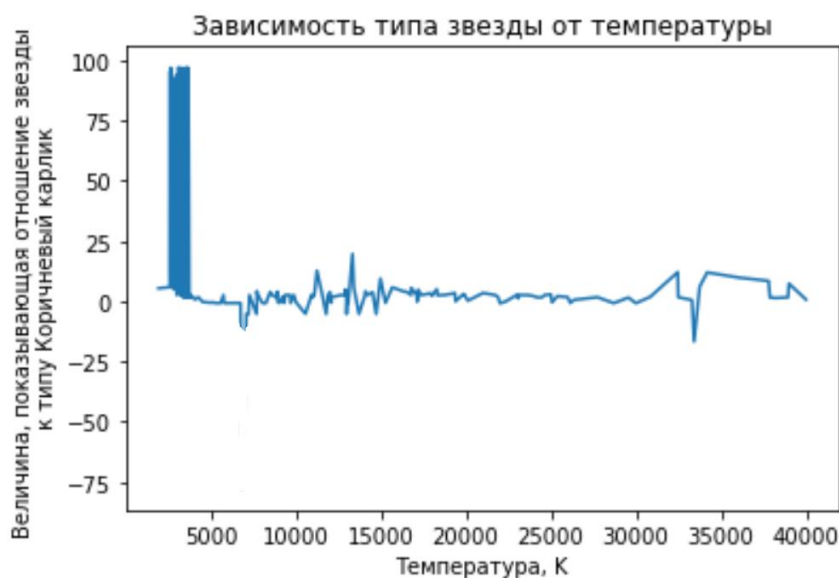
Далее сравним прогнозные значения с фактическим: построим столбчатую диаграмму (см. рис. 11). Как можно заметить на рисунке ниже, нейросетевая модель отнесла звезду к типу сверхгигантов (фактически звезда и являлась Сверхгигантом). Модель в незначительной степени ошибается.



**Рисунок 11. Прогноз модели и фактическое значение**

На следующем шаге была выполнена оценка качества модели: были вычислены такие показатели, как RMSE, MSE, MAE и MdAE. Среднеквадратичная ошибка составила 11,6, значения остальных показателей составили 5,3 и 2,09, соответственно.

Далее был использован метод «замораживания» для того, чтобы показать типа звезды «Коричневый карлик» от температуры. Как можно заметить по рисунку ниже (рис. 12) при температуре менее 4000 К модель в большинстве случаев относит звезду к типу «Коричневый карлик». При увеличении температуры модель считает, что звезда в меньшей степени относится к данному типу (значения колеблются около 0).



**Рисунок 12. Зависимость типа звезды от температуры**

Таким образом, была построена модель на python с использованием методов библиотеки Keras. Полученная модель по качеству получилась хуже, чем нейросеть, построенная в «Нейросимулятор 5»: модель больше ошибается. Кроме того, следует отметить, что построение нейросетевой модели при помощи инструмента «Нейросимулятор 5» гораздо проще: при помощи интерфейса пользователя можно построить нейросеть, обучить её, протестировать, вычислить значимость входных параметров, спрогнозировать результаты. Также важным преимуществом является факт того, что пользователю не обязательно нужно знать язык программирования для построения нейросетевой модели.

**Заключение.** В данной исследовательской работе была построена нейросетевая система прогнозирования типа звезды на основе её температуры, цвета, абсолютной величины, светимости, спектрального класса и радиуса с точностью, достигающей 100%. Категориальные признаки (цвет и спектральный класс) были разбиты на отдельные входные параметры. С целью проверки адекватности построенной модели были построены графики, сопоставляющие ожидаемое и спрогнозированное значения типов звезд. Средняя относительная ошибка составила 0,002% для типов «Коричневый карлик» и «Белый карлик», для остальных выходных параметров ошибка равна нулю.

Кроме того, в работе был использован метод сценарного прогнозирования: были построены графики зависимостей прогнозируемого типа звезды от изменения наиболее значимых входных параметров (температуры, абсолют-

ной величины, радиуса и светимости) с фиксацией остальных переменных. Следует отметить, что получившиеся зависимости соответствовали закономерностям в реальном мире. Таким образом, описанная в данной исследовательской работе нейросетевая модель может быть использована для прогнозирования типа звезды в дальнейших исследованиях небесных тел.

### Библиографический список

1. Star-Type Classifier. Kaggle [Электронный ресурс]. Режим доступа: <https://www.kaggle.com/yash161101/star-type-classifier> (дата обращения 20.02.2022).
2. Accuracy score of 98 using random forest. Kaggle [Электронный ресурс]. Режим доступа: <https://www.kaggle.com/sachinsharma1123/accuracy-score-of-98-using-random-forest> (дата обращения: 20.02.2022).
3. Star type prediction >99% accuracy. Kaggle [Электронный ресурс]. Режим доступа: <https://www.kaggle.com/pylajus/star-type-prediction-99-accuracy> (дата обращения: 20.02.2022).
4. Трефил Дж. Диаграмма Герцшпрунга-Рассела. Энциклопедия «Двести законов мироздания» [Электронный ресурс]. Режим доступа: [https://elementy.ru/trefil/21098/Diagramma\\_GertsshprungaRassela](https://elementy.ru/trefil/21098/Diagramma_GertsshprungaRassela) (дата обращения 26.02.2022).
5. Star dataset to predict star type. Kaggle [Электронный ресурс]. Режим доступа: <https://www.kaggle.com/deepu1109/star-dataset> (дата обращения 11.02.2022).
6. Светимость в астрономии. Астронет [Электронный ресурс]. Режим доступа: <http://www.astronet.ru/db/msg/1188710> (дата обращения: 26.02.2022).
7. Абсолютная звездная величина. Астронет [Электронный ресурс]. Режим доступа: <http://www.astronet.ru/db/msg/1162088> (дата обращения: 26.02.2022).
8. Спектральные класс звезд. Астронет [Электронный ресурс]. Режим доступа: <http://www.astronet.ru/db/msg/1162991> (дата обращения: 26.02.2022).
9. Звезды. Kvant.Space [Электронный ресурс]. Режим доступа: <http://kvant.space/zvyozdy> (дата обращения: 26.02.2022).
10. Star dataset to predict star types. Kaggle [Электронный ресурс]. Режим доступа: <https://www.kaggle.com/deepu1109/star-dataset> (дата обращения 08.02.2022).
11. Черепанов Ф.М., Ясницкий Л.Н. Нейросимулятор 5.0. Свидетельство о регистрации программы для ЭВМ RUS 2014618208. Заявка № 2014614649 от 05.05.2014.
12. Ясницкий Л.Н. Интеллектуальные системы : учебник. – М.: Лаборатория знаний, 2016. – 221 с

# NEURAL NETWORK SYSTEM FOR PREDICTING A STAR TYPE

*Zhuikova Svetlana K.*

National Research University “Higher School of Economics”  
Studencheskaya str., 38, Perm, Russia, 614070, Russia, skzhuykova@edu.hse.ru

The article describes the development of a neural network system for predicting a star type based on its physical properties (temperature, color, luminosity and others). With the help of the developed model, a study of the star’s properties and their main types were carried out. Moreover, patterns and significance of the mentioned above features were identified.

**Keywords:** artificial intelligence, neural network, star type, classification, Hertzsprung-Russell diagram.

УДК 004.032.26

## НЕЙРОСЕТЕВАЯ СИСТЕМА ПРОГНОЗИРОВАНИЯ ВЫЖИВАНИЯ ЧЕЛОВЕКА В КАТАСТРОФЕ НА ПРИМЕРЕ ТИТАНИКА

*Калинина Маргарита Олеговна*

Пермский государственный национальный исследовательский университет,  
ПМИ. 614990, Россия, г. Пермь, ул. Букирева, 15, ruinmath@bk.ru

В статье представлено описание разработки нейросетевой системы для прогнозирования судьбы человека в катастрофе. Система позволяет с большой точностью предсказать выживет ли человек в каком-либо крушении на основании данных о пассажирах Титаника. С помощью разработанной интеллектуальной системы проведено исследование предметной области, выявлены закономерности, имеющие практическое значение.

**Ключевые слова:** искусственный интеллект, нейросетевые технологии, прогнозирование, катастрофа, Титаник.

**Введение.** Крушение «Титаника» – одно из самых печально известных кораблекрушений в истории.

15 апреля 1912 года, во время своего первого рейса, широко известный “непотопляемый” «Титаник» затонул после столкновения с айсбергом. К сожалению, спасательных шлюпок, для всех находившихся на борту, не хватило, в результате чего погибло 1502 из 2224 человек, в число которых входили пассажиры и члены экипажа.

Хотя в выживании был определен элемент удачи, похоже, что у некоторых групп людей было больше шансов выжить, чем у других.

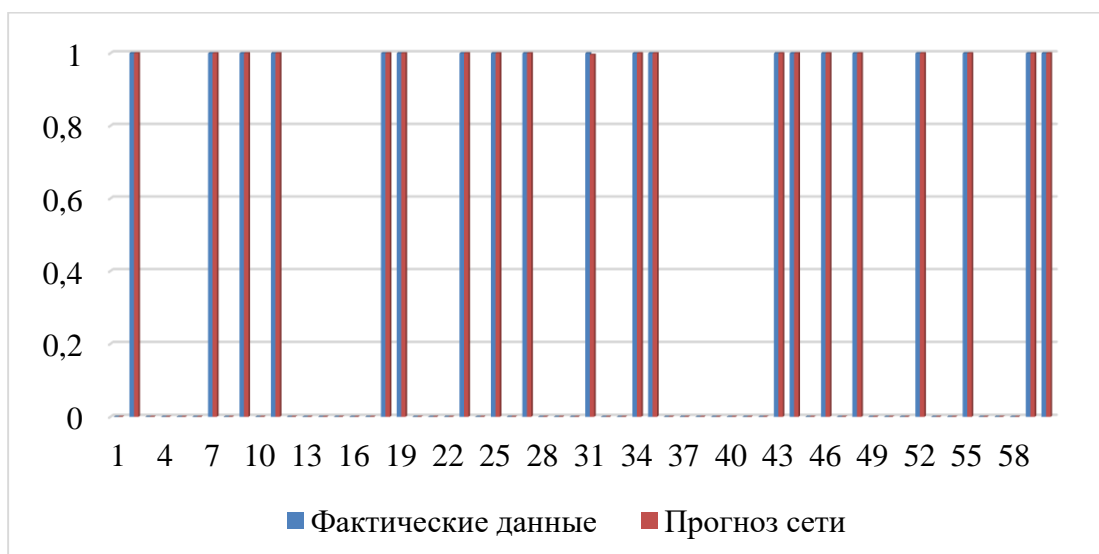
В данной работе будет построена прогностическая модель, которая отвечает на вопрос: “У каких людей было больше шансов выжить?”, используя данные о пассажирах.

Для создания нейросетевой системы были выбраны следующие параметры: X1 – класс билета, X2 – пол пассажира, X3 – возраст, X4 – количество братьев и сестер / супругов на борту, X5 – количество родителей / детей на борту, X6 – пассажирский тариф, X7 – порт посадки. Выходной параметр – вывод о том, выживет человек или нет.

Обучающее множество было взято с Интернет-ресурса [1]. Перед переходом к проектированию и обучению нейросетевой модели, была выполнена очистка исходного множества от столбцов с элементами, содержащими буквы (номер билета, номер каюты) и строк с пустыми ячейками, также было выполнено избавление от выбросов по методике [4]. В частности, из 714 примеров 138 было удалено. Данное множество было разделено на обучающее, валидирующее и тестирующее в соотношении 70%, 20% и 10% соответственно.

Проектирование, обучение, тестирование нейронной сети выполнялись с помощью программы «Нейросимулятор 5» [2] по методике [3]. В результате оптимизации спроектированная нейронная сеть представляла собой персептрон, который имеет семь входных нейронов, один выходной и один скрытый слой с двумя нейронами. В качестве функции обучения для входных нейронов и скрытых слоёв был выбран тангенс гиперболический.

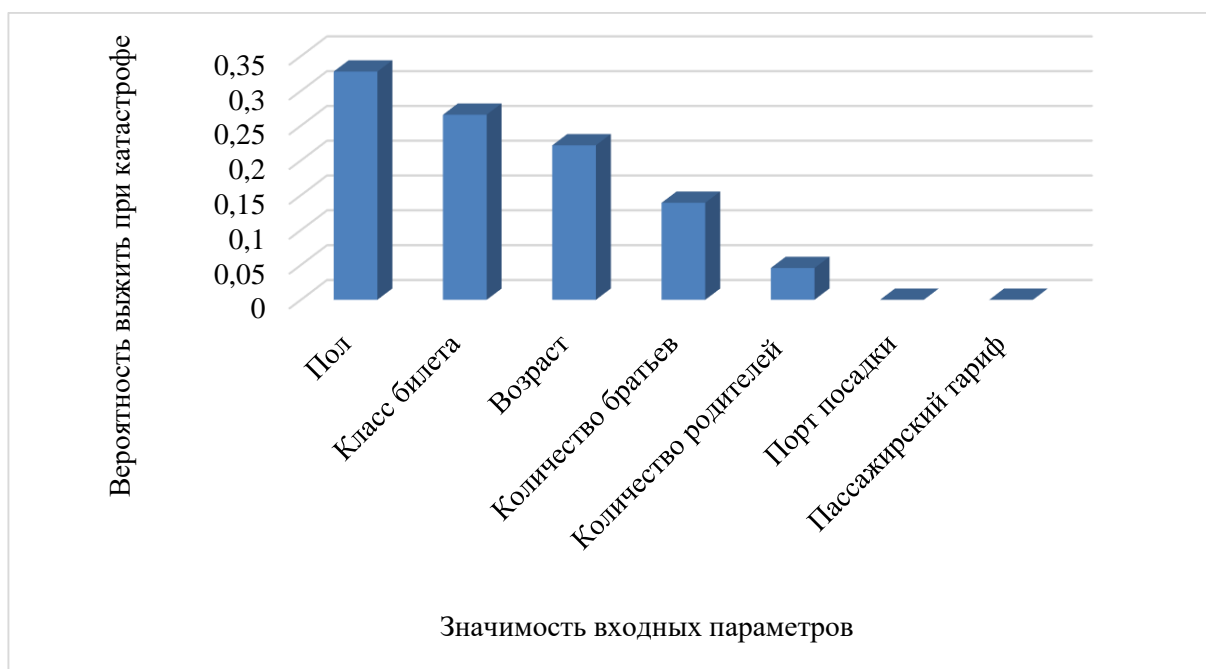
Для оценки корректной работы спроектированной нейронной сети использовалось тестирующее множество, состоящее из 61 примера. Средняя относительная ошибка тестирования составила 0,007%, что можно считать отличным результатом. На рисунке 1 представлена гистограмма, демонстрирующая разницу между фактическим и прогнозируемым результатом нейронной сети.



**Рисунок 1. Результат тестирования нейронной сети**

Из результатов, изображенных на рисунке 1, можно сделать вывод об адекватной работе спроектированной нейронной сети.

Оценка значимости входных параметров была выполнена с помощью программы «Нейросимулятор 5», результат отображен на рисунке 2.



**Рисунок 2. Значимость входных параметров нейросетевой модели**

Как видно из рисунка 2, наиболее значимыми являются пол, класс билета, возраст, и т.д. Как и ожидалось, наиболее влиятельным параметром является пол, так как в первую очередь всегда спасают женщин и детей.

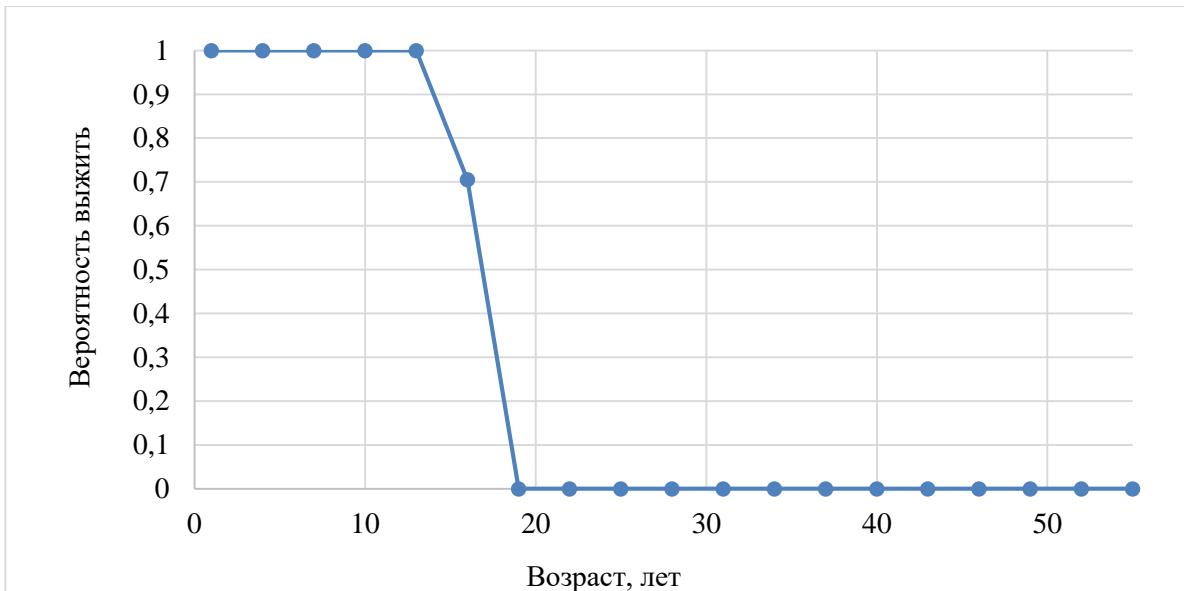
Далее было проведено исследование полученных зависимостей между входными параметрами и выжившими в катастрофе. Исследование производилось с помощью метода «замораживания» [3], суть которого заключается в варьировании значения одного параметра и фиксации значений всех других параметров. Для этой цели был отобран «нейтральный пример», про который нейросеть не может с уверенностью сказать, у кого из игроков значительное преимущество.

На рисунке 3 показан график зависимости прогнозируемого выжившего от возраста. В качестве параметра был взят мужчина, с увеличением возраста можно проследить уменьшение вероятности выжить.

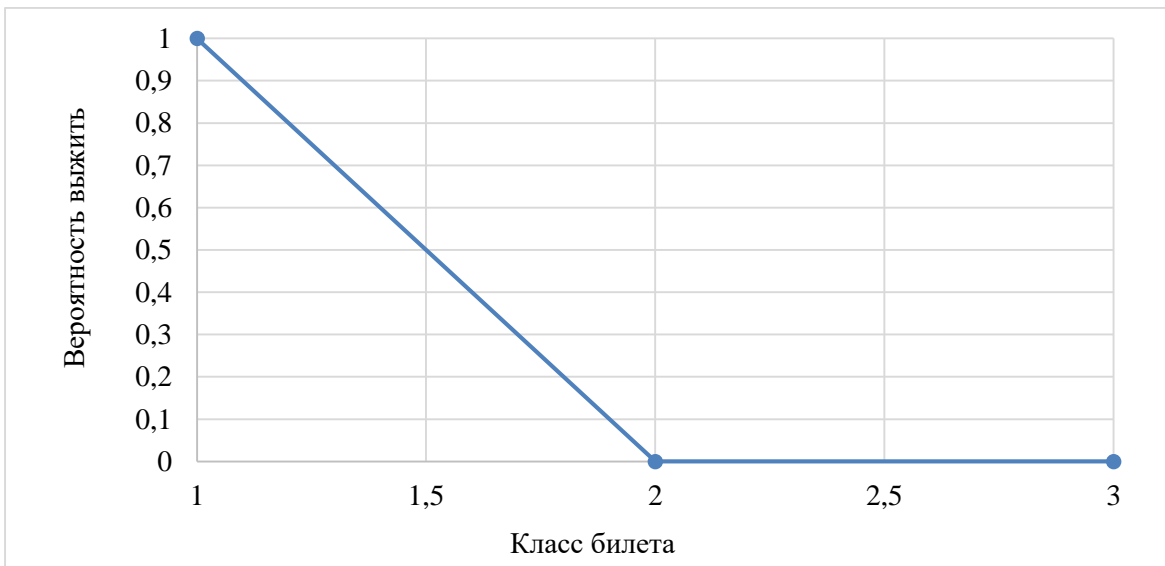
На рисунке 4 продемонстрирована зависимость прогнозируемого выжившего от класса билета. Можно заметить, что с уменьшением класса вероятность выжить падает.

На рисунке 5 изображен график зависимости прогнозируемого выжившего от пола. Как видно из графика, вероятность выжить у женщины больше, чем у мужчины.

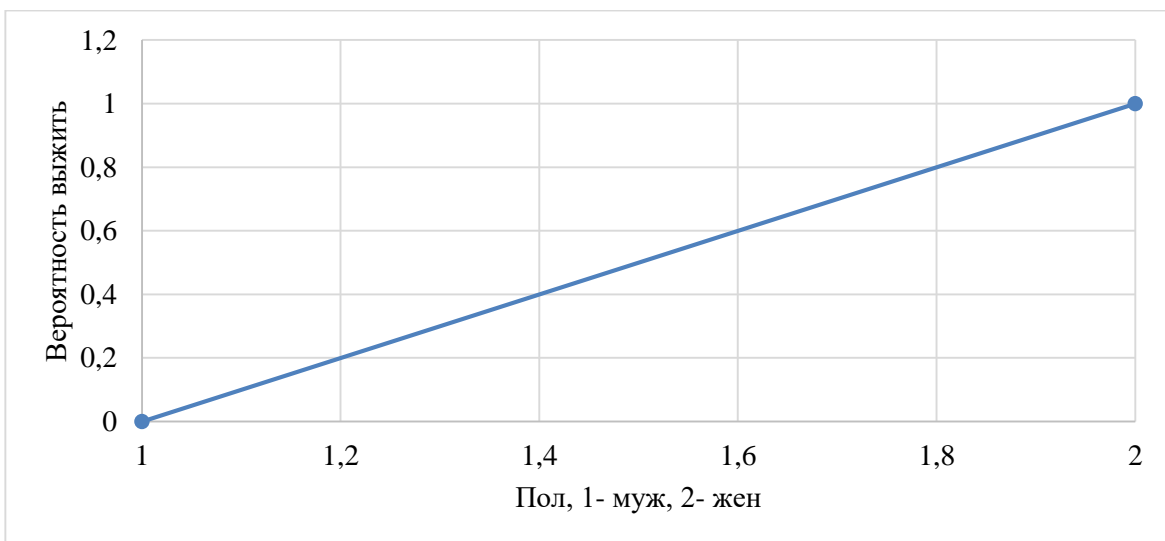
На рисунке 6 продемонстрирована зависимость прогнозируемого победителя матча от количества братьев и сестер / супругов на борту. Как видно из графика, с увеличением количества резко уменьшается вероятность выжить.



**Рисунок 3. Зависимость прогнозируемого выжившего от возраста**



**Рисунок 4. Зависимость прогнозируемого выжившего от класса билета**

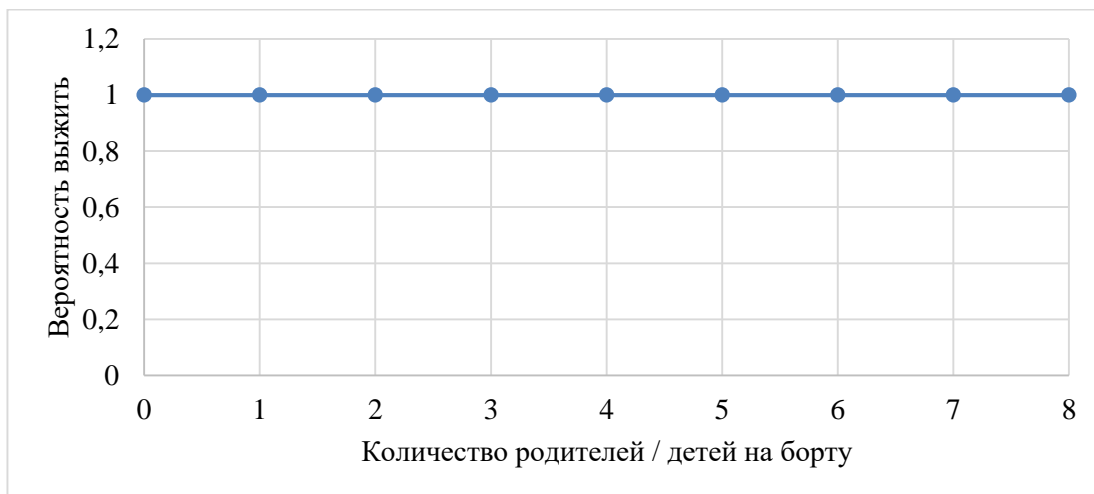


**Рисунок 5. Зависимость прогнозируемого выжившего от пола**



**Рисунок 6. Зависимость прогнозируемого победителя от количества братьев и сестер / супругов на борту**

На рисунке 7 продемонстрирована зависимость прогнозируемого выжившего от количества родителей / детей на борту. Можно заметить, что вероятность выжить с увеличением количества не уменьшается, это можно связать с тем, что спасают в первую очередь детей с их родителями.



**Рисунок 7. Зависимость прогнозируемого выжившего от количества родителей / детей на борту**

Полученные результаты исследований не противоречат реальности, что подтверждает, что спроектированную нейронную сеть можно считать пригодной для прогнозирования результатов выживания при катастрофе.

Данная работа была также выполнена в Google Collab с использованием Keras [5]. В ходе выполнения были выявлены следующие отличия: в Keras требуется писать много кода, в то время как в Нейросимуляторе всё это скрыто и есть удобный интерфейс; в Keras больше возможностей, но в то же время в нём сложнее разобраться, чем в Нейросимуляторе.



Результат в Keras оказался хуже, чем в отечественном аналоге. Коэффициент детерминации в нём равен 0,999 что на 0,001 меньше, чем в Нейросимуляторе.

Пользователю, который только начинает знакомиться с нейронными сетями не надо задумываться о том, что лежит в глубине программы, поэтому лучше начинать с Нейросимулятора.

**Заключение.** Построена система нейросетевого прогнозирования выживших при катастрофе. Спроектированная нейросетевая модель учитывает 7 параметров: класс билета, пол, возраст, количество братьев и сестер / супругов на борту, количество родителей / детей на борту, пассажирский тариф, порт посадки. Методом сценарного прогнозирования построены графики зависимостей прогнозируемого выжившего от изменения входных параметров. Применение такого набора параметров в модели позволяет с высокой точностью прогнозировать выживших при различных катастрофах.

### Библиографический список

1. kaggle.com – множество людей на Титанике. [Электронный ресурс]. Режим доступа: <https://www.kaggle.com/c/titanic/data>
2. Черепанов Ф.М., Ясницкий Л.Н. Нейросимулятор 4.0. Свидетельство о регистрации программы для ЭВМ RUS 2014612546. Заявка № 2014610341 от 15.01.2014.
3. Ясницкий Л.Н. Введение в искусственный интеллект. М.: Издательский центр «Академия», 2005. – 176 с.
4. Yasnitsky L.N. Algorithm for searching and analyzing abnormal observations of statistical information based on the Arnold – Kolmogorov – Hecht-Nielsen theorem // International Journal of Advanced Trends in Computer Science and Engineering. 2020. 9(2). Pp. 1814-1819.
5. [https://colab.research.google.com/drive/1a\\_ah61VQD\\_xlPwp4yDgbtK83rnIX5gJ-?usp=sharing](https://colab.research.google.com/drive/1a_ah61VQD_xlPwp4yDgbtK83rnIX5gJ-?usp=sharing)

## NEURAL NETWORK SYSTEM FOR PREDICTING THE WINNER OF A SNOOKER MATCH

*Kalinina Margarita O.*

Perm State University

Str. Bukireva, 15, Perm, Russia, 614990, [ruinmath@bk.ru](mailto:ruinmath@bk.ru)

The article describes the development of a neural network system for predicting survivors of a disaster. The system allows you to accurately predict survivors based on data on age, gender, ticket class, number of relatives on board, ticket fare and port of embarkation. With the help of the developed intellectual system, a study of the subject area was carried out, patterns of practical importance were identified.

**Keywords:** artificial intelligence, neural network, forecasting, catastrophe, Titanic.

# **ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ В ПРОБЛЕМАХ БЕЗОПАСНОСТИ**

## НЕЙРОСЕТЕВАЯ СИСТЕМА ДЕТЕКТИРОВАНИЯ МЕДИЦИНСКИХ МАСОК НА ЛИЦАХ ЛЮДЕЙ

*Понькин Никита Александрович,  
Гильмутдинов Ринат Рашитович*

Пермский государственный национальный исследовательский университет,  
ПМИ. 614990, Россия, г. Пермь, ул. Букирева, 15, ruinmath@bk.ru

В статье представлено описание разработки нейросетевой системы для распознавания (обнаружение или детекция объекта) и классификации медицинских масок на лицах людей (или её отсутствия) на фото- или видеокадрах. Система позволяет с большой точностью определить, есть ли на лице человека маска, и правильно ли она надета. С помощью разработанной интеллектуальной системы проведено исследование предметной области, выявлены закономерности, имеющие практической значение.

**Ключевые слова:** алгоритмы детекции объектов и их классификации, распознавание образов, искусственный интеллект, нейросетевые технологии, детекция, медицинская маска, лица.

**Введение.** Одной из областей применения объектовой видео аналитики является безопасность. В условиях пандемии появляется требование обязательного ношения масок во многих регионах России. В такой ситуации целесообразно контролировать и детектировать использование медицинских масок и иных средств индивидуальной защиты персоналом торговых точек и медицинских учреждений. Опыт показывает, что использование нейросетевых моделей может помочь в решении задач в этих сферах. Поэтому актуальность задачи детекции людей, лиц и масок на них очевидна.

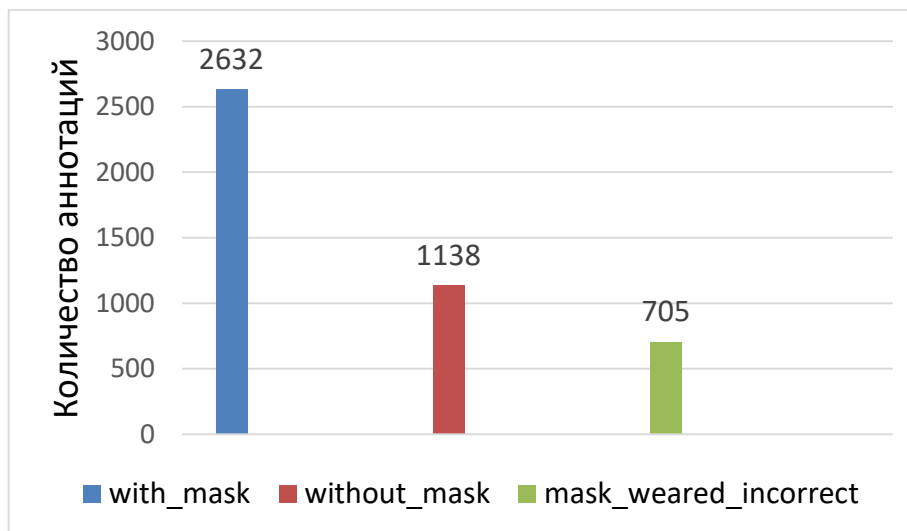
Первая работа в области распознавания образов в России была выполнена одним из основоположников современной теории информации – академиком А.А. Харкевичем [1]. В ней рассматриваются так называемые «читающие автоматы», а математическим аппаратом являлась теория статистических решений. Слово «образ» использовалось для обозначения напечатанного или написанного от руки знака, изображающего букву или цифру.

Задача детекции объектов — задача, в рамках которой необходимо выделить несколько объектов на изображении посредством нахождения координат их ограничивающих рамок и классификации этих ограничивающих рамок из множества заранее известных классов. В отличие от классификации с локализацией, число объектов, которые находятся на изображении, заведомо неизвестно.

В зависимости от формы области применения, дальнейшая детекция возможна в виде: ограничивающих прямоугольников, многоугольников, битовых масок, в виде точек интереса и пр. Существенно важны: топология области применения (ее форма, площадные зависимости и пр.) и/или особенно-

сти координатных зависимостей «особых» точек. В нашей задаче были выбраны ограничивающие прямоугольники – bounding boxes.

Помимо формы предсказания, необходимо было определить формат набора данных (способ организации данных и аннотаций). Выбор пал на MS COCO и YOLO.



**Рисунок 1. Баланс классов**

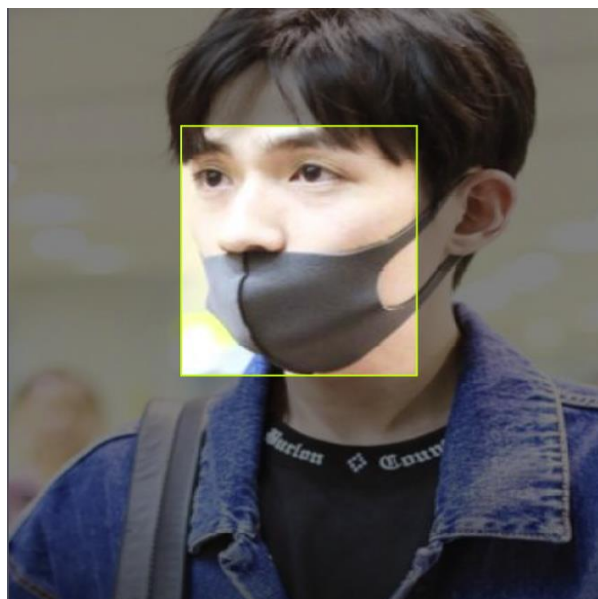


**Рисунок 2. Пример изображения с классом объектов with\_mask**

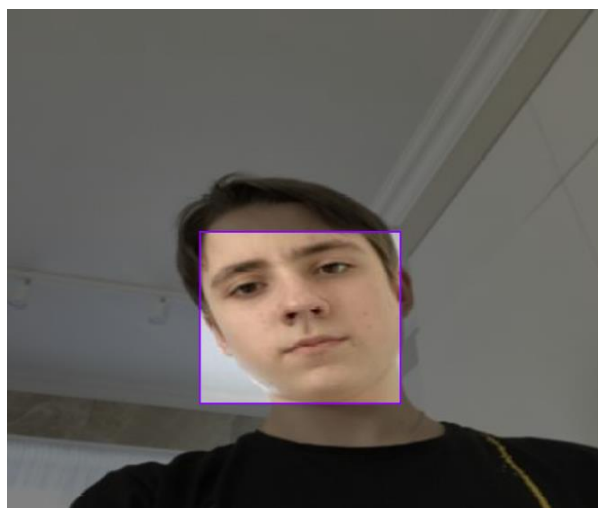
Основная цель настоящей работы заключается в сборе множества публичных данных, а также создание и обучение нейросетевой модели на этих данных. Конечным результатом должна быть нейросетевая система, способная обнаруживать медицинскую маску на лице человека на видеокадре.

Обучающее множество было собрано с публичного интернет-ресурса Kaggle [2]. Данные в задаче представлены классическим для технологии машинного зрения образом – в виде цветных RGB снимков. Дополнительная

обработка, аугментирование и аннотирование выполнялись с помощью сервиса Roboflow [3]. Таким образом, объем итогового множества включает в себя 1 231 фотографию с 4 475 аннотациями. Данное множество было разделено на обучающее, валидирующее и тестирующее в соотношении 76%, 15% и 9% соответственно (в валидирующее и тестовое множества аугментированные данные не попали). Для создания нейросетевой системы был выбран набор данных, в котором все изображения поделены на 3 основных класса: `with_mask` – маска надета, `without_mask` – маски нет, `mask_wearred_incorrect` – маска надета неправильно.



**Рисунок 3. Пример изображения с классом объектов `mask_wearred_incorrect`**



**Рисунок 4. Пример изображения с классом объектов `without_mask`**

Проектирование, обучение и тестирование нейронной сети выполнялись с помощью языка программирования Python, фреймворка PyTorch и библиотеки Detectron2. Изучив особенности поставленной задачи и необходимую литературу [6], была выбрана архитектура нейронной сети Faster R-CNN [7].

Процесс работы Faster R-CNN с подаваемым изображением можно описать так:

1. Извлечение карты признаков изображения с помощью нейронной сети;
2. Генерация на основе полученной карты признаков гипотез – определение приблизительных координат ограничивающих прямоугольников и наличия объекта любого класса;
3. Сопоставление координат гипотез с картой признаков, полученной на первом шаге;
4. Классификация гипотез (уже на определение конкретного класса) и дополнительное уточнение координат

Для генерации гипотез используется, по сути, небольшая отдельная нейронная сеть – **Region Proposal Network**.

В изначальной концепции R-CNN, от которой произошла Faster R-CNN, каждая предложенная гипотеза по отдельности обрабатывается с помощью CNN – такой подход стал своеобразным бутылочным горлышком (сеть была слишком долгой и в обучении, и в предсказании). Для решения этой проблемы авторами был разработан **Region of Interest (RoI)** слой. Этот слой позволяет единожды обрабатывать изображение целиком с помощью нейронной сети, получая на выходе карту признаков, которая далее используется для обработки каждой гипотезы.

Основной задачей RoI слоя является сопоставление координат гипотез (координаты ограничивающих рамок) с соответствующими координатами карты признаков. Делая «срез» карты признаков, RoI слой подает его на вход полносвязному слою для последующего определения класса и поправок к координатам.

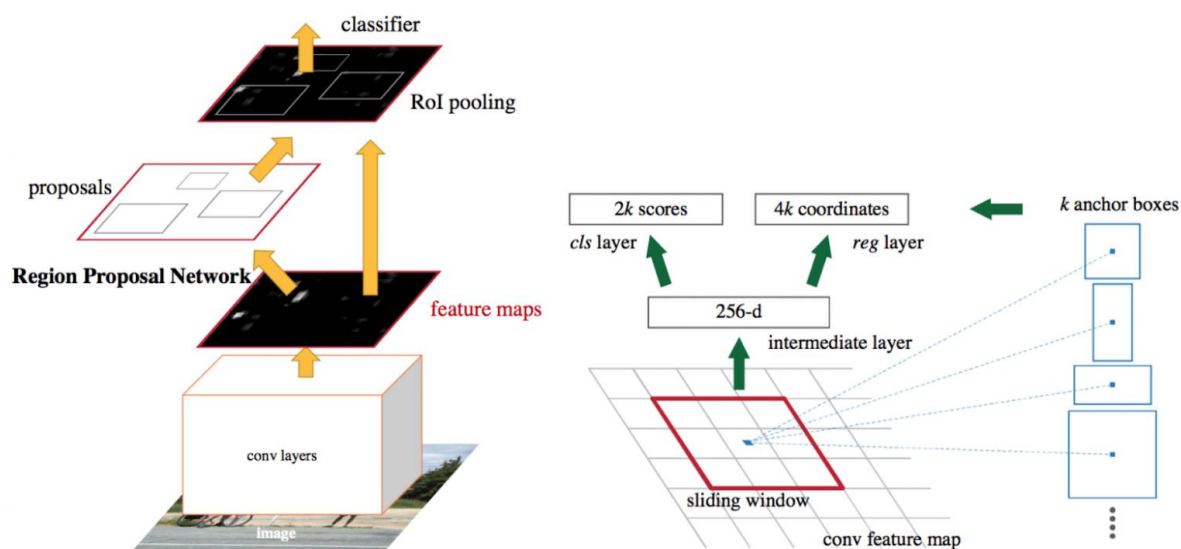
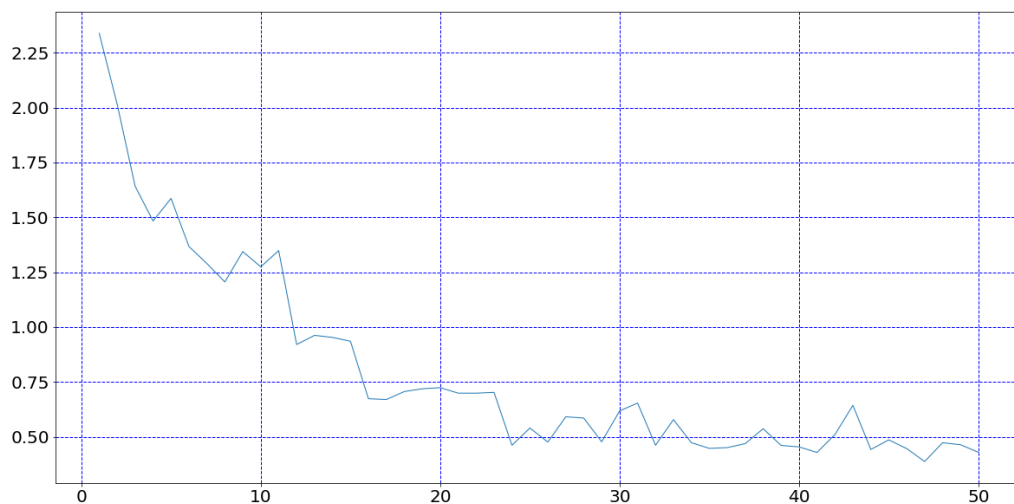
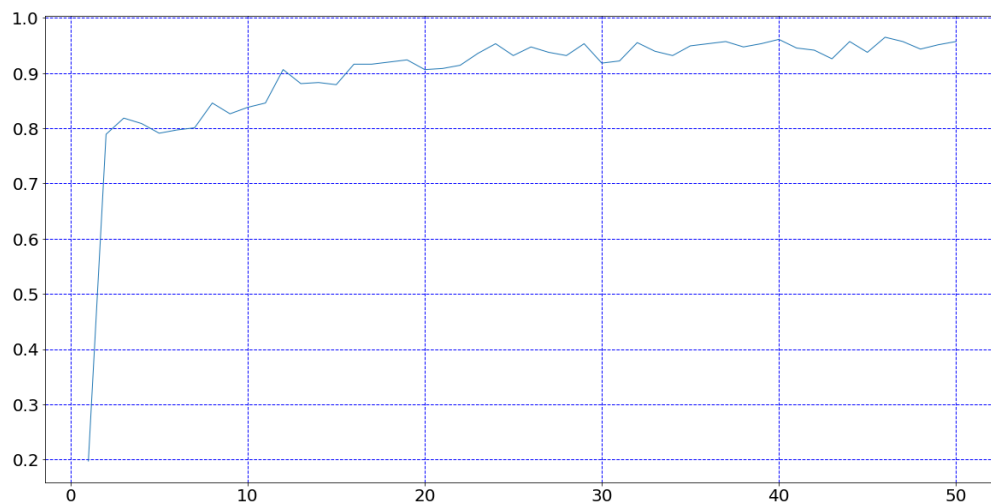


Рисунок 5. Архитектура Faster-RCNN

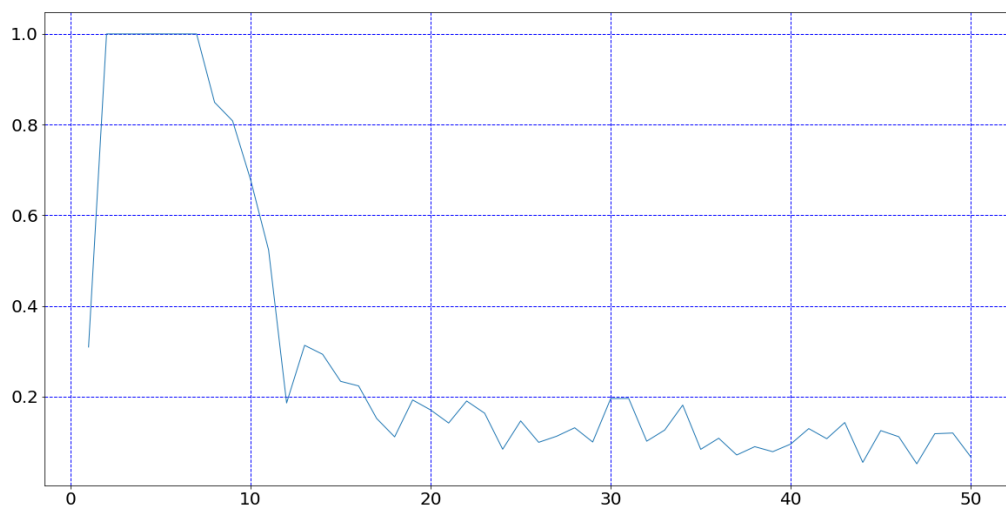
В одновременном обучении сети для задач регрессии ограничивающих рамок и классификации применяется специальная лосс-функция, объединяющая в себе кросс-энтропию классификации объектов, а также регрессионный лосс координат прямоугольников.



**Рисунок 6. Изменение лосса с процессом обучения (на валидирующем множестве)**



**Рисунок 7. Изменение точности предсказания классов боксов с процессом обучения (на валидирующем множестве)**



**Рисунок 8. Изменение доли False Negative (непредсказанных, проигнорированных сетью) боксов с процессом обучения (на валидирующем множестве)**

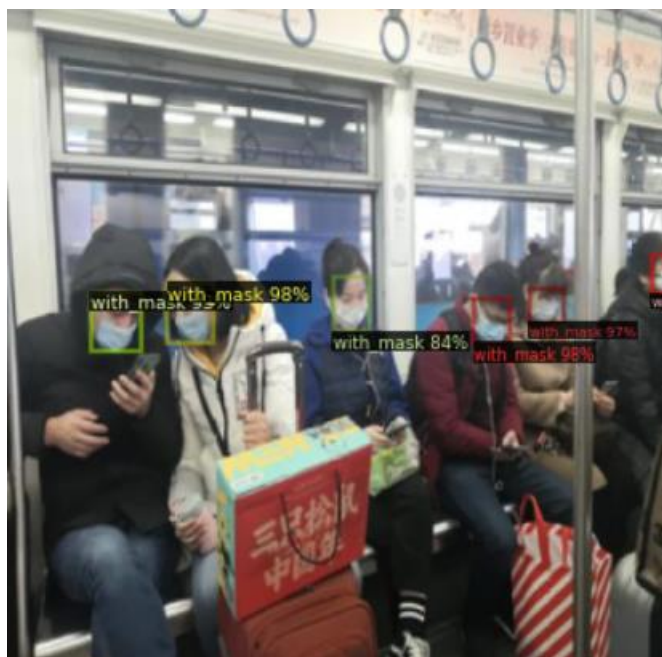
Во время использования нейронной сети для предсказаний прохождения изображения через неё выглядит так:

1. Изображение поступает на вход нейронной сети, генерируя карту признаков.

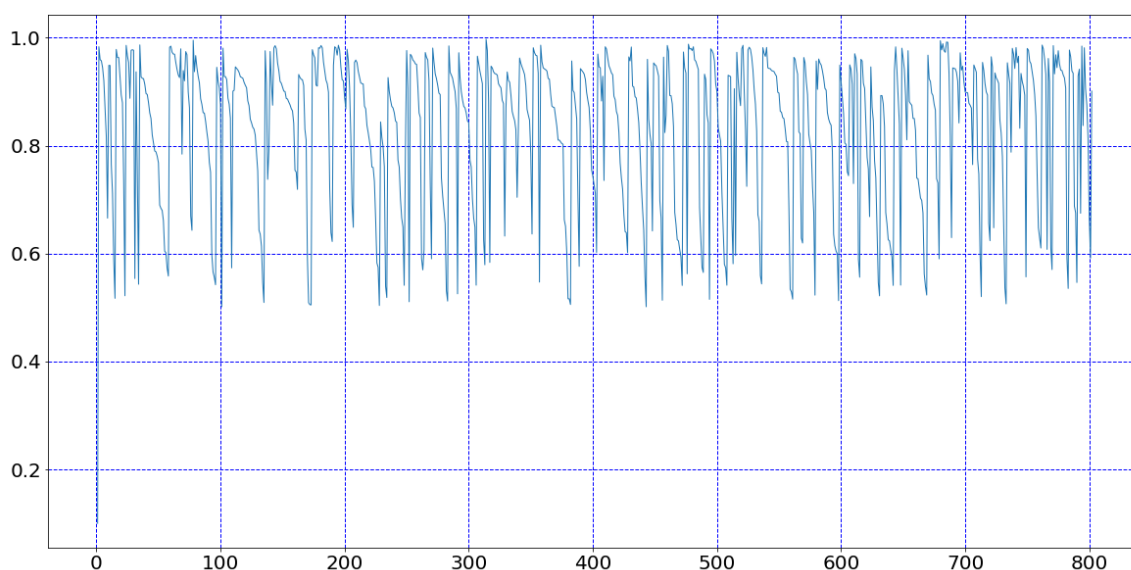
2. Каждая ячейка карты признаков обрабатывается с помощью RPN, выдавая в результате поправки к положению координат боксов и вероятность наличия объекта любого класса.

3. Соответствующие предсказанные рамки далее на основе карты признаков и RoI слоя поступают в дальнейшую обработку сети.

4. На выходе получаем уже конкретный класс объектов и их точное положение на изображении.



**Рисунок 9. Результат на тестовом множестве**



**Рисунок 10. Точность классификации объектов на тестовом множестве**



Для оценки корректной работы спроектированной нейронной сети использовалось тестирующее множество, состоящее из 115 примеров. Среднее значение метрики precision 0.776, метрики recall – 0.804, что можно считать приемлемым результатом.

Полученные результаты исследований не противоречат реальности, что подтверждает, что спроектированную нейронную сеть можно считать пригодной для обнаружения медицинских масок на лицах людей на фото- видео кадрах.

**Заключение.** Построена нейросетевая система обнаружения медицинских масок на лицах людей на фото- видео кадрах. Собран и дополнительно обработан набор данных, включающий в себя 3 класса данных: with\_mask – маска надета, without\_mask – маски нет, mask\_worn\_incorrect – маска надета неправильно. Проведены реальные тесты с использованием построенной нейросетевой системы.

### Библиографический список

1. Харкевич А.А. Опознавание образов. – «Радиотехника», 1959, т.14, № 5, с.12-22.
2. Kaggle.com – система организации конкурсов по исследованию данных, а также социальная сеть специалистов по обработке данных и машинному обучению. [Электронный ресурс]. Режим доступа: <https://www.kaggle.com/>
3. Roboflow.com – сервис для разработчиков компьютерного зрения, позволяющая улучшить методы сбора данных, их предварительной обработки и обучения моделей. [Электронный ресурс]. Режим доступа: <https://app.roboflow.com/>
4. Статья на тему «Самая сложная задача в Computer Vision. Трекинг.» 15 июня 2020. [Электронный ресурс]. Режим доступа: <https://habr.com/ru/company/recognitor/blog/505694/>
5. Статья на тему «3 метода детектирования объектов с Deep Learning: R-CNN, Fast R-CNN и Faster R-CNN» 24 июля 2020. [Электронный ресурс]. Режим доступа: <https://clck.ru/ZPXAR>
6. Ясницкий Л.Н. Введение в искусственный интеллект. М.: Издательский центр «Академия», 2005. – 176 с.
7. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks, 4 июня 2015. [Электронный ресурс]. Режим доступа: <https://arxiv.org/abs/1506.01497>

# NEURAL NETWORK SYSTEM FOR DETECTING MEDICAL MASKS ON PEOPLE'S FACES

*Ponkin Nikita A.,  
Gilmutdinov Rinat R.*

Perm State University

Str. Bukireva, 15, Perm, Russia, 614990, ruinmath@bk.ru

This article describes the development of a neural network system for recognition (object detection or detection) and classification of medical masks on people's faces (or its absence) on photo- or video-frames. The system allows you to determine with great accuracy whether there is a mask on the person's face, and whether it is correctly worn. With the help of the developed intelligent system conducted a study of the subject area, identified patterns of practical relevance.

**Keywords:** algorithms for object detection and classification, pattern recognition, artificial intelligence, neural network technology, detection, medical mask, faces.

УДК 004.032.26

## ОБУЧЕНИЕ АРХИТЕКТУРЫ YOLOV3 ДЛЯ ДЕТЕКТИРОВАНИЯ ОРУЖИЯ НА ВИДЕО С КАМЕР ВИДЕОНАБЛЮДЕНИЯ

*Гладкий Сергей Леонидович*

ООО «ВИПАКС+»,

614000, Россия, г. Пермь, ул. Краснова, 24, lrndlrnd@mail.ru

*Жуланов Вячеслав Николаевич*

Пермский национальный исследовательский политехнический университет,  
614990, Россия, г. Пермь, Комсомольский пр., 29, 590slavv@gmail.com

В статье представлено обучение глубокой нейронной сети для детектирования оружия с камер видеонаблюдения. Применены способы улучшения детектирующих свойств нейронной сети, одним из которых является добавление отрицательных примеров в датасет. Выполнен подсчёт метрик точности и процента ложных срабатываний. Проведено сравнение получаемых результатов на каждом из этапов работы, которое показало увеличение точности детектирования и сокращение количества ложных срабатываний. Полученная нейронная сеть позволяет автоматизировать оповещение сотрудников службы безопасности о попадании огнестрельного оружия в поле зрения камер видеонаблюдения, что ускоряет процесс реагирования на опасную ситуацию.

**Ключевые слова:** глубокое обучение, свёрточные нейронные сети, детектирование, оружие, видеонаблюдение.

**Введение.** Поставленная задача заключалась в том, чтобы обучить детектор оружия на основе нейронной сети (НС), способный работать в режиме реального времени, для использования его в системе видеонаблюдения. В свободном доступе было найдено несколько уже готовых решений, удовлетворяющих условию работы в реальном времени. Были проведены тесты точности детектирования, по итогам которых была выбрана лучшая НС. Данная НС (далее – базовая НС) основана на архитектуре You Only Look Once (YOLO).

YOLO относится к типу one – short detectors. Нейронные сети этого типа отличаются высокой скоростью работы за счёт однократного прохождения изображения через сеть, однако точность получаемых результатов может быть ниже [1-2].

Датасет, на котором была обучена базовая НС состоит из 8 тысяч изображений, по 4 тысячи на класс. Эксперты сходятся во мнении, что оптимальным количеством изображений на класс является 10 тысяч и более. Поэтому использованный датасет является недостаточным для полноценного обучения НС. К тому же объекты на этих изображениях находились вблизи и имели большой размер, тем самым не подходили для обучения НС работающей в системе видеонаблюдения, где объекты находятся вдали и имеют малый размер. Пример изображения из датасета представлен на рисунке 1.



**Рисунок 1. Изображения из датасета**

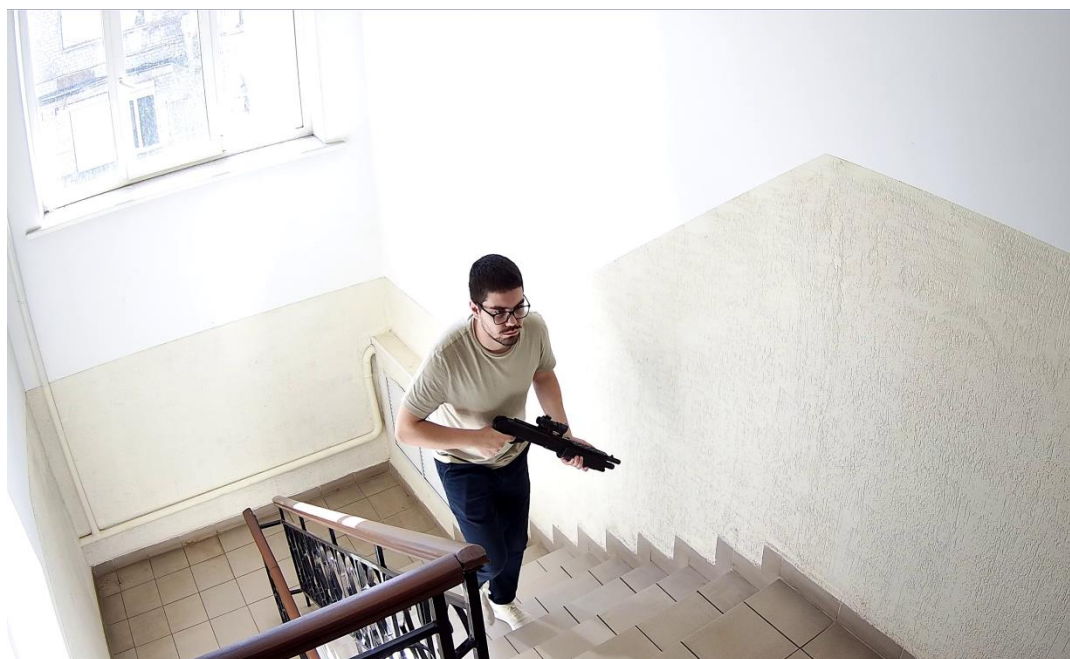
Проведённые испытания показали, что базовая НС имеет низкую точность детектирования и даёт большое количество ложных (false positive) срабатываний в условиях работы в системе видеонаблюдения.

Для увеличения точности детектирования и уменьшения количества ложных срабатываний, было принято решение создать новый датасет из видео с камер видеонаблюдения, на которых присутствует оружие. Тем самым новый датасет должен быть более подходящим для обучения НС, способной решать поставленную задачу.

**Основная часть.** Первым этапом обучения НС является создание датасета. Было принято решение разделять оружие на два класса: короткоствольное оружие (КО) (пистолеты) и длинноствольное оружие (ДО) (автоматы, ружья). Такое разделение является достаточным для определения наличия оружия на камере видеонаблюдения. Разделение оружия на большее количество классов в данном случае не представляется оптимальным, поскольку на изображениях с камеры видеонаблюдения значительно сложнее выделить мелкие детали различных моделей оружия. К тому же, при разделении на множество классов, задача создания датасета усложняется, так как для каждого класса необходимо создать свою выборку изображений.

Для создания датасета был составлен перечень различных поз, в которых обычно человек держит оружие. Также были собраны правдоподобные муляжи оружия двух классов. В итоге было отснято 69 видео в различных сценах, из которых получилось более 22 тысяч изображений с разрешением 1920x1080 и 2592x1944 пикселей, по 11 тысяч изображений на класс.

Во время обучения, НС основанная на архитектуре YOLO, принимает на вход изображения с разрешением 416x416 пикселей. Если подать кадр из видео с разрешением 1920x1080 или 2592x1944 пикселей, НС уменьшит изображение с сохранением соотношения сторон, а оставшееся свободное место заполнится чёрными пикселями. Вследствие этого, НС выделит меньше отличительных свойств детектируемого объекта. Чтобы этого не происходило, вокруг объекта на изображении вырезалась область размером 800x800 или 1200x1200 пикселей, если изображение имело разрешение 1920x1080 или 2592x1944 пикселей соответственно. Далее эта область сжималась до размера 416x416 пикселей. В таком формате область детектирования подавалась на вход НС.



**Рисунок 2. Исходное изображение 2592x1944 пикселей**

Это решение позволяет обучать НС с такими размерами объектов, которые будут встречаться ей при работе в системе видеонаблюдения, что, соответственно, должно положительно повлиять на показатели точности детектирования. Примеры изображений представлены на рисунках 2 и 3.



**Рисунок 3. Область детектирования 416x416 пикселей**

Вторым этапом было обучение новой НС. Оно проходило на основе базовой НС с применением переносимого обучения (transfer learning). Такой подход позволяет получить более высокие показатели точности детектирования за меньшее количество итераций [3].

Был произведён подсчёт и сравнение результатов тестирования с доверительным порогом (confidence threshold) равным 0.25. Результаты представлены в таблице 1.

Для подсчёта ложных срабатываний были собраны видео из реальных условий эксплуатации. Данные видео были получены с камер видеонаблюдения, установленных на входе в различных учреждениях, в условиях, максимально приближенных к реальным. Кадры из видео подавались в НС на детектирование. Процент ложных срабатываний от общего количества кадров был взят как один из показателей качества НС. Результаты представлены в таблице 1.

**Таблица 1**

**Результаты тестирования НС**

Нейросеть	Точность, % (КО)	Точность, % (ДО)	mAP, %	Ложные срабатывания, %
Базовая	58.14	76.42	69.46	5.58
Новая	97.50	97.25	97.38	15.55

Как видно из таблицы, точность детектирования у новой нейросети выше. Она лучше детектирует длинноствольное оружие и значительно лучше короткоствольное, по сравнению с базовой НС. Однако, при тестировании на ложные срабатывания новая НС показывает результат хуже, чем базовая НС.

С целью уменьшения количества ложных срабатываний, было применено дополнение датасета отрицательными примерами (negative samples) [4-5], то есть, изображениями без искомого объекта (оружия), но с объектами, на которых произошло ложное срабатывание (люди, ручные сумки, телефоны и т.д.). В качестве отрицательных примеров использовались кадры из видео без оружия, которые не были использованы для подсчёта процента ложных срабатываний.

Отрицательные примеры были добавлены в датасет, который использовался для обучения новой НС. Получившиеся результаты тестирования нейронных сетей на датасете с отрицательными примерами (при значении доверительного порога 0.25) представлены в таблице 2.

**Таблица 2**

**Результаты тестирования НС на датасете с отрицательными примерами**

Нейросеть	Точность, % (КО)	Точность, % (ДО)	mAP, %	Ложные срабатывания, %
Базовая	56.44	74.19	65.31	5.58
Новая	96.93	97.02	96.97	0.57

Как видно из таблицы, добавление отрицательных примеров в датасет значительно уменьшило процент ложных срабатываний.

**Заключение.** Поставленные задачи, согласно которым новая НС должна превзойти базовую НС по показателям точности детектирования и проценту ложных срабатываний, были выполнены с помощью создания нового датасета, выделения области детектирования с искомым классом объектов и добавления отрицательных примеров в датасет. Были проведены тесты, согласно которым новая НС имеет точность на 30% выше и в 9,8 раз меньше ошибается, по сравнению с базовой НС. Новая НС внедрена в аппаратно-программный комплекс интеллектуального видеонаблюдения *Domination*.

**Библиографический список**

1. Research of YOLO Architecture Models in Book Detection – <https://www.atlantis-press.com/article/125946016.pdf>, 2020. – 228 p.
2. Rosebrock A. Deep Learning for Computer Vision with Python. ImageNet Bundle. – [pyimagesearch.com](http://pyimagesearch.com), 2017. – 250 p.
3. Transfer Learning With Yolo V3, Darknet, and Google Colab – <https://medium.com/@cunhafh/transfer-learning-with-yolo-v3-darknet-and-google-colab-7f9a6f9c2afc>
4. Comparison of RetinaNet, SSD, and YOLO v3 for real-time pill identification, 2021. – 3 p. <https://bmcmmedinformdecismak.biomedcentral.com/counter/pdf/10.1186/s12911-021-01691-8.pdf>
5. Darknet FAQ – [https://www.coderun.ca/programming/darknet\\_faq/](https://www.coderun.ca/programming/darknet_faq/)

# YOLOV3 ARCHITECTURE TRAINING FOR WEAPONS DETECTION ON CCTV VIDEO

*Gladkiy Sergey L.*

VIPAX+ Ltd, 614000, Russia, Perm, Krasnova str. 24,  
lrndlrnd@mail.ru

*Zhulanov Vyacheslav N.*

Perm National Research Polytechnic University,  
29 Komsomolsky Ave., Perm, 614990, Russia  
590slavv@gmail.com

This paper presents the training of a deep neural network for detecting weapons from CCTV cameras. Ways to improve the detection properties of the neural network have been applied, one of which is to add negative examples to the dataset. The accuracy metrics and percentage of false alarms have been calculated. The comparison of the obtained results at each stage of the work was done and it showed the increase of detection accuracy and decrease of the number of false positives. The obtained neural network allows to automate the notification of security officers when a firearm is detected on a video surveillance camera. This speeds up the process of responding to the dangerous situation.

**Keywords:** deep learning, convolution neural networks, detection, weapons, surveillance.

УДК 004.93'12

## РАЗРАБОТКА ПРОГРАММЫ МОДЕЛИРОВАНИЯ АЛГОРИТМОВ СИСТЕМ РАСПОЗНАВАНИЯ ЛИЦ

*Липин Юрий Николаевич*

*Сторожев Сергей Александрович*

Пермский национальный исследовательский политехнический университет,  
614000, Россия, г. Пермь, ул. Комсомольский проспект, 29,  
ur-lip193530@yandex.ru

На рынке информационных продуктов предложено значительное количество проектов распознавания лиц. Все они решают в общем случае одну задачу – “Свой-Чужой” с закрытыми кодами, платные, не масштабируемые, со своими значениями ошибок 1-го и 2-го рода. Реальность такова, особенно, в учебном процессе студентов, что нужен подобный продукт, позволяющий оперативно реагировать на требования практики, например, распознавание близнецов, одного человека разного возраста и т.д. Предполагается, что авторы программ и алгоритмисты одно и то же лицо, знакомые с сущностью предмета исследования до мелочей.

**Ключевые слова:** распознавание, Золотое сечение, Фурье, вектор.



**Введение.** Для функционирования системы распознавания необходимо наличие двух субстанций – базы данных и самой программы распознавания. Базу данных можно создать с помощью своих или государственных фотолабораторий, также с помощью записанных видеопотоков с наличием специальных программ (Нами создана такая программа). Программы распознавания можно отнести к трем классам: 1) программы по следующему принципу работы, «пошлите фото и мы определим пол, возраст, найдем подобные лица»; 2) программы для режима пропуска; 3) программы для анализа и синтеза алгоритмов распознавания, работы для разных баз данных; для обучения студентов данного профиля. В данной статье рассматривается только третий вариант. Что касается алгоритмов и методов распознавания, то отмечаем следующее. Можно использовать алгоритмы зашитые в библиотеках (Numpy, OpenCv, Dlib, OpenFace), само программирование сводится к созданию интерфейсов вызова модулей, возможно программирование этих модулей на том или ином языке. Что касается математических методов распознавания собственных векторов, шаблонов лиц, метод графов, метод главных компонент (PCA), Active Appearance Models (ARM,ASM). с привлечением математики Габора, теория массового обслуживания, нейронные сети). Вырезание лиц из видеопотока в режиме on-line или из групповых фото (Метод Виолы-Джонсона, признаки Хаара, Алгоритм Лукаса-Канаде). Распознавания лиц (метод EigenFace, FisherFaces, LBP, FaceNet). Все эти методы сложны, трудоемки, требующие значительных вычислительных ресурсов. Мы используем свой подход, в частности, метод золотого сечения (МЗС), не требующий никаких сложных вычислений и вычислительных ресурсов. Заданный рисунок (фото) преобразуется фильтром Собеля, по черно-белому изображению программа находит ширину между бровями, равной ширине между внутренними уголками глаз, ширине носа, с 1.68 ширине губ, подбородок -уровень бровей, деленный пополам, равно положению носа и минус 0.32 от половины лица – положение губ. В итоге программа строит сетку 6X6 с получением 25 прямоугольников, по каждому из них, после преобразования Фурье, вычисляется вектор лица размером 25 значений с 26-м значением, как среднее по всему вектору [1-5].

**Результат работы программы.** Считанное фото из базы данных (левое фото на рис. 1) или из видео потока подвергается автоматической коррекции яркости и контрастности с выравниванием цветов по каналам RGB. По 26 значению вектора сортируются. Это позволяет при любом объеме базы для идентификации выбрать только шесть значений.

Дана база 24, размер входного потока Probe= 100,включающий 24 фото из базы данных, распознавание: N = 99..100% 24, N = 96..98% 39; N = 93..95% 20; N = 90..92% 9; N = 86..89% 5; N = 81..85% 2; N = 76..80% 1; N = 72..75% 0; N = 68..71% 0; N = 1..67% 0. Всего распознано по процентам чел.= 100; Время Распознавания на одно фото мсек. = 2; Время создания одного вектора и на одно фото мсек= 552; Все опознаны в 100%; Для данного примера ошибки FAR, FRR равны нулю.

Алгоритм создает массивы векторов базы данных и входного потока с записью на диск и позволяет при необходимости считывать эти базы, получать



результат идентификации с целью проведения дополнительного анализа. Это означает, что в базе данных хранятся только вектора с идентификаторами личности, база данных с фото отправляется в отдельный архив. Кроме того, имеется возможность идентифицировать отдельно любой объект из базы.

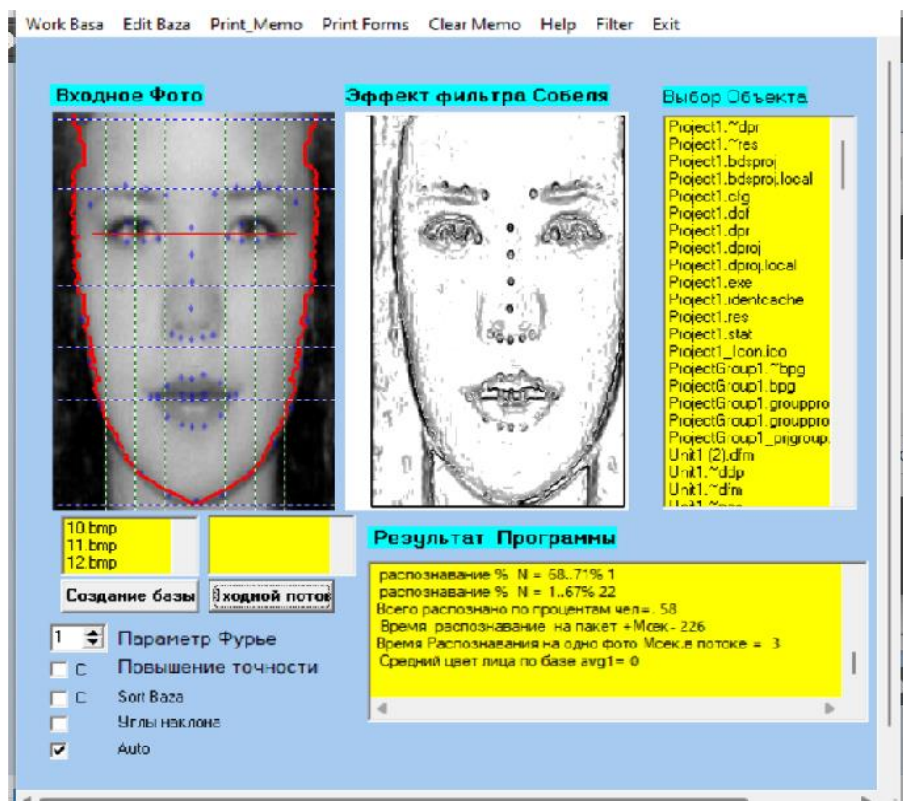


Рисунок 1. Результат работы программы

**Идентификация близнецов.** Распознавание идентичных близнецов с помощью изображений их лиц является сложной задачей в биометрии. Оно направлено на выявление важности экологических и генетических влияний на черты, фенотипы и расстройства. Исследование близнецов считается ключевым инструментом в поведенческой генетике и в других областях, от биологии до психологии. Исследования близнецов являются частью более широкой методологии, используемой в генетике поведения, которая использует все данные, которые являются генетически информативными – исследования братьев и сестер, исследования усыновления, родословная и т.д. Эти исследования были использованы для отслеживания черт, начиная от личного поведения до презентации тяжелых психических заболеваний, таких как шизофрения.

Близнецы являются ценным источником для наблюдений, потому что они позволяют изучать влияние окружающей среды и различный генетический состав: "идентичные" или монозиготные (MZ) близнецы имеют по существу 100% своих генов, что означает, что большинство различий между близнецами (например, рост, восприимчивость к скуке, интеллект, депрессия и т. Д.) связаны с опытом, который имеет один близнец, но не другой близнец [6]. "Братские" или дизиготные (DZ) близнецы разделяют только около 50% своих генов, так же, как и любой другой брат. Близнецы также разделяют многие ас-

пекты своей среды (например, среда матки, стиль воспитания, образование, богатство, культура, сообщество), потому что они рождаются в одной семье. Наличие данного генетического или фенотипического признака только у одного члена пары идентичных близнецов (так называемый диссонанс) обеспечивает мощное окно в воздействие окружающей среды на такой признак.

Близнецы также полезны для демонстрации важности уникальной среды (специфичной для одного близнеца или другого) при изучении представления признаков. Изменения в уникальной среде могут быть вызваны событием или происшествием, которое затронуло только одного близнеца. Это может варьироваться от травмы головы или врожденного дефекта, который один близнец выдержал, в то время как другой остается здоровым. Большие трудности вызывает разработка алгоритма определения идентификация близнецов, прожившую разную жизнь, они на изображении могут иметь форматы, которые алгоритмы идентифицируют, как имеющие высокие Far, свой идентифицируется, как чужой. Однако в криминологии они должны идентифицироваться, как одни и те же личности.

В [7] предложен довольно сложный метод, который использует слияние на уровне признаков, слияние на уровне баллов и слияние на уровне принятия решений с анализом главных компонент, гистограммой ориентированных градиентов и локальными экстракторами признаков бинарных шаблонов. В данной статье и в своей работе автор отказался от использования сложных математических алгоритмов и принял методику тестирования возникших идей на собственных программах продуктах.

Результат работы программы в обычном режиме. Дана база равная по одному представителю из пары, т.е. 6 фото и входной поток из 12 фото. Получен результат в виде, где идентифицированы только 6 близнецов из базы, остальные представители имеют малые проценты распознавания: распознавание % N = 99..100% 6; N = 96..98% 0; N = 93..95% 0; N = 90..92% 0; N = 86..89% 0; N = 81..85% 0; N = 76..80% 1; N = 72..75% 0; N = 68..71% 1; N = 1..67% 4. Данный пример без коррекции означает полное отсутствие распознавания. После корректировки данного алгоритма в режиме настройки можем получить другие результаты: распознавание % N = 99..100% 12; N = 96..98% 0; N = 93..95% 0; N = ; N = 68..71% 0; N = 1..67% 0, Всего распознано по процентам чел = 12.

Результат позволяет утверждать правомочность алгоритма в решении данной проблемы и не только в ней и в задаче идентификации людей разного возраста. Возникла идея, что не все части, элемента лиц близнецов изменяются в одинаковой степени и что-то остается почти без изменения. По результатам тестирования по разработанной программе эта предпосылка подтвердилась. Повышение процента распознавания достигается за счет равенства единицы всех значений вектора из 25, кроме отвечающих за глаза, брови нос, часть губ.

**Распознавание лиц разного возраста.** Распознавание лиц по возрасту – очень сложная международная проблема в области распознавания лиц. Как мы все знаем, изображения одного и того же человека в разном возрасте будут иметь очень большие различия, и эти различия серьезно повлияют на точность

распознавания лиц разных возрастов. Между несколькими лицами одного и того же человека в разном возрасте существуют очень существенные различия, что серьезно влияет на производительность существующих моделей глубокого распознавания лиц. Например, в [8] предложен новый алгоритм глубокого обучения, который разлагает глубинные особенности на возрастные компоненты и компоненты идентичности. Tencent AI Lab [9, 10] В данной статье используется алгоритм позволяющий повысить эффективность распознавания людей разного возраста не за счет использования сложнейшей математики, а за счет доказательности в меньшей изменчивости с возрастом области глаз и бровей, здравого смысла и эффекта рациональности. Для примера идентифицируются 12 лиц одного человека разных возрастов, из них только один имеет 100% как сам на себя. N = 99..100% 1; N = 96..98% 0; N = 93..95 0; N = 90..92% 2; N = 86..89% 6; N = 81..85% 2; N = 76..80% 1; N = 72..75% 0; N = 68..71% 0; N = 1..67% 0. Получены отрицательные результаты. С учетом коррекции алгоритма повышения распознаваемости получены результаты: N = 99..100% 1; N = 96..98% 9; N = 93..95% 2; N = 90..92% 0; N 86..89% 0; N = 81..85% 0; N = 76..80% 0; N = 72..75% 0; N = 68..71% 0; N = 1..67% 0.

**Влияние очков на процент распознавания.** Очевидно, необходимо ожидать влияние очков на распознавание (база фото представлена на рис. 2).



**Рисунок 2. Образцы с наличием очков**

Рассмотрим результаты: N = 99..100% 3; N = 96..98% 1; N = 93..95% 0; N = 90..92% 0; N = 86..89% 0; N = 81..85% 0; N = 76..80% 0; N = 72..75% 0; N = 68..71% 1; N = 1..67% 1

После коррекции: N=99..100% 4; N=96..98% 0; N= 93..95% 0; N = 90..92% 1; N= 86..89% 1; N= 81..85% 0; N = 76..80% 0; N = 72..75% 0; N = 68..71% 0; N = 1..67% 0. Вывод – светлые очки на результат распознавания влияют очень незначительно.

**Влияние углов наклона на распознавание.** Большое влияние на коэффициент распознавания личности по фотопортрету оказывает ракурс фотосъемки. В большинстве исследований и работ накладываются ограничения на допустимый угол наклона и поворота головы. Создана база данных из 22 фото от 0..35 гр. В алгоритме имеется модуль определения углов наклона и их коррекции. Программа дает положительные результаты в пределах наклонов головы до 35 гр.

**Модуль StopApp создания базы фотографий.** Программа формирования базы фото выполнена на основании библиотеки Dlib в среде Python и в виде модуля (\*.exe) вставлена в Borland delphi7\_Sience\_Edition\_2000(Delphi\_For\_

Windows\_10\_25450639) с использованием модуля ShellApi в Uses. Изображения должны быть представлены в формате “.jpg”. Образцы фото могут иметь следующие форматы (см. рис. 3).



**Рисунок 3. Образцы для формирования базы фото**

Результат работы программы приведен на рисунке 4.



**Рисунок 4. Результат работы программы**

Исходные фото могут быть заданные в неверном разрешении. Программа позволяет изменять размер картинка на 200x300 пикселей для корректной работы алгоритма.

Программа имеет возможность работать с видеопотоком.

**Выводы.** В замечательной статье [11] и исходя из здравого смысла следует, что мозг человека создан не мозгом человека, а иными высшими неизвестными структурами, то в ближайшее время, а может и никогда не получиться возможности мозга по распознаванию перетранслировать в наши алгоритмические программные структуры. Однако на основании многолетнего опыта в этой проблеме, согласно [12] и иным публикациям некоторые выводы можно сделать. Нельзя использовать абсолютные значения точек лица, расстояния между пикселями, любителям разных сеток лиц -углов фигур, разных отношений площадей, полиномов кривых. Овал лица (см. рис.1.) может использоваться, но его использование, как отношение числа пикселей к 200x300 сразу снижает процент распознавания, поскольку очень трудно подобрать фильтр для всего множества различных лиц;

В предлагаемом алгоритме используется всеобщее природное (снежинка, лицо, тело человека – художники Леонардо и Дюрер) правило золотого сечения (ЗС). Алгоритм находит только горизонтальное расположение бровей (при углах наклона используется модуль возврата лиц в вертикаль), далее по правилу ЗС автоматом строится сетка 6x6 линий. На ширине лица на уровне глаз уместается 4 ширины глаз, линия носа- половина расстояния до бровей от подбородка, и.т. д., ширина носа, равная расстоянию между глазами, бровями и ширине глаз сразу дает ошибки. Гармоники Фурье берутся для вектора только большие по значению от половины по максимальному значе-

нию в каждой из 25 фигур (Своего рода фильтр шума). Кроме того модуль гармоника формируется в степенной зависимости и логарифмической зависимости от координат пикселей, вернее от освещенности.

### **Библиографический список**

1. Аракелян Г.Б. Математика и история золотого сечения. -М.: Логос, 2014, 404 с. -ISBN 978-5-98704-663-0.
2. Бендукидзе А.Д. Золотое сечение // Квант. -1973. -№8.
3. Васютинский, Н.А. Золотая пропорция / Н.А. Васютинский. – Москва : Молодая гвардия, 1990. – 240 с. – (Эврика).
4. Власов В.Г. Золотое сечение, или Божественная пропорция / В.Г. Власов // Новый энциклопедический словарь изобразительного искусства: в 10 т. Т.3. – СПб.: Азбука-Классика, 2005. – С. 725–732.
5. Мазель Л., Опыт исследования золотого сечения в музыкальных построениях в свете общего анализа форм, "Музыкальное образование", 1930, № 2.
- А. Росс, К. Нандакумар и А. К. Джайн, Справочник по Мультибиометрии (Springer-Verlag, 2006).
6. А. Болотникова, Х. Демирель, Г. Анбарджафари Система распознавания лиц на основе ансамбля в реальном времени для гуманоидов НАО с использованием локального бинарного шаблона. Аналоговый интеграл. Circ. Sig. Процесс. 92(3), 1–9 (2017).
7. Габитоскопия и портретная экспертиза : учебник / А. М. Зинин, И. Н. Подволоцкий ; под ред. Е. Р. Российской — М. : Норма: ИНФРА-М, 2017.
8. Зинин, А. М. Внешность человека в криминалистике и судебной экспертизе : монография / А. М. Зинин. — М. : Юрлитинформ, 2015.
9. Зинин, А. М. Фоторобот: монография / А. М. Зинин. — М. : Юрлитинформ, 2015.
10. Распознавание лиц человеческим мозгом: 19 фактов, о которых должны знать исследователи компьютерного зрения. [Электронный ресурс]. Режим доступа: <https://habr.com/ru/post/136483/>
11. Gee A. Estimating gaze from a single view of a face / A. Gee, R. Cipolla // Image and Vision Computing. — 1994. — Vol. 12, № 10. — P. 639 — 647. Хайкин С. Нейронные сети : полный курс.

## **DEVELOPMENT OF A PROGRAM FOR SIMULATION OF ALGORITHMS OF FACE RECOGNITION SYSTEMS**

*Lipin Yury Nikolaevich*

*Storozhev Sergey Alexandrovich*

Perm National Research Polytechnic University

Komsomolsky prospect, 29, Perm, Russia, 614000, h1941@yandex.ru

A significant number of face recognition projects have been proposed on the information product market. All of them generally solve one problem – "Friend or Foe" with closed codes, paid, not scalable, with their own error values of the 1st and

2nd kind. The reality is, especially in the educational process of students, that a similar product is needed that allows you to quickly respond to the requirements of practice, for example, recognition of twins, one person of different ages, etc. It is assumed that the authors of programs and algorithms are one and the same person, who are familiar with the essence of the subject of research to the smallest detail.

**Keywords:** recognition, golden section, Fourier, vector.

УДК 004.032.26

## НЕЙРОСЕТЕВАЯ СИСТЕМА ОПРЕДЕЛЕНИЯ СКЛОННОСТИ ЧЕЛОВЕКА К СЕРИЙНЫМ УБИЙСТВАМ

*Аникина Инесса Владимировна, Бурылова Александра Алексеевна*

Пермский государственный национальный исследовательский университет,  
КМБ. 614990, Россия, г. Пермь, ул. Букирева, 15, anikinaiv710@gmail.com,  
burylova.alexandraa@yandex.ru .

В статье представлено описание разработки нейросетевой системы для выявления склонности человека к совершению серии убийств. Система позволяет с большой точностью определить является человек серийным убийцей или нет.

**Ключевые слова:** криминалистика; серийный убийца; маньяк; насилие; раскрытие; преступление; искусственный интеллект; нейронная сеть; математическая модель.

**Введение.** В современном обществе среди криминалистов нет единого мнения в вопросе выбора параметров, позволяющих выделить признаки, отличающие маньяка-убийцу от нормального человека, что затрудняет создание эффективных компьютерных программ, предназначенных для использования в следственной практике. В статье описан опыт разработки нейронной сети, обучаемой на данных известных серийных убийц США XX-XXI веков, а также публичных лиц США, включающих их биологические, социальные и психологические параметры. Выполнена оценка погрешности разработанной нейросетевой математической модели. Показана ее адекватность и проведено исследование, в результате которого получена сравнительная оценка степени влияния различных факторов на результат моделирования — предрасположенность человека к совершению серии убийств.

Был выбран следующий набор показателей для нейросетевой системы:

**X1**— пол: 1 — мужской, 2 — женский; **X2**— психическое расстройство: 1 — сильное, 2 — среднее, 3 — отсутствует; **X3** — факт насилия в детстве: 1 — физическое, 2 — сексуальное, 3 — психологическое, 4 — физическое, психологическое и сексуальное 5 — отсутствует; **X4**— наличие родителей: 1 — отсутствует мать (или рано лишился матери), 2 — отсутствует отец



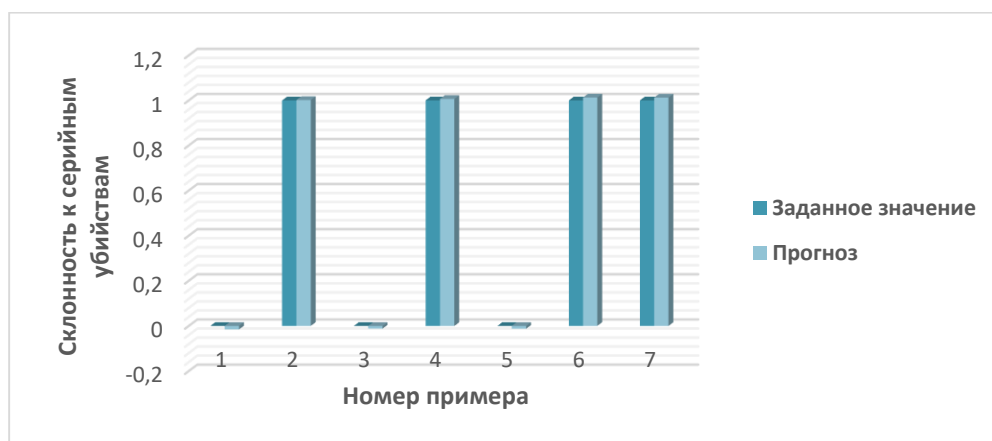
(или рано лишился отца), 3 — отсутствуют оба родителя (или рано лишился обоих родителей), 4 — родители есть; **X5** — социальный статус родителей: 1 — рабочий класс, 2 — интеллигенция, 3 — бизнесмены, 4 — бедные; **X6** — психологическое расстройство у родителей: 1 — у матери, 2 — у отца, 3 — у обоих родителей, 4 — отсутствует; **X7** — склонность к алкоголизму у родителей: 1 — у отца, 2 — у матери, 3 — у обоих родителей, 4 — склонности к алкоголизму нет; **X8** — семейный статус: 1 — женат (замужем), 2 — разведен (-а), 3 — гражданский брак, 4 — холост (не замужем); **X9** — дети: 1 — есть, 2 — нет; **X10** — ранее судим: 1 — да, 2 — нет.

Ранее было уже проведено подобное исследование [4]. Мы решили добавить к исследованным ранее параметрам новые, которые на наш взгляд также могут повлиять на склонность человека к совершению серии убийств. Нами были добавлены следующие параметры: факт насилия в детстве, наличие психического расстройства у родителей и наличие судимости ранее.

Обучающее множество было собрано вручную с интернет-ресурсов [1, 5]. Объем итогового множества включает в себя 41 пример. Данное множество было разделено на обучающее (83%) и тестирующее (17%). Перед переходом к проектированию и обучению нейросетевой модели, была выполнена очистка исходного множества от противоречивых примеров, выбросов.

Проектирование, обучение, тестирование нейронной сети выполнялись с помощью программы «Нейросимулятор 5» [2] по методике [3]. В результате оптимизации спроектированная нейронная сеть представляла собой персептрон, который имеет десять входных нейронов, один выходной и один скрытый слой с тремя нейронами.

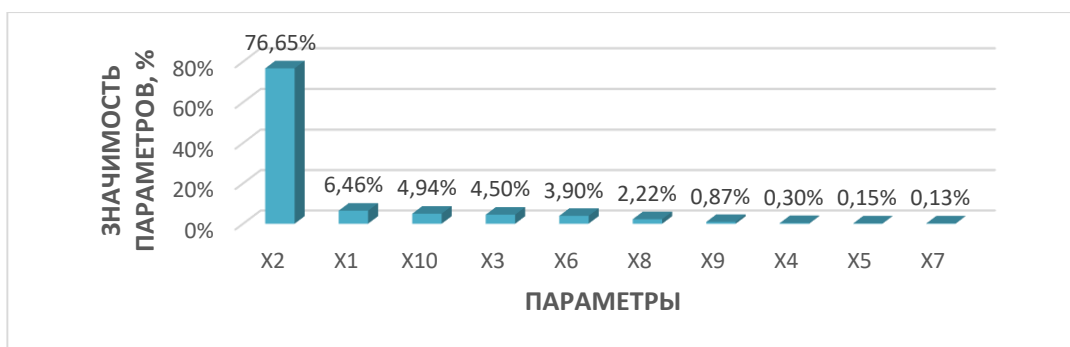
Для оценки корректной работы спроектированной нейронной сети использовалось тестирующее множество, состоящее из 7 примеров. Средняя относительная ошибка тестирования составила 1,04%. На рисунке 1 представлена гистограмма, демонстрирующая разницу между фактической и прогнозируемой нейросетью склонностью к серийным убийствам.



**Рисунок 1. Результат тестирования нейронной сети**

Из результатов, изображенных на рисунке 1, можно сделать вывод об адекватной работе спроектированной нейронной сети.

Оценка значимости входных параметров была выполнена с помощью программы «Нейросимулятор 5», результат отображен на рисунке 2.



**Рисунок 2. Значимость входных параметров нейросетевой модели**

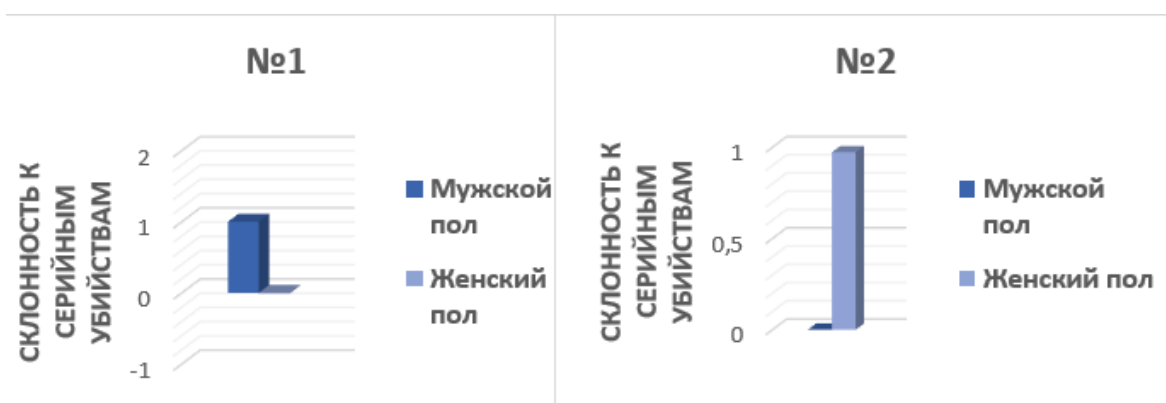
Как видно из рисунка 2, наиболее значимыми параметрами являются степень психического расстройства человека, пол человека, был ли человек ранее судим, факт насилия в детстве и наличие психического расстройства у родителей. Добавленные нами параметры (X10, X3, X6) являются достаточно значимыми в данном исследовании. Наиболее влиятельным параметром оказалась степень психического расстройств человека.

Далее было проведено исследование полученных зависимостей между входными параметрами и склонности к серийным убийствам. Исследование производилось с помощью метода «замораживания» [3], суть которого заключается в варьировании значения одного параметра и фиксировании значений всех других параметров. Для этой цели был отобраны примеры с наименьшими отклонениями от начальных данных:

**Таблица 1**

**«Примеры с наименьшими отклонениями от начальных данных»**

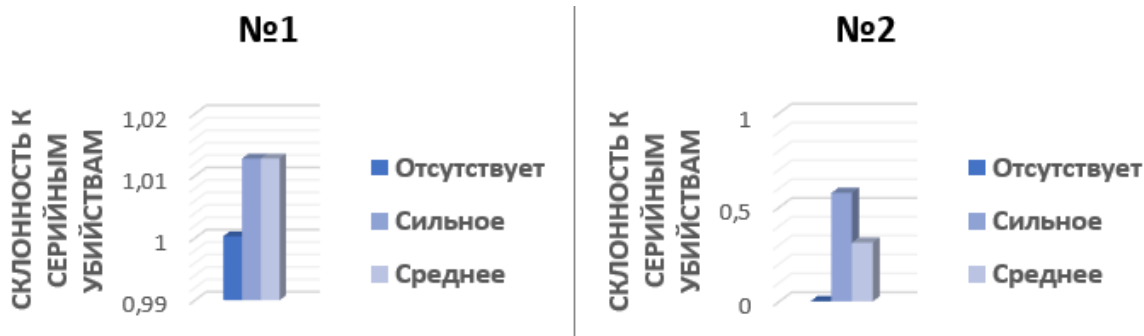
	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	Результат прогноза
№1	1	3	5	4	1	4	4	4	2	1	1,0003
№2	1	3	1	4	3	4	4	1	1	1	0,0014



**Рисунок 3. Зависимость склонности к совершению серии убийств от пола**

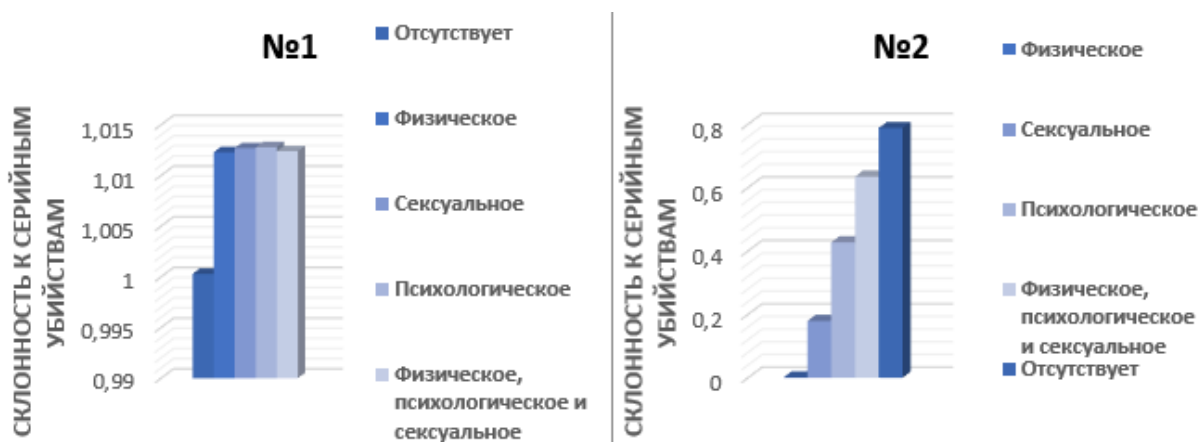
В данных примерах пол кардинально меняет результат прогноза.





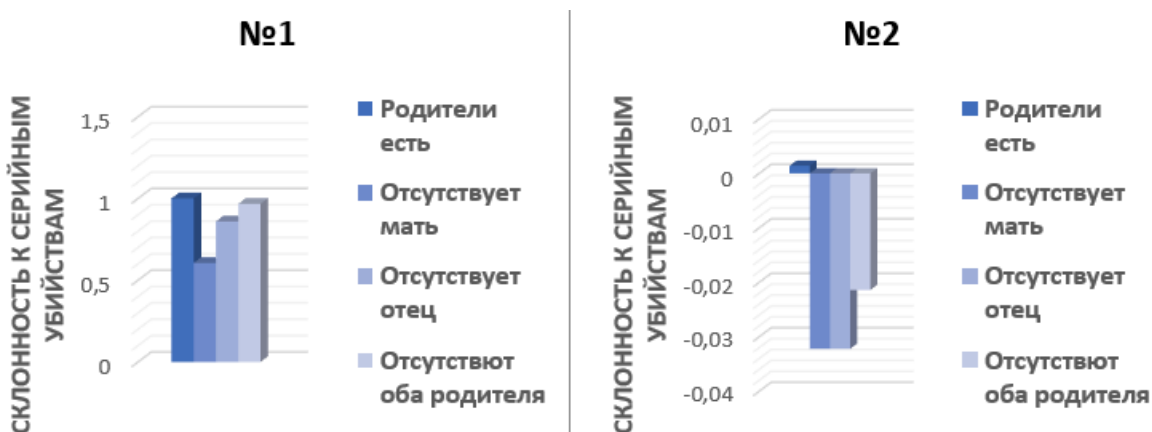
**Рисунок 4. Зависимость склонности к совершению убийств от степени психологического расстройства**

В данных примерах от степени психологического расстройства достаточно сильно меняет результат прогноза для 2 примера.



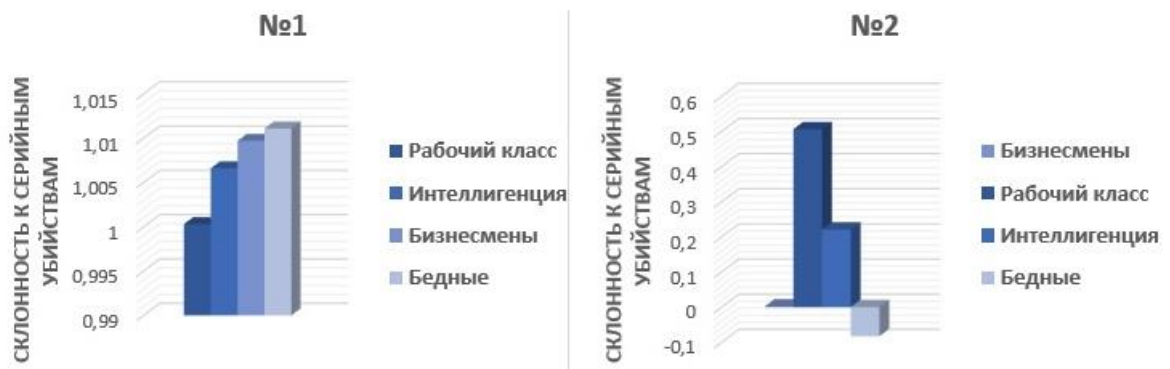
**Рисунок 5. Зависимость склонности к совершению убийств факта насилия в детстве**

В данных примерах вид насилия или его отсутствие достаточно сильно меняет результат прогноза для 2 примера.



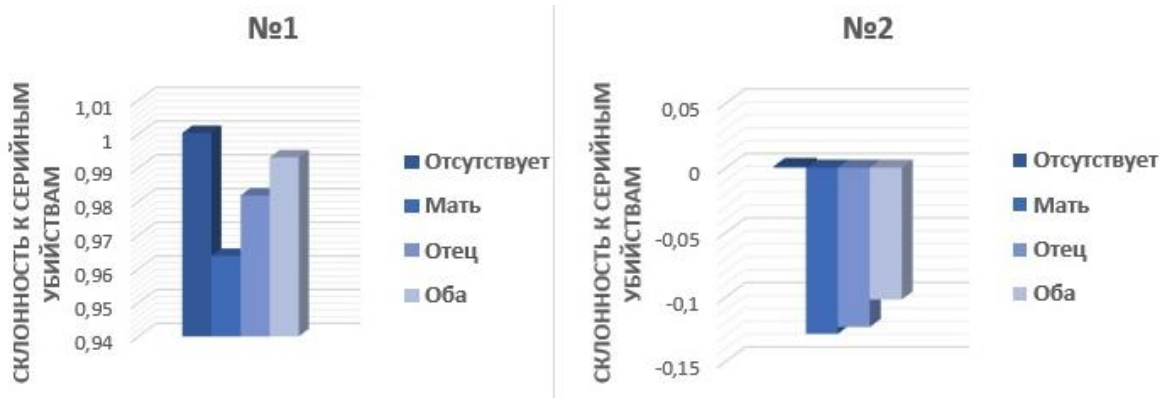
**Рисунок 6. Зависимость склонности к совершению убийств от наличия родителей**

В данных примерах наличие родителей сильно меняет результат прогноза для 1 примера.

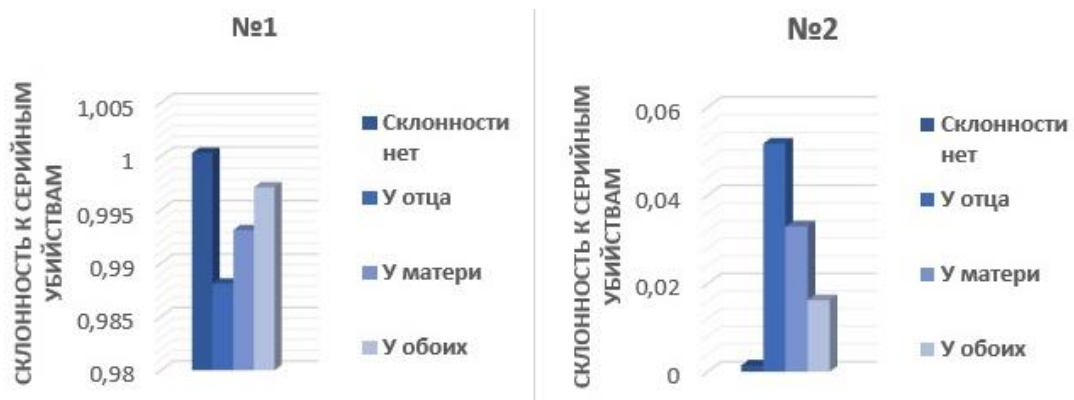


**Рисунок 7. Зависимость склонности к совершению убийств от социального статуса родителей**

В данных примерах социальный статус родителей достаточно сильно меняет результат прогноза для 2 примера.



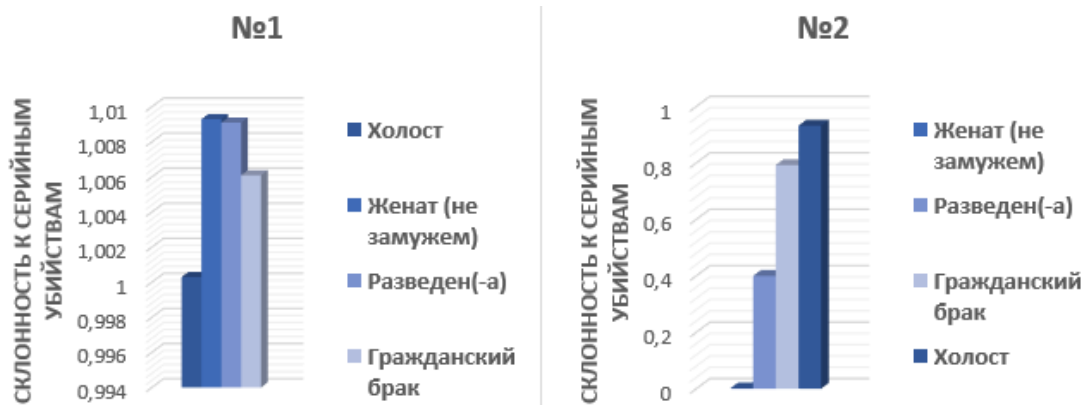
**Рисунок 8. Зависимость склонности к совершению убийств от наличия психического расстройства у родителей**



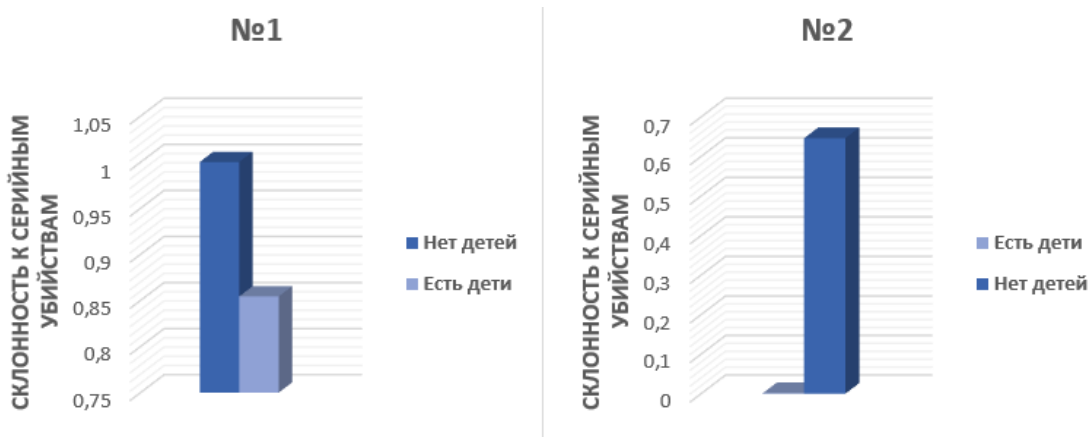
**Рисунок 9. Зависимость склонности к совершению убийств от наличия склонности к алкоголизму у родителей**

В данных примерах семейный статус достаточно сильно меняет результат прогноза для 2 примера.

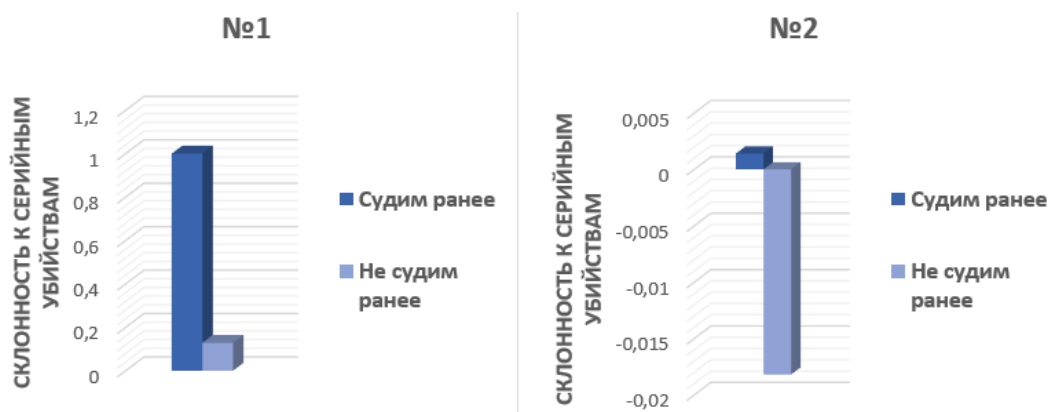
В данных примерах вид насилия или его отсутствие достаточно сильно меняет результат прогноза для обоих примеров.



**Рисунок 10. Зависимость склонности к совершению убийств от семейного статуса**



**Рисунок 11. Зависимость склонности к совершению убийств от наличия детей**



**Рисунок 12. Зависимость склонности к совершению убийств от наличия судимости**

Полученные результаты исследований не противоречат реальности, что подтверждает, что спроектированную нейронную сеть можно считать пригодной для прогнозирования склонности к серийным убийствам.

**Заключение.** Построена система нейросетевого прогнозирования склонности к серийным убийствам. Спроектированная нейросетевая модель учитывает 10 параметров: пол, степень психологического расстройства, факт насилия в детстве, наличие родителей, социальный статус родителей, нали-

чие психологического расстройства у родителей, наличие склонности к алкоголизму у родителей, семейные статус, наличие детей и наличие судимости. Параметры добавленные нами, такие как факт насилия в детстве, наличие психологического расстройства у родителей и наличие судимости, показали свою эффективность и были выявлены как значимые. Методом сценарного прогнозирования построены графики зависимостей прогнозируемой склонности к серийным убийствам от изменения входных параметров. Применение такого набора параметров в модели позволяет с высокой точностью прогнозировать склонности к серийным убийствам.

### Библиографический список

1. Ru.wikipedia.org – свободная энциклопедия [Электронный ресурс]. Режим доступа: [https://ru.wikipedia.org/wiki/%D0%9A%D0%B0%D1%82%D0%B5%D0%B3%D0%BE%D1%80%D0%B8%D1%8F:%D0%A1%D0%B5%D1%80%D0%B8%D0%B9%D0%BD%D1%8B%D0%B5\\_%D1%83%D0%B1%D0%B8%D0%B9%D1%86%D1%8B\\_%D0%A1%D0%A8%D0%90](https://ru.wikipedia.org/wiki/%D0%9A%D0%B0%D1%82%D0%B5%D0%B3%D0%BE%D1%80%D0%B8%D1%8F:%D0%A1%D0%B5%D1%80%D0%B8%D0%B9%D0%BD%D1%8B%D0%B5_%D1%83%D0%B1%D0%B8%D0%B9%D1%86%D1%8B_%D0%A1%D0%A8%D0%90)
2. Черепанов Ф.М., Ясницкий Л.Н. Нейросимулятор 4.0. Свидетельство о регистрации программы для ЭВМ RUS 2014612546. Заявка № 2014610341 от 15.01.2014.
3. Ясницкий Л.Н., Сичинава З.И. Нейросетевые алгоритмы анализа поведения респондентов // Нейрокомпьютеры: разработка, применение. 2011. № 10. С. 59-64.
4. Ясницкий Л.Н., Ваулева С.В., Сафонова Д.Н., Черепанов Ф.М. Использование методов искусственного интеллекта в изучении личности серийных убийц // Криминологический журнал Байкальского государственного университета экономики и права. 2015. Т. 9. № 3. С. 423-430.
5. u.wikipedia.org – свободная энциклопедия [Электронный ресурс]. Режим доступа: <https://ru.wikipedia.org/wiki/>

## NEURAL NETWORK SYSTEM FOR DETERMINING A PERSON'S PROPENSITY TO SERIAL MURDER

*Anikina Inessa V., Burylova Alexandra A.*

Perm State University

Str. Bukireva, 15, Perm, Russia, 614990,

[anikinaiv710@gmail.com](mailto:anikinaiv710@gmail.com), [burylova.alexandraa@yandex.ru](mailto:burylova.alexandraa@yandex.ru)

The article describes the development of a neural network system to identify a person's propensity to commit a series of murders. The system allows you to determine with great accuracy whether a person is a serial killer or not.

**Keywords:** criminology; serial killer; maniac; violence; disclosure; crime; artificial intelligence; neural network; mathematical model.

# **ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ В СПОРТЕ**

## НЕЙРОСЕТЕВАЯ СИСТЕМА ОЦЕНКИ ЭФФЕКТИВНОСТИ БАСКЕТБОЛЬНОГО ИГРОКА

*Батаев Богдан Викторович*

Национальный исследовательский университет «Высшая школа экономики»  
614990, Россия, г. Пермь, ул. Бульвар Гагарина, 37а,  
bataev.bogdan@list.ru

В статье представлено описание разработки нейросетевой системы для оценки эффективности баскетбольного игрока. Система позволяет с большой точностью оценить способности игрока на основании данных его предыдущих выступлений. С помощью разработанной интеллектуальной системы проведено исследование предметной области, выявлены закономерности, имеющие практического значения.

**Ключевые слова:** искусственный интеллект, нейросетевые технологии, оценка эффективности, баскетбол.

**Введение.** Оценка эффективности игроков в каком-либо виде спорта является крайне сложной задачей, и баскетбол среди всех видов спорта представляет наибольшую сложность. Тем не менее, опыт Пермской научной школы искусственного интеллекта [1-3], убедительно показывает, что правильно натренированная нейросетевая система может добиться успешности значительной части прогнозов. Анализ и использование оценок эффективности игроков в игре может быть полезно для тренеров команд, частных лиц, букмекерских компаний для увеличения прибыли компании.

Изучая литературные источники, выяснилось, что работ на тему оценки эффективности игрока в баскетбол достаточно мало, но они есть. Так, Джон Холлинджер разработал свою систему рейтингов для оценки эффективности игроков в баскетбол PER (с англ. Player Efficiency Rating) [4]. Среди критериев он выделял следующие:  $tm$  – коэффициент превосходства игрока над своей командой;  $lg$  – коэффициент превосходства игрока над лигой в целом;  $min$  – количество сыгранных минут;  $3P$  – количество результативных 3-очковых бросков;  $FG$  – количество результативных бросков с игры;  $FT$  – количество результативных штрафных бросков;  $VOP$  – значение владения (в данном случае в сравнении с лигой в целом);  $RB$  – количество подборов:  $ORB$  – в атаке,  $DRB$  – в защите,  $TRB$  – всего,  $RBP$  – соотношение подборов в атаке и защите. Остальные сокращения – стандартные сокращения статистики НБА:  $AST$  – количество передач;  $TO$  – количество потерь;  $FTA$  – количество штрафных бросков;  $STL$  – количество перехватов;  $BLK$  – количество блокшотов;  $PF$  – количество персональных фолов;  $PTS$  – количество набранных очков. В данной работе автор даже учитывает преимущества команд, кото-

рые играют в быстром темпе, следовательно, больше владеют мячом и имеют больше возможностей для эффективных действий в лиге. Также учитывается время, проведенное в игре, чтобы верно оценивать эффективность игроков в запасе.

Но данная система рейтинга не является идеальной. PER в большей степени учитывает атакующие действия игрока, нежели действия в защите. 2 защитных статистических показателя (блок-шоты и перехваты) могут не отобразить реальной ценности игрока в защите и что PER не является абсолютно надёжным показателем эффективности игрока в защите. К примеру, Брюс Боуэн, включавшийся 8 раз в сборную всех звёзд защиты НБА, регулярно получал PER ниже 10.

Некоторые утверждают, что PER придаёт чрезмерный вес вкладу игрока в отведённое игровое время или относительно дублирующего состава команды, тем самым недооценивая слишком разносторонних для стартового состава игроков.

Также существует мнение, что PER стимулирует количество неэффективных бросков. По словам спортивного экономиста и аналитика Дэйва Берри, если каждое двухочковое попадание имеет стоимость 1,65 очка, трёхочковое попадание имеет стоимость 2,65 очка, а пропущенный бросок с игры стоит команде 0.72 очка, то игрок не получает минус к PER, если забросит хотя бы 30,4% двухочковых бросков и 21.4% трёхочковых бросков. Если игрок превышает данный процент реализации, а практически каждый игрок НБА набирает данный процент по двухочковым броскам, чем больше он бросает, тем выше значение. Таким образом, игрок может быть неэффективным снайпером и просто поднять свой PER за счёт большого количества бросков.

Основная цель настоящей работы заключается в сборе множества публичных данных о результатах игроков в баскетбол на основе их статистических показателей, а также создание и обучение нейросетевой модели на этих данных. Конечным результатом должна быть реализована нейросетевая система, способная оценивать эффективность игроков в баскетбол на основе собранных данных.

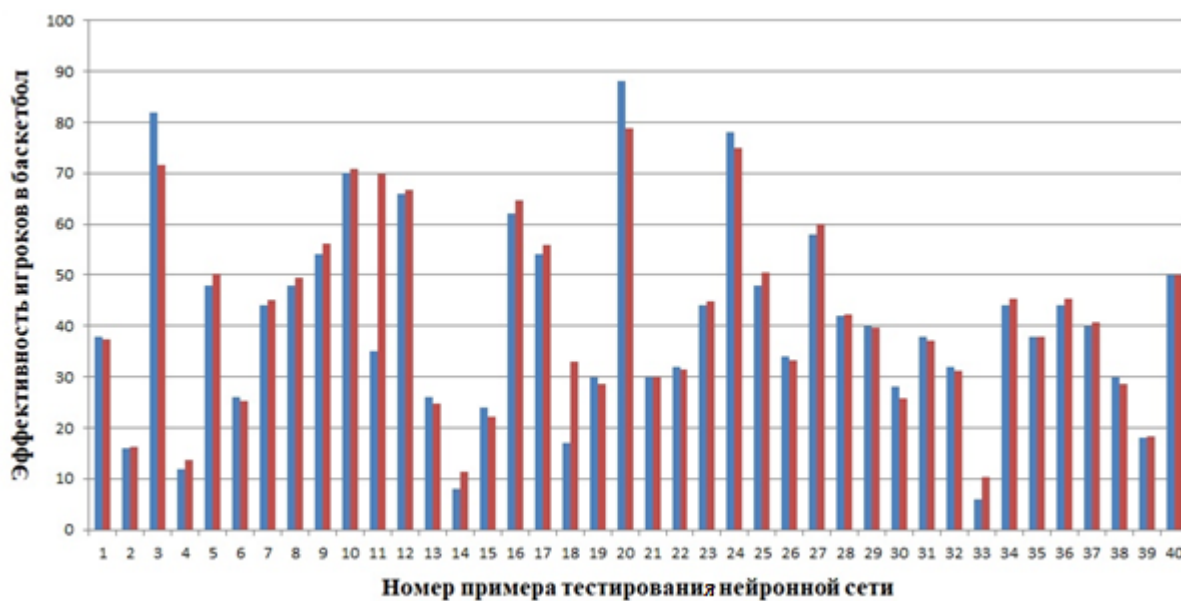
Для создания нейросетевой системы были выбраны следующие параметры: X1 – матч проводится дома или нет (0 или 1), X2 – победа в матче (0 или 1), X3 – очки команды в матче, X4 – очки противника в матче, X5 – минуты, проведенные игроком в игре, X6 – количество заброшенных мячей игроком, X7 – количество попыток забросить, X8 – количество заброшенных 3-х очковых в матче, X9 – количество попыток забросить 3-х очковый в матче, X10 – количество заброшенных штрафных в матче, X11 – количество попыток забросить штрафные в матче, X12 – количество подборов при нападении в матче, X13 – количество подборов при защите в матче, X14 – количество помощи в матче, X15 – количество отборов в матче, X16 – количество блоков в матче, X17 – потери мяча в матче, X18 – количество нарушений в матче, X19 – очки, принесенные игроком в матче. Выходной параметр – эффективность игрока в матче.

Обучающее множество было собрано вручную. Перед переходом к проектированию и обучению нейросетевой модели, была выполнена очистка

исходного множества от дубликатов. Таким образом, объем итогового множества включает в себя 133 примера. Данное множество было разделено на обучающее и тестирующее в соотношении 85% и 15% соответственно. Большая часть данных была собрана из интернет-ресурсов [4-5].

Проектирование, обучение, тестирование нейронной сети выполнялись с помощью программы «Нейросимулятор 5» [7] по методике [8]. В результате оптимизации, спроектированная нейронная сеть представляет собой персептрон, который имеет 19 входных нейронов, один выходной и один скрытый слой с одним нейроном.

Для оценки корректной работы спроектированной нейронной сети использовалось тестирующее множество, состоящее из 40 примеров.



**Рисунок 1. Результат тестирования нейронной сети**

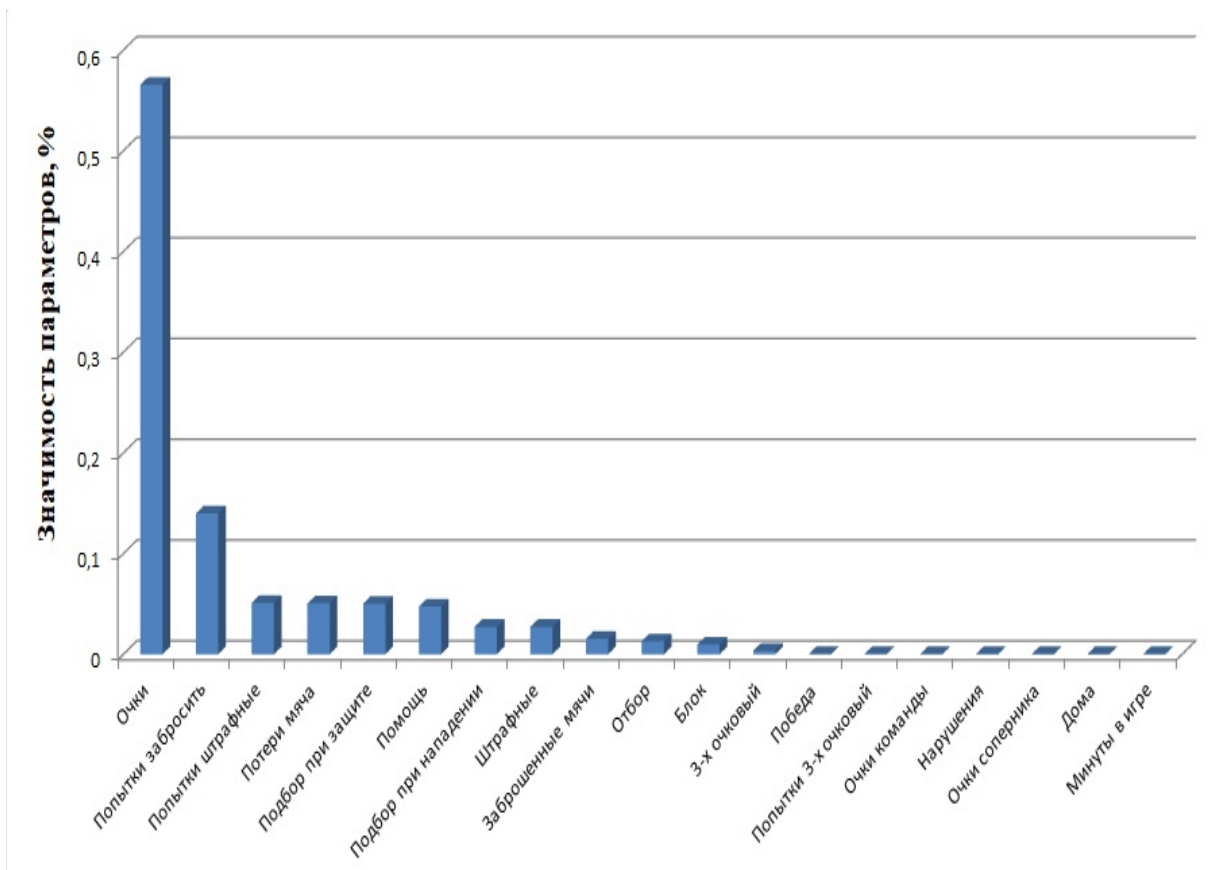
Средняя относительная ошибка тестирования составила 3.620%, что можно считать приемлемым результатом. На рисунке 1 представлена гистограмма, демонстрирующая разницу между настоящей и прогнозируемой нейросетью эффективностью игроков.

Из результатов, изображенных на рисунке 1, можно сделать вывод об адекватной работе спроектированной нейронной сети.

Оценка значимости входных параметров была выполнена с помощью программы «Нейросимулятор 5», результат отображен на рисунке 2.

Как видно из рисунка 2, наиболее значимыми являются набранные очки в матче, попытки забросить, попытки забросить штрафные, потери мяча, подбор при защите, помощь и т.д. Как и ожидалось, наиболее влиятельным параметром являются набранные игроком очки. Это объясняется тем, что в баскетболе крайне важно реализовывать владение мечом и доводить эти ситуации до результата, который отображается на табло.

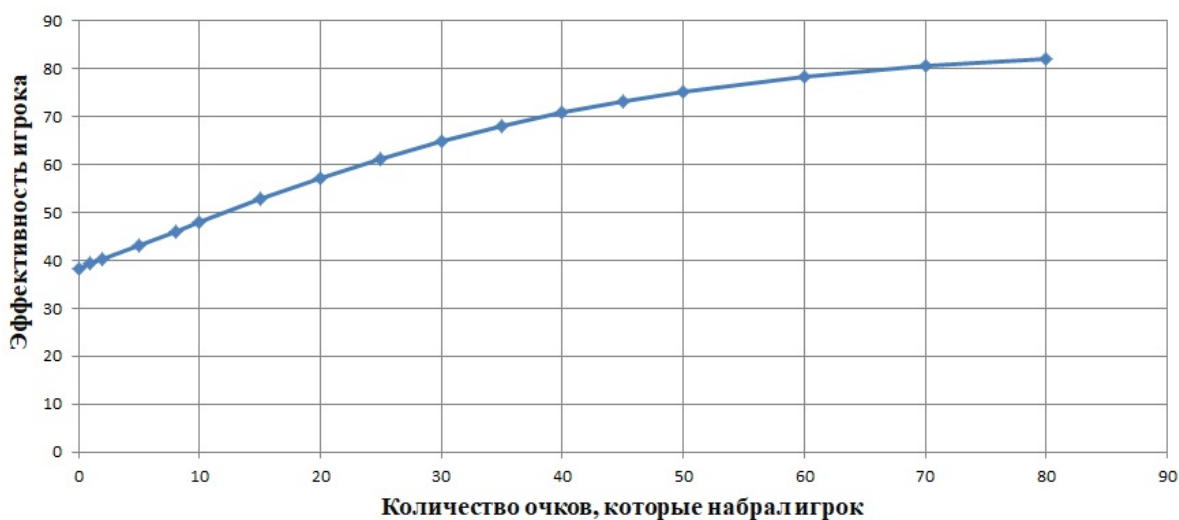




**Рисунок 2. Значимость входных параметров нейросетевой модели**

Далее было проведено исследование полученных зависимостей между входными параметрами и эффективностью игроков. Исследование производилось с помощью метода «замораживания» [1-3, 7], суть которого заключается в варьировании значения одного параметра и фиксировании значений всех других параметров.

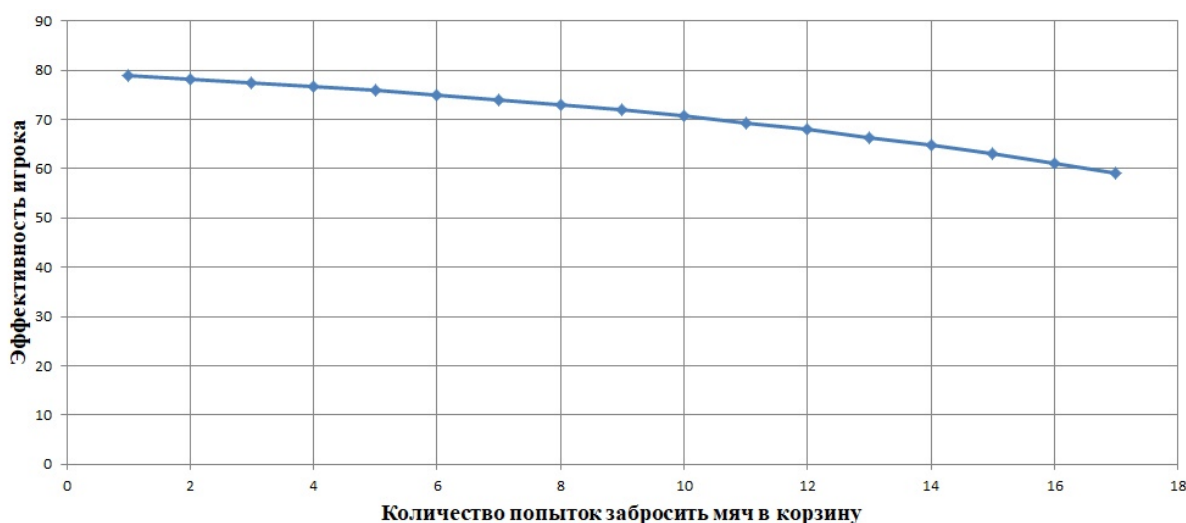
На рисунке 3 показан график зависимости эффективности игрока от количества очков, которые он принес команде в матче.



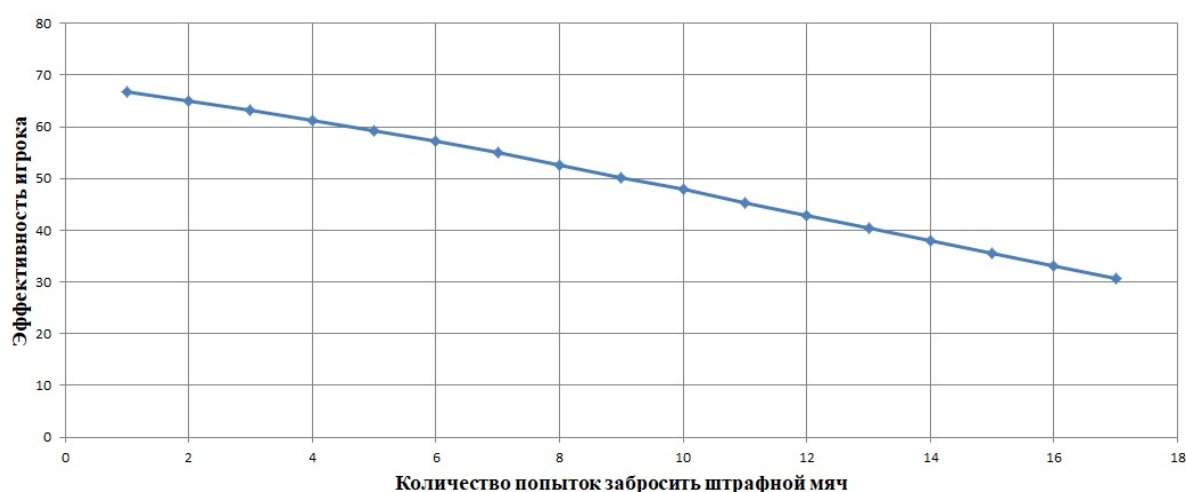
**Рисунок 3. Зависимость эффективности игрока от количества очков, которые он принес команде**

На рисунке 4 представлена зависимость эффективности игрока от количества попыток забросить мяч. Можно заметить, что эффективность игрока становится меньше от того, что у него возрастает количество неудачных попыток забросить мяч. Можно заметить, что эффективность игрока снижается немного сильнее после 10 попытки забросить мяч.

На рисунке 5 изображен график зависимости эффективности игрока от количества попыток забросить штрафной мяч с учетом того, что количество набранных очков не изменилось. Как видно из графика, чем больше попыток было у игрока, тем меньше у него эффективность при условии, что количество очков осталось тем же.

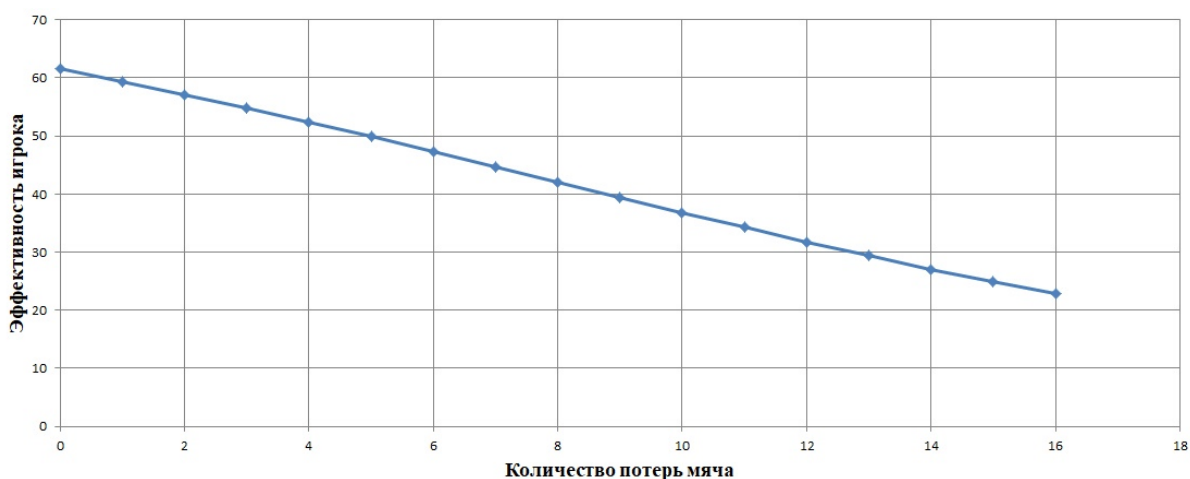


**Рисунок 4. Зависимость эффективности игрока от количества попыток забросить мяч**



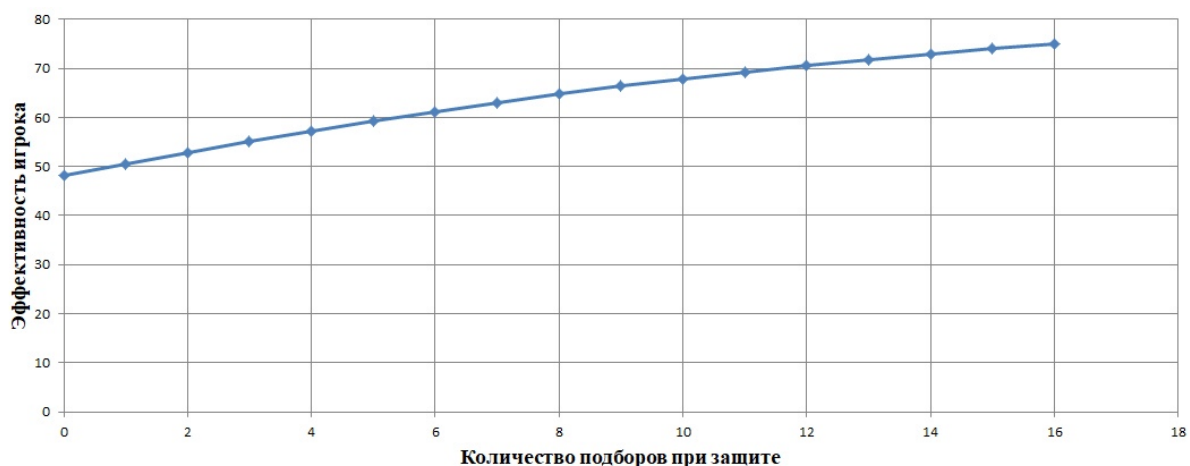
**Рисунок 5. Зависимость эффективности игрока от попыток забросить штрафной мяч**

На рисунке 6 представлена зависимость эффективности игрока от количества случаев потери владения мячом. Как видно из графика, это прямая зависимость – чем больше случаев потери мяча, тем меньше игрок получит очков эффективности.



**Рисунок 6. Зависимость шансов сборных на победу на чемпионате мира от попадания или непопадания команды в первую корзину**

На рисунке 7 представлена зависимость эффективности игрока от количества подборов мяча при защите кольца. Этот пример показывает, что эффективность игрока увеличится с ростом количества его подборов мяча при защите.



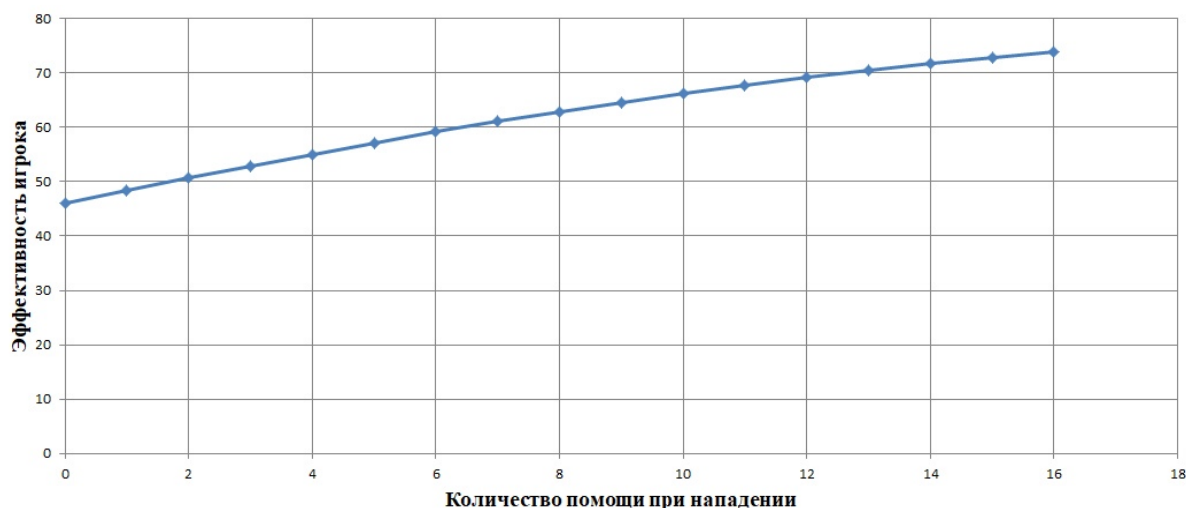
**Рисунок 7. Зависимость эффективности от количества подборов при защите**

На рисунке 8 представлена зависимость эффективности игрока от количества помощи при нападении. Можно сделать вывод, количество помощи при нападении положительно влияет на оценку эффективности игрока.

Полученные результаты исследований не противоречат реальности, что подтверждает, что спроектированную нейронную сеть можно считать пригодной для оценивания эффективности игроков в баскетбол.

**Заключение.** Построена система нейросетевой оценки эффективности баскетбольного игрока. Спроектированная нейросетевая модель учитывает 19 параметров: матч проводится дома или нет (0 или 1), победа в матче (0 или 1), очки команды в матче, очки противника в матче, минуты, проведенные игроком в игре, количество заброшенных мячей игроком, количество

попыток забросить, количество заброшенных 3-очковых в матче, количество попыток забросить 3-очковый в матче, количество заброшенных штрафных в матче, количество попыток забросить штрафные в матче, количество подборов при нападении в матче, количество подборов при защите в матче, количество помощи в матче, количество отборов в матче, количество блоков в матче, потери мяча в матче, количество нарушений в матче, очки, принесенные игроком в матче.



**Рисунок 8. Зависимость эффективности от количества помощи при нападении**

Методом сценарного прогнозирования построены графики зависимостей оценки эффективности баскетбольного игрока от изменения входных параметров. Применение такого набора параметров в модели позволяет с высокой точностью оценить эффективность баскетболиста.

### Библиографический список

1. Ясницкий Л.Н., Павлов И.В., Черепанов Ф.М. Прогнозирование результатов олимпийских игр 2014 года в неофициальном командном зачете методами искусственного интеллекта // Современные проблемы науки и образования. 2013. № 6. С. 985.
2. Ясницкий Л.Н., Vnukova O.V. Прогноз результатов олимпиады-2014 в мужском одиночном фигурном катании методами искусственного интеллекта // Современные проблемы науки и образования. 2014. № 1. С. 189.
3. Ясницкий Л.Н., Кировоса А.В., Ратегова А.В., Черепанов Ф.М. Прогноз результатов чемпионата мира-2015 по легкой атлетике методами искусственного интеллекта // Современные проблемы науки и образования. 2014. № 4. С. 624.
4. Basketball reference – Calculating PER. [Электронный ресурс]. Режим доступа: <https://www.basketball-reference.com/leaders/>
5. NBA – National Basketball Association. [Электронный ресурс]. Режим доступа: <https://www.nba.com/>

6. Черепанов Ф.М., Ясницкий Л.Н. Нейросимулятор 4.0. Свидетельство о регистрации программы для ЭВМ RUS 2014612546. Заявка № 2014610341 от 15.01.2014.

7. Ясницкий Л.Н. Введение в искусственный интеллект. М.: Издательский центр «Академия», 2005. – 176 с.

## NEURAL NETWORK SYSTEM FOR PREDICTING CHANCE OF WORLD CUP WINNERS

*Bataev Bogdan V.*

National Research University “Higher School of Economics”  
Str. Gagarin Boulevard, 37a, Perm, Russia, 614990, bataev.bogdan@list.ru

This article describes the development of a neural network system for predicting the chances of national teams to win the World Cup. The system allows predicting the chances of national teams to win the World Championships with great accuracy, based on the previous performances of teams at previous world championships. With the help of the developed intelligent system the study of the subject area has been carried out and the regularities which have practical importance have been revealed.

This article describes the development of a neural network system for evaluating the effectiveness of a basketball player. The system allows to assess the player's abilities with great accuracy based on the data of his previous performances. With the help of the developed intelligent system the study of the subject area has been carried out, patterns of practical importance were identified.

**Keywords:** artificial intelligence, neural network technology, efficiency rating, basketball.

УДК 004.032.26

## НЕЙРОСЕТЕВАЯ СИСТЕМА ПРОГНОЗИРОВАНИЯ ПОЛУЧЕНИЯ НАГРАДЫ САМОГО ЦЕННОГО ИГРОКА СЕЗОНА

*Соболев Дмитрий Алексеевич*

Национальный исследовательский университет «Высшая школа экономики»,  
614990, Россия, г. Пермь, ул. Студенческая, 38, dasobolev@edu.hse.ru

В статье представлено описание разработки нейросетевой системы для прогнозирования получения награды самого ценного игрока сезона в баскетбольной лиге «NBA». Система позволяет с большой точностью предсказать обладателя данной награды, главной персональной награды данной баскетбольной лиги, на основании данных о личных результатах по ходу продвижения в сезоне. С помощью разработанной интеллектуальной си-

стемы проведено исследование предметной области, выявлены закономерности, имеющие практическое значение.

**Ключевые слова:** нейронная сеть, искусственный интеллект, прогноз, самый ценный игрок, персептрон, тестовое множество.

**Введение.** Награда самого ценного игрока сезона, как не сложно догадаться является одной из самых престижных в главной баскетбольной лиге, а для кого-то и более ценной чем заветное чемпионство его команды. Но по каким же именно критериям определяется победитель данной заветной номинации. С целью получения данной информации и была проведена данная работа.

Целью данной работы стало создание системы прогнозирования получения награды самого ценного игрока сезона. Данная цель также подразумевает определение ключевых факторов успеха и их влияние на получение заветной награды. В данной статье представлен анализ существующих работ по этому направлению, выявлены тезисы, которые были применены при проектировании системы. В процессе работы в нейросимуляторе была разработана система, с помощью которой можно выполнять прогнозы получения данного трофея. В основе системы лежит нейронная сеть, обученная на результатах 8 сезонов самой престижной баскетбольной лиги в мире и в частности номинантов на получение данного титула. Итогом работы стала система, способная дать прогноз в отношении того, получит ли тот или иной игрок заветную статуэтку, а также показать влияние различных параметров на итоговый результат и выявить самые важные из них.

**Анализ литературы.** Киноиндустрия — это молодая, но в то же время перспективная и непрерывно растущая отрасль мировой экономики.

В качестве основных параметров для просмотра значимости игрок на претендентство на титул самого ценного игрока отмечают среднее количество блоков за игру [1], и ведь действительно данный атрибут является влиятельным по большому списку критериев. Во-первых современный баскетбол всегда был направлен на зрелищность происходящего, а неожиданно заблокированный мяч в красивом положении всегда радуется глаз зрителя, которые впоследствии и определяют результат. Вторым же момент важности данного элемента является сохранение преимущества своей команды что напрямую влияет на общую результативность команды, что в свою очередь ведет к большому количеству матчей, а соответственно большому количеству возможностей набирать заветные очки в других элементах игры.

Не трудно догадаться что, как и в любой другой спортивной игре количество забитых оппоненту очков влияет на оценку результативности игрока. Это также подмечает [2] в своей работе. В частности, отмечается процент забитых средних бросков ибо данный вид является наиболее уравновешенным между уровнем пользы и сложность выполнения. В связи с ускорением игры даже игроки позиций еще 10-15 лет предпочитающие забивать из под кольца в современных реалиях обязаны тренировать и использовать свои средние броски.

Таким образом, исследователи сходятся на том, что именно забитые средние броски, а также среднее количество блокировок мяча за игру высту-

пают в качестве наиболее важных критериев. Однако несмотря на все это можно продолжать делать попытки в поиске наиболее релевантных параметрах для оценки шансов на получение титула самого ценного игрока и продвигаться все глубже в изучение данной предметной области.

**Методика прогнозирования.** Для того, чтобы обучить нейронную сеть, были собраны данные о претендентах на получение данного трофея за 8 лет. Главным источником данных стал сайт с большим количеством наборов разнообразных данных Kaggle [3]. Из выбранного набора данных были выделены наиболее важные параметры для проведения дальнейшей работы над ними. Входные параметры представлены ниже:

- $x_1$  – Количество средних бросков за игру
- $x_2$  – Количество дальних бросков за игру
- $x_3$  – Количество штрафных бросков за игру
- $x_4$  – Минуты игрового времени
- $x_5$  – Среднее количество очков за игру
- $x_6$  – Среднее количество подборов за игру
- $x_7$  – Среднее количество передач за игру
- $x_8$  – Среднее количество отборов мяча за игру
- $x_9$  – Среднее количество блокирования мяча за игру
- $x_{10}$  – Средний процент средних бросков
- $x_{11}$  – Средний процент дальних бросков
- $x_{12}$  – Средний процент штрафных бросков
- $x_{13}$  – Количество очков MVP за сезон

Выходным параметром является получение награды самого ценного игрока сезона.

$d_1$  – 0 – не получит, 100 – получит.

Примеры были разбиты на три множества: обучающее, валидирующее и тестирующее. Для обучения нейронной сети были отобраны 69 примеров, валидирующее множество состоит из 16 примеров, для проведения проверки также было отобрано 19 примеров. Тестирующее множество позволяет проверить корректность работы нейросети. На рисунке 1 представлена часть обучающего множества, а на рисунке 2 – тестирующее множество.

	x1	x2	x3	x4	x5	x6	x7	x8	x9	x10	x11	x12	x13	D1
player	fga	fg3a	fta	mp_per_g	pts_per_g	trb_per_g	ast_per_g	stl_per_g	blk_per_g	fg_pct	fg3_pct	ft_pct	mwp_p	MVP
Dwyane Wade	16,20	2,70	8,60	37,10	25,50	6,40	4,60	1,50	1,10	0,50	0,31	0,76	240,00	0,00
Manu Ginobili	12,70	5,50	5,10	30,30	17,40	3,70	4,90	1,50	0,40	0,43	0,35	0,87	250,00	0,00
Derick Rose	19,00	3,30	7,70	34,00	27,60	10,20	6,00	2,00	3,20	0,63	0,65	0,77	1000,00	100,00
Blake Griffin	16,80	0,30	8,50	38,00	22,50	12,10	3,80	0,80	0,50	0,51	0,29	0,64	350,00	0,00
Rajon Rondo	9,90	0,60	1,90	37,20	10,60	4,40	11,20	2,30	0,20	0,48	0,23	0,57	276,00	0,00
Tony Parker	13,70	0,90	3,90	32,40	17,50	3,10	6,60	1,20	0,00	0,52	0,36	0,77	123,00	0,00
Chris Paul	11,60	2,30	4,80	36,00	15,90	4,10	9,80	2,40	0,10	0,46	0,39	0,88	58,00	0,00
LeBron James	16,00	3,20	4,50	35,20	27,30	10,20	8,30	1,90	3,20	0,62	0,64	0,79	1074,00	100,00
Kevin Durant	19,70	5,20	7,60	38,60	28,00	8,00	3,50	1,30	1,20	0,50	0,39	0,86	889,00	0,00
Chris Paul	14,80	3,60	5,00	36,40	19,80	3,60	9,10	2,50	0,10	0,48	0,37	0,86	385,00	0,00
Kobe Bryant	23,00	4,90	7,80	38,50	27,90	5,40	4,60	1,20	0,30	0,43	0,30	0,85	352,00	0,00
Tony Parker	14,80	1,00	4,70	32,10	18,30	2,90	7,70	1,00	0,10	0,48	0,23	0,80	331,00	0,00
Kevin Love	19,30	5,10	8,40	39,00	26,00	13,30	2,00	0,90	0,50	0,45	0,37	0,82	58,00	0,00

**Рисунок 1. Часть обучающего множества**



	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12	X13
player	fga	fg3a	fta	mp_per_g	pts_per_g	trb_per_g	ast_per_g	stl_per_g	blk_per_g	fg_pct	fg3_pct	ft_pct	mwp_p
Giannis Antetokounmpo	15,70	2,30	7,70	35,60	22,90	8,80	5,40	1,60	1,90	0,52	0,27	0,77	7,00
John Wall	18,40	3,50	6,80	36,40	23,10	4,20	10,70	2,00	0,60	0,45	0,33	0,80	7,00
Anthony Davis	21,20	4,50	7,30	36,80	28,60	10,20	5,50	1,40	3,20	0,62	0,65	0,79	958,00
Kevin Durant	16,50	5,00	6,20	33,40	25,10	8,30	4,80	1,10	1,60	0,54	0,38	0,88	2,00
DeMar DeRozan	20,90	1,70	8,70	35,40	27,30	5,20	3,90	1,10	0,20	0,47	0,27	0,84	1,00
James Harden	20,10	10,00	10,10	35,40	30,40	5,40	8,80	1,80	0,70	0,45	0,37	0,86	165,00
LeBron James	19,30	5,00	6,50	36,90	27,50	8,60	9,10	1,40	0,90	0,54	0,37	0,73	367,00
Anthony Davis	22,00	4,30	7,70	36,80	28,60	10,20	9,60	1,50	3,20	0,62	0,65	0,79	945,00
Damian Lillard	19,40	8,60	7,40	36,60	26,90	4,50	6,60	1,10	0,40	0,44	0,36	0,92	207,00
Russell Westbrook	21,10	4,10	7,10	36,40	25,40	10,10	6,30	1,80	0,30	0,45	0,30	0,74	76,00
Giannis Antetokounmpo	18,70	1,90	8,50	36,70	26,90	10,00	4,80	1,50	1,40	0,53	0,31	0,76	75,00
Kevin Durant	18,00	6,10	5,90	34,20	26,40	6,80	5,40	0,70	1,80	0,52	0,42	0,89	66,00
DeMar DeRozan	17,70	3,60	7,00	33,90	23,00	3,90	5,20	1,10	0,30	0,46	0,31	0,83	32,00
LaMarcus Aldridge	18,00	1,20	5,30	33,50	23,10	8,50	2,00	0,60	1,20	0,51	0,29	0,84	6,00
Jimmy Butler	15,60	3,40	7,20	36,70	22,20	5,30	4,90	2,00	0,40	0,47	0,35	0,85	5,00
Stephen Curry	16,90	9,80	5,90	32,00	26,40	5,10	6,10	1,60	0,20	0,50	0,42	0,92	5,00
Joel Embiid	16,80	3,40	7,40	30,30	22,90	11,00	3,20	0,60	1,80	0,48	0,31	0,77	4,00
Victor Oladipo	17,90	5,80	4,90	34,00	23,10	5,20	4,30	2,40	0,80	0,48	0,37	0,80	2,00

Рисунок 2. Тестирующее множества

Проектирование, обучение, тестирование нейронной сети выполнялись с помощью программы «Нейросимулятор 5» [4] по методике [5]. В результате оптимизации спроектированная нейронная сеть представляла собой персептрон, который имеет 13 входных нейронов, один выходной и один скрытый слой. Персептрон представлен на рисунке 3.

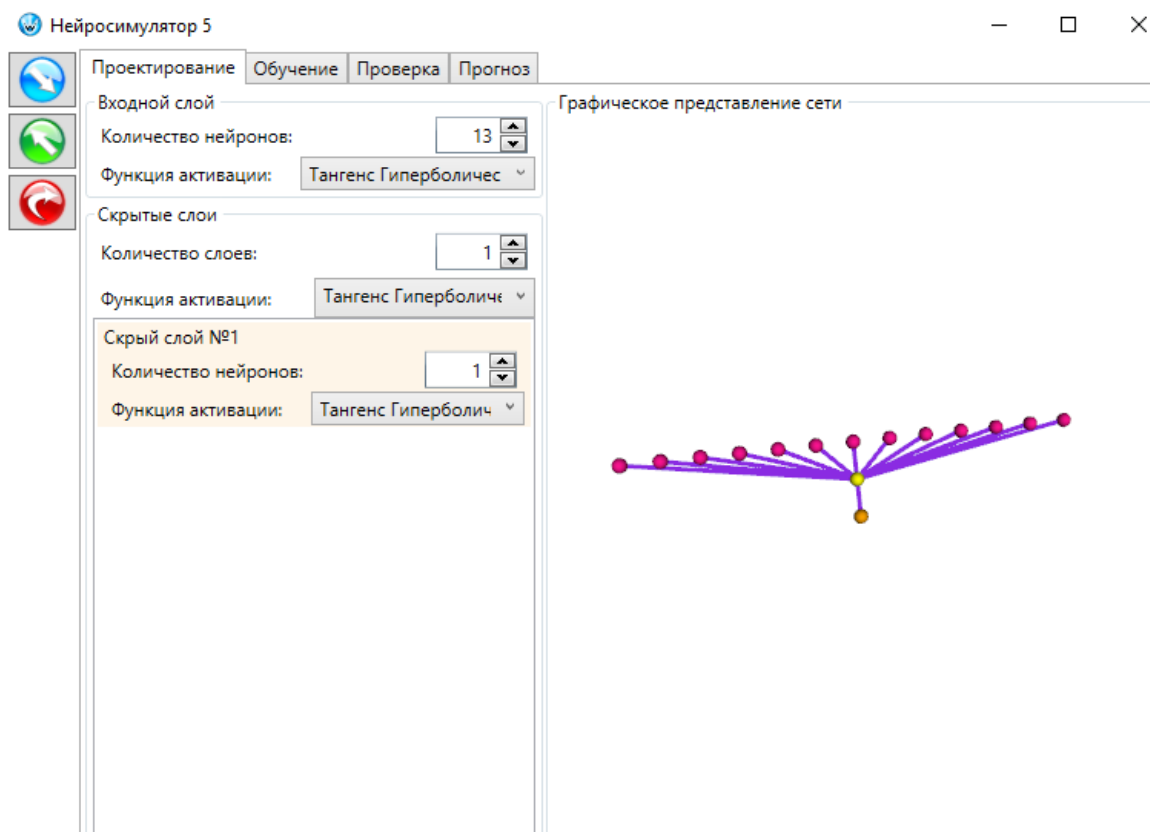
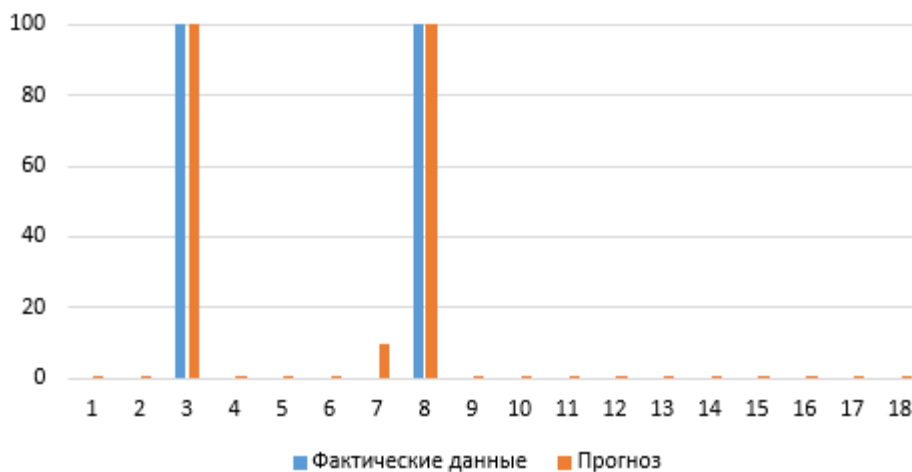


Рисунок 3. Персептрон

Для оценки корректной работы спроектированной нейронной сети использовалось тестирующее множество, состоящее из 13 примеров. На рисунке



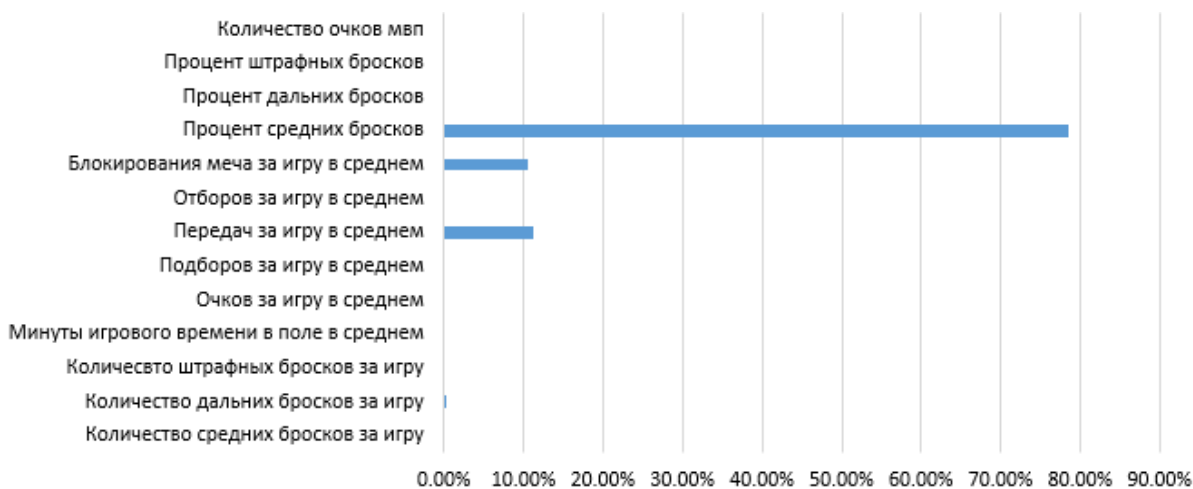
ке 4 представлена гистограмма, демонстрирующая разницу между фактическим и прогнозируемым нейросетью победителями в гонке за титул лучшего игрока сезона.



**Рисунок 4. Результат тестирования нейронной сети**

Из результатов, изображенных на рисунке 4, можно сделать вывод об адекватной работе спроектированной нейронной сети.

Оценка значимости входных параметров была выполнена с помощью программы «Нейросимулятор 5», результат отображен на рисунке 6

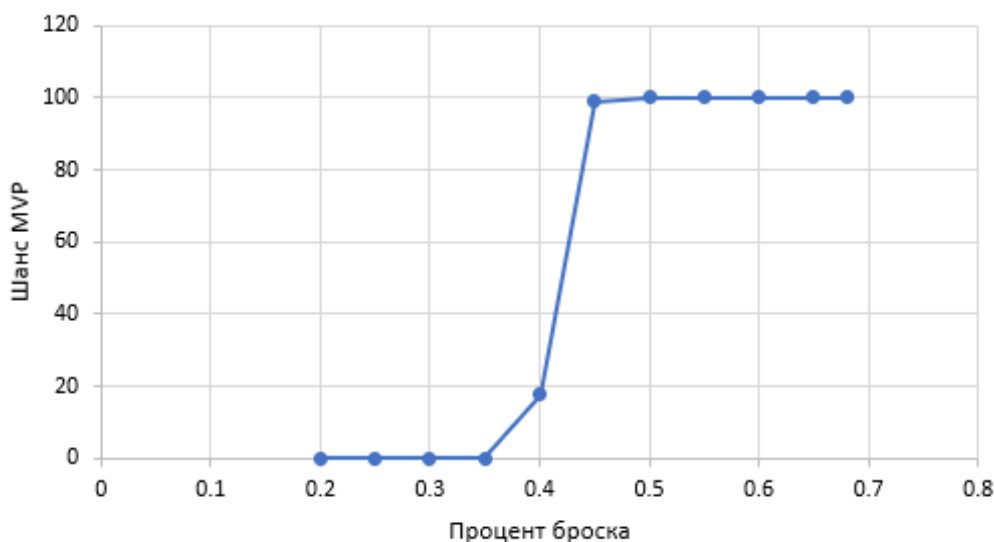


**Рисунок 5. Значимость входных параметров нейронной сети**

Как видно из рисунка 5, самым значимым параметром является процент средних бросков, после чего идут параметры среднего количества передач за игру и среднего количества блокирования мяча.

Применение методов нейросетевого моделирования позволяет не только выполнять прогнозы, но и ставить над моделями эксперименты, пытаться изменять эти прогнозы в наиболее благоприятную для спортсменов сторону. Так, варьируя входные параметры обученной нейронной сети и производя вычисления, можно сформировать конкретный список рекомендаций, способствующий получению данного трофея.

В качестве первого примера для рассмотрения обратимся к параметру процент средних бросков и одному из лучших и наиболее известных игроков Леброну Джеймсу. В его последнем сезоне получения завистного титула его процент средних бросков за сезон составил 68 процентов. Отредактируем данный параметр и проверим насколько именно он повлиял на его итоговый результат. Данный показатель был изменен от наименьшего к наибольшему, как и представлено на рисунке 6.



**Рисунок 6. Зависимость процента среднего броска на получение титула самого ценного игрока**

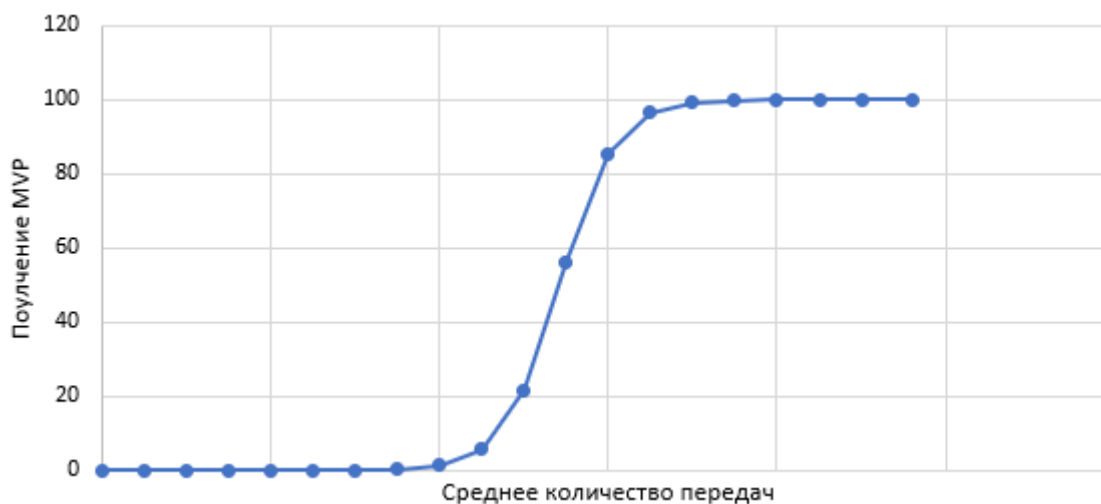
Исходя из рисунка 6 можно сделать вывод, что связь между процентом среднего броска и получением титула самого ценного игрока года обладает прямой зависимостью. В целом это подтверждается и самой моделью, построенной в нейросимуляторе, где параметр процента среднего броска за сезон был самым важным среди других входных.

Другим немаловажным параметром является процент передач за игру в среднем. Из представленного множества примеров наивысшим по данному показателю является один и последних обладателей данного титула Энтони Девис. Центровой команды Лос-Анджелес Лейкерс давно зарекомендовал себя как отличного пасующего игрока создавая моменты для реализации атак своими товарищами по команде. Давайте же посмотрим, как данный фактор повлиял на его получение звания самого ценного игрока сезона. Результаты представлены на рисунке 7.

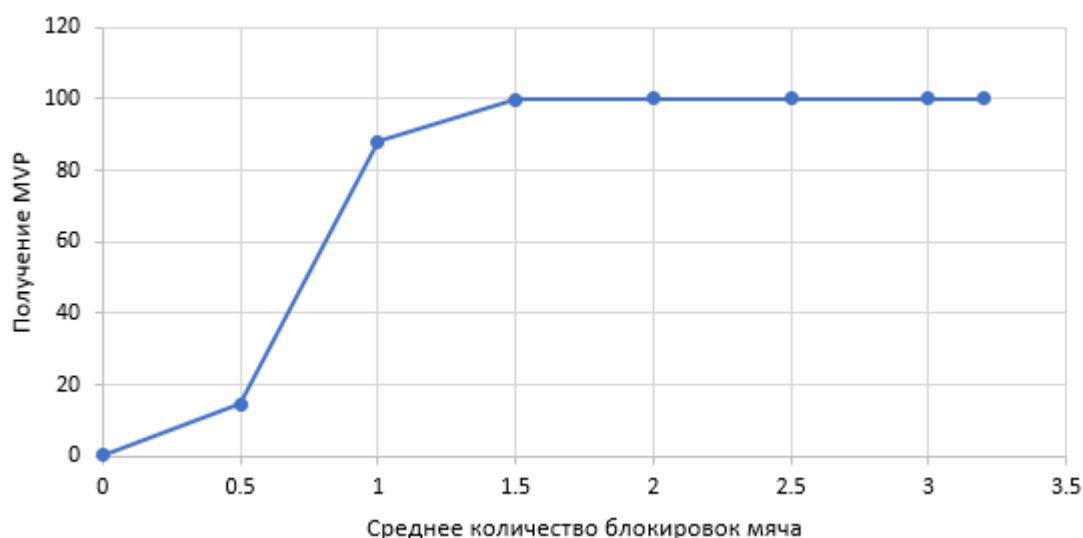
Исходя из рисунка видна прямая зависимость между двумя данными параметрами.

И последним из списка выделенных параметров является среднее значение блокирования мяча за матч. Так как с одинаковыми результатами по данному параметру титул получало несколько игроков, то я позволю себе самому выбрать одного из них. Мой выбор пал на одного из самых перспективных, но так и не раскрывших свой потенциал из-за травмы, игрока – Дерика Роуза. На его примере мы рассмотрим влияние данного па-

раметра на получения титула самого ценного игрока года. Результаты представлены на рисунке 8.



**Рисунок 7. Зависимость параметра процент передач на получение титула самого ценного игрока**



**Рисунок 8. Зависимость параметра блокирования за матч на получение титула самого ценного игрока**

Исходя из рисунка 8 можно сделать вывод, что связь между средним количеством блоков мяча за игру и шансом получения награды самый ценный игрок года присутствует.

Полученные результаты исследований не противоречат реальности, что подтверждает, что спроектированную нейронную сеть можно считать пригодной.

**Заключение.** Таким образом, в ходе данной работы была разработана нейросетевая модель, предназначенная для прогнозирования получения титула самого ценного игрока.

Данная модель может быть в дальнейшем применена для аналитического прогнозирования будущего обладателя статуса самого ценного игрока сезона.

## Библиографический список

1. Ясницкий Л.Н., Павлов И.В., Черепанов Ф.М. Прогнозирование результатов олимпийских игр 2014 года в неофициальном командном зачете методами искусственного интеллекта // Современные проблемы науки и образования. 2013. № 6. С. 985.
2. Ясницкий Л.Н., Кировоса А.В., Ратегова А.В., Черепанов Ф.М. Прогноз результатов чемпионата мира-2015 по легкой атлетике методами искусственного интеллекта // Современные проблемы науки и образования. 2014. № 4. С. 624.
3. Ясницкий Л.Н. Черепанов Ф.М. Нейросимулятор 4.0. Свидетельство о регистрации программы для ЭВМ RUS 2014612546. Заявка № 2014610341 от 15.01.2014.
4. Ясницкий Л.Н. Черепанов Ф.М. Искусственный интеллект. Методическое пособие. – М.: БИНОМ. Лаборатория знаний. 2012. – 216 с.
5. Ясницкий Л.Н., Внукова О.В., Черепанов Ф.М. Прогноз результатов олимпиады-2014 в мужском одиночном фигурном катании методами искусственного интеллекта // Современные проблемы науки и образования. 2014. [Электронный ресурс]. Режим доступа: <https://science-education.ru/ru/article/view?id=11339> (дата обращения: 08.03.2022).
6. Yasnitsky L.N. Algorithm for searching and analyzing abnormal observations of statistical information based on the Arnold – Kolmogorov – Hecht-Nielsen theorem // International Journal of Advanced Trends in Computer Science and Engineering. 2020. 9(2). Pp. 1814-1819. DOI: 10.30534/ijatcse/2020/139922020 <http://www.warse.org/IJATCSE/static/pdf/file/ijatcse139922020.pdf> Caio Brighenti. Using Data Science to Predict the Next NBA MVP. URL: <https://towardsdatascience.com/using-data-science-to-predict-the-next-nba-mvp-30526e0443da>
7. Прогнозирование следующего MVP НБА с помощью машинного обучения. [Электронный ресурс]. Режим доступа: <https://ichi.pro/ru/prognostirovanie-sledyuusego-mvp-nba-s-pomos-u-masinnogo-obucenia-108170294656885> (дата обращения: 15.03.2022).
8. Социальная сеть специалистов по обработке данных и машинному обучению. [Электронный ресурс]. Режим доступа: <https://www.kaggle.com/datasets/danchyy/nba-mvp-votings-through-history> (дата обращения: 15.03.2022).

## NEURAL NETWORK SYSTEM FOR MVP PREDICTION

*Sobolev Dmitry Alexeevich*

National Research University “Higher School of Economics”,  
614990, Russia, Perm, Str. Studencheskaya, 38, [dasobolev@edu,hse.ru](mailto:dasobolev@edu,hse.ru)

The article presents a description of the development of a neural network system for predicting the result of most valuable player contest. The system makes it possible to predict with great confidence the winner MVP race. With the

help of a study of the developed intellectual system, carried out within the subject area, observations of practical significance were identified.

**Key words:** neural network, artificial intelligence, forecast, most valuable player, perceptron, test task.

УДК 004.032.26

## НЕЙРОСЕТЕВАЯ СИСТЕМА ПРОГНОЗИРОВАНИЯ ПОБЕДИТЕЛЯ БАСКЕТБОЛЬНОГО МАТЧА НА ОСНОВЕ РЕЗУЛЬТАТИВНОСТИ ИГРОКА

*Белоусов Александр Александрович*

Пермский государственный национальный исследовательский университет,  
ПМИ. 614990, Россия, г. Пермь, ул. Букирева, 15, a-a-belousov@mail.ru

В статье представлено описание разработки нейросетевой системы для прогнозирования результатов баскетбольных матчей на основе результативности игрока. Система позволяет с большой точностью предсказать победителя следующего матча. С помощью разработанной интеллектуальной системы проведено исследование предметной области, выявлены закономерности, имеющие практической значение.

**Ключевые слова:** искусственный интеллект, нейросетевые технологии, прогнозирование, баскетбол, НБА.

**Введение.** В отличие от азартных игр, результаты матчей в спорте гораздо меньше зависят от случая. Поэтому прогнозирование победителя в спортивных соревнованиях является более точным.

При анализе литературных и интернет-источников выяснилось, что работ на данную тему с использованием нейронных сетей не было. За последнее время данные о результативности игрока не менялись, значит можно добавлять более старые данные, но в данном примере использовались данные игроков НБА, имеющих амплуа – «атакующий защитник», данные были взяты за последние матчи двух игроков с сайта [4]: Лука Дончич и Донован Митчелл.

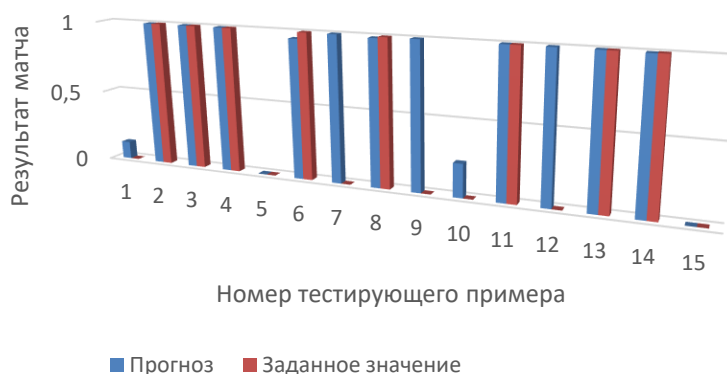
Основная цель настоящей работы заключается в сборе множества публичных данных об игроках и их матчах, а также создание и обучение нейросетевой модели на этих данных. Конечным результатом должна быть нейросетевая система, способная прогнозировать исход матча на основе вклада игрока.

Для создания нейросетевой системы были выбраны следующие параметры: X1 – Количество набранных очков, X2 – количество результативных передач, X3 – количество перехватов мяча, X4 – количество потерь мяча, X5 – количество подборов под кольцом, X6 – количество блокшотов, X7 – количество совершенных фолов. Выходной параметр – результат матча.

Обучающее множество было собрано с помощью программы-парсера с интернет-ресурсов, а также вручную. Перед переходом к проектированию и обучению нейросетевой модели, была выполнена очистка исходного множества от противоречивых примеров, выбросов, дубликатов. Например, везде, где стояли пропуски на сайте были поставлены 0, то есть игрок не совершил этого действия. Таким образом, объем итогового множества включает в себя 114 примеров. Данное множество было разделено на обучающее и тестирующее, 99 и 15 примеров соответственно.

Проектирование, обучение, тестирование нейронной сети выполнялись с помощью программы «Нейросимулятор 5» [5] по методике [6]. В результате оптимизации спроектированная нейронная сеть представляла собой персептрон, который имеет семь входных нейронов, один выходной и один скрытый слой с тремя нейронами.

Для оценки корректной работы спроектированной нейронной сети использовалось тестирующее множество, состоящее из 15 примеров. Средняя относительная ошибка тестирования составила 9%, что можно считать приемлемым результатом.



**Рисунок 1. Результат тестирования нейронной сети**

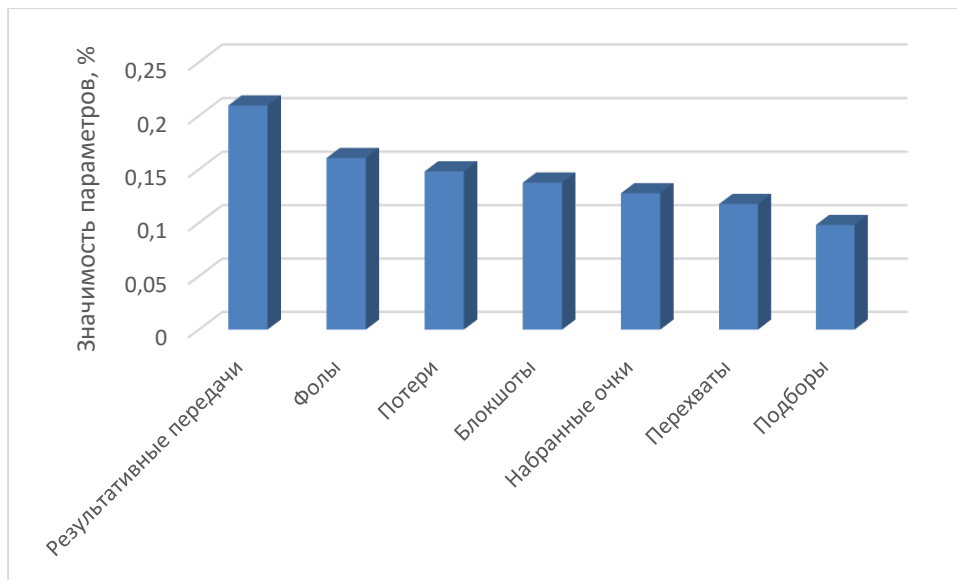
На рисунке 1 представлена гистограмма, демонстрирующая разницу между фактическим и прогнозируемым нейросетью победителями матчей

Из результатов, изображенных на рисунке 1, можно сделать вывод об адекватной работе спроектированной нейронной сети.

Оценка значимости входных параметров была выполнена с помощью программы «Нейросимулятор 5», результат отображен на рисунке 2.

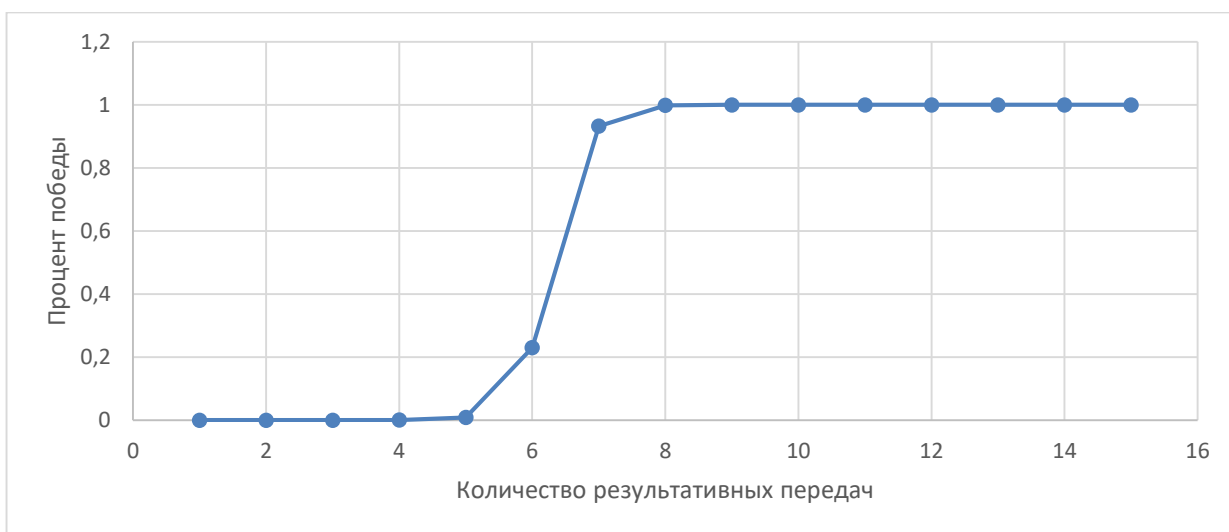
Как видно из рисунка 2, наиболее значимым параметром является количество результативных передач.

Далее было проведено исследование полученных зависимостей между входными параметрами и победителем матча. Исследование производилось с помощью метода «замораживания» [1-3, 6], суть которого заключается в варьировании значения одного параметра и фиксации значений всех других параметров. Для этой цели был отобран «нейтральный пример», про который нейросеть не может с уверенностью сказать, в каком матче исход predetermined.



**Рисунок 2. Значимость входных параметров нейросетевой модели**

На рисунке 3 показан график зависимости прогнозируемого победителя от результата результативных передач. В том случае, когда игрок совершает большее количество результативных передач за игру.

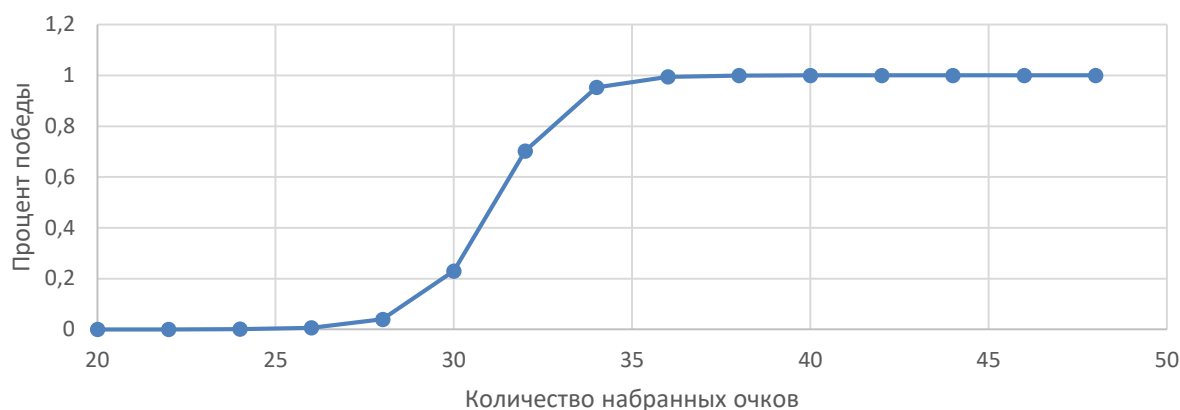


**Рисунок 3. Зависимость прогнозируемого победителя от количества результативных передач**

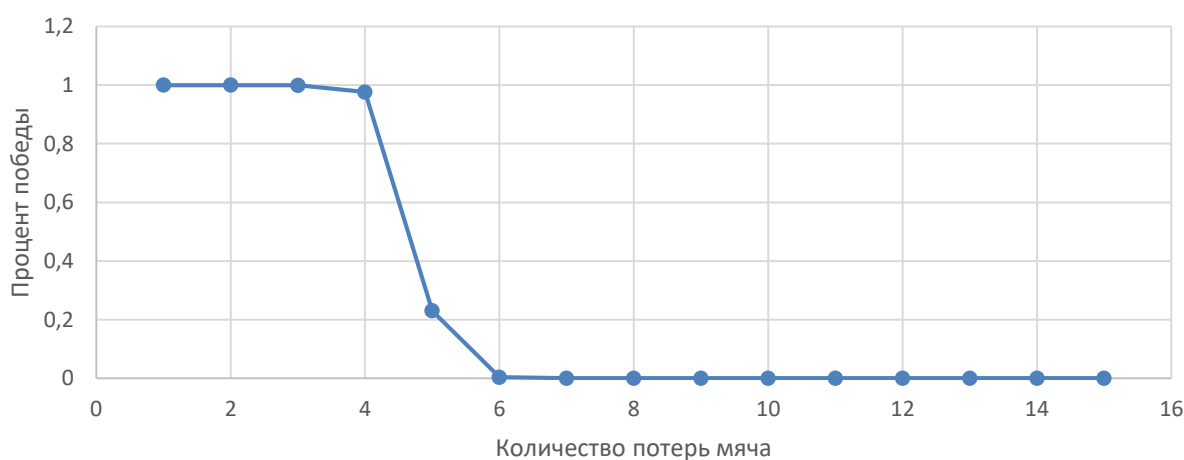
На рисунке 4 продемонстрирована зависимость прогнозируемого победителя от количества набранных очков.

На рисунке 5 изображен график зависимости победителя от количества потерь мяча. Как видно из графика, чем меньше мячей потеряет игрок, тем выше шанс на победу его команды.

Полученные результаты исследований не противоречат реальности, что подтверждает, что спроектированную нейронную сеть можно считать пригодной для прогнозирования результатов матчей профессиональных игроков в снукер.



**Рисунок 4. Зависимость прогнозируемого победителя от количества набранных очков**



**Рисунок 5. Зависимость прогнозируемого победителя от количества потерь мяча**

**Заключение.** Построена система нейросетевого прогнозирования победителя матча между командами на основе результативности игрока. Спроектированная нейросетевая модель учитывает 7 параметров: набранные очки, результативные передачи, перехваты мяча, потери мяча, подборы, блокшоты, фолы. Методом сценарного прогнозирования построены графики зависимостей прогнозируемого победителя матча от изменения входных параметров. Применение такого набора параметров в модели позволяет с высокой точностью прогнозировать победителя матчей. Данный набор параметров может быть изменен для прогнозирования результатов матчей непрофессиональных команд.

### **Библиографический список**

1. Ясницкий Л.Н., Павлов И.В., Черепанов Ф.М. Прогнозирование результатов олимпийских игр 2014 года в неофициальном командном зачете методами искусственного интеллекта // Современные проблемы науки и образования. 2013. № 6. С. 985.
2. Ясницкий Л.Н., Vnukova O.V. Прогноз результатов олимпиады-2014 в мужском одиночном фигурном катании методами искусственного интеллекта // Современные проблемы науки и образования. 2014. № 1. С. 189.



3. Ясницкий Л.Н., Кировоса А.В., Ратегова А.В., Черепанов Ф.М. Прогноз результатов чемпионата мира-2015 по легкой атлетике методами искусственного интеллекта // Современные проблемы науки и образования. 2014. № 4. С. 624.

4. [https://www.championat.com/basketball/\\_nba/tournament/4715/statistic/player/point/](https://www.championat.com/basketball/_nba/tournament/4715/statistic/player/point/) – результаты игроков НБА сезон 2021/2022

5. Черепанов Ф.М., Ясницкий Л.Н. Нейросимулятор 4.0. Свидетельство о регистрации программы для ЭВМ RUS 2014612546. Заявка № 2014610341 от 15.01.2014.

6. Ясницкий Л.Н. Введение в искусственный интеллект. М.: Издательский центр «Академия», 2005. – 176 с.

## NEURAL NETWORK SYSTEM FOR BASKETBALL MATCH WINNER PREDICTION BASED ON PLAYER PERFORMANCE

*Belousov Alexandr A.*

Perm State University

Str. Bukireva, 15, Perm, Russia, 614990, a-a-belousov@mail.ru

The article presents a description of the development of a neural network system for predicting the results of basketball matches based on the player's performance. The system allows you to predict the winner of the next match with great accuracy. With the help of the developed intellectual system, a study of the subject area was carried out, patterns of practical importance were identified.

**Keywords:** artificial intelligence, neural network technologies, forecasting, basketball, NBA.

УДК 004.032.26

## НЕЙРОСЕТЕВАЯ МОДЕЛЬ ПРОГНОЗИРОВАНИЯ ЗАНЯТОГО КИБЕРСПОРТСМЕНОМ МЕСТА В ГОЛОСОВАНИИ ЗА ЛУЧШЕГО ИГРОКА ГОДА

*Коробов Никита Андреевич*

Национальный исследовательский университет «Высшая школа экономики»,  
614070, Россия, г. Пермь, ул. Студенческая, 38, nik3a@mail.ru

В статье представлено описание разработки нейросетевой системы для прогнозирования результатов голосования лучшего игрока года среди киберспортсменов в компьютерной игре «CS GO». В ходе работы система была спроектирована и обучена с помощью существующей статистики результатов киберспортсменов, были проанализированы параметры, влияющие на место в голосовании и определена их значимость. С помощью раз-

работанной интеллектуальной системы возможно формирование полезных рекомендаций для киберспортсмена.

**Ключевые слова:** искусственный интеллект, нейросетевые технологии, прогнозирование, киберспорт, cs go.

**Введение.** В настоящее время киберспорт уже плотно закрепился в мировом сообществе, с каждым годом соревнования по киберспорту собирают все больше призовых и зрителей. Как и в обычном спорте, в киберспорте результаты основных наград для спортсменов плотно зависят от их статистики за сезон. В последнее время, а если быть точным в последние пять лет, статистика игроков в игре «CS GO» ведется точно и подробно. Изучение данной области поможет киберспортсменам увеличить свои шансы на звание лучшего игрока года путем улучшения статистики по наиболее значимым факторам.

При анализе литературных источников выяснилось, что работ на тему прогнозирования результатов лучшего игрока года в компьютерной игре «CS GO» нет. Однако данный тип работы достаточно популярен в большом спорте. Например, прогнозирование результата «Чемпионата Европы» по футболу, состязаний по бегу, соревнований по спортивной ходьбе, результата хоккейных матчей, соревнований по плаванию [1-5]. Из-за отсутствия работ данной тематики в киберспорте, в частности в игре «CS GO», данная тема и была выбрана в качестве исследования.

Основная цель настоящей работы заключается в сборе множества публичных данных об игроках, а также создание и обучение нейросетевой модели на этих данных. Конечным результатом должна быть нейросетевая система, способная прогнозировать лучшего игрока года.

Для создания нейросетевой системы были выбраны следующие параметры:

X1 – рейтинг игрока, показывающий выступил ли игрок выше или ниже среднего числа, где 1.00 является средним значением (RATING 2.0).

X2 – усредненное количество смертей. Рассчитывается по формуле: количество смертей / количество раундов (DPR).

X3 – процент раундов, в которых игрок совершил убийство, помощь, выжил или был разменян (KAST %).

X4 – Измеряет влияние мульти убийств, открывающих убийств и клатчей (ИМПАСТ).

X5 – среднее количество урона, который наносит игрок своему сопернику за раунд. Рассчитывается по формуле: количество общего урона / количество сыгранных раундов (ADR).

X6 – статистический показатель, который показывает количество убийств, который игрок совершает в среднем за игру. Рассчитывается по формуле: количество убийств / количество сыгранных раундов (KPR).

X7 – количество сыгранных за сезон карт (Maps played).

X8 – процент убийств в голову (Headshot %).

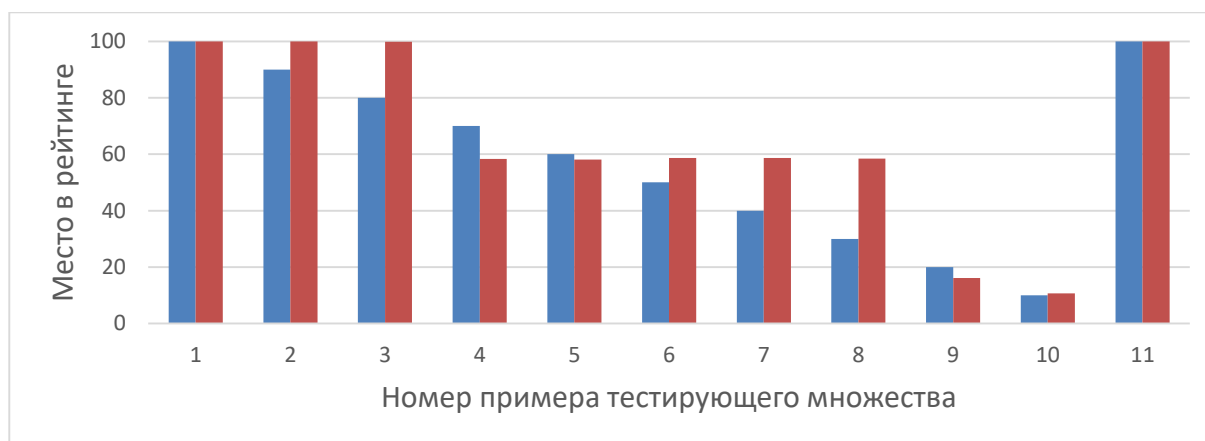
X9 – количество наград лучшего игрока главных чемпионатов за сезон (Major MVP). Возможный максимум за сезон равен 2 (100%).

X10 – количество выигранных главных чемпионатов за сезон (Major winner). Возможный максимум за сезон равен 2 (100%).

Обучающее множество было собрано с интернет-ресурса [6]. Так как статистические данные в данной области быстро устаревают, были взяты актуальные данные за последние 5 лет. Множество примеров, состоящее из 51 записи, было поделено на обучающее, предназначенное для обучения нейронной сети, и тестовое, предназначенное для проверки нейронной сети. В тестовое множество вошло 11 записей.

Проектирование, обучение, тестирование нейронной сети выполнялись с помощью программы «Нейросимулятор 5» [7] по методике [8]. В результате оптимизации спроектированная нейронная сеть представляет собой персептрон, который имеет десять входных нейронов, один выходной и один скрытый слой с двумя нейронами.

Для оценки корректной работы спроектированной нейронной сети использовалось тестирующее множество. Средняя относительная ошибка тестирования составила 10.4%, что можно считать приемлемым результатом. На рисунке 1 представлена гистограмма, демонстрирующая разницу между фактическим и прогнозируемым. Где 100 – 1 место, 90 – 2 место, 80 – 3 место и т.д. до 10 места.

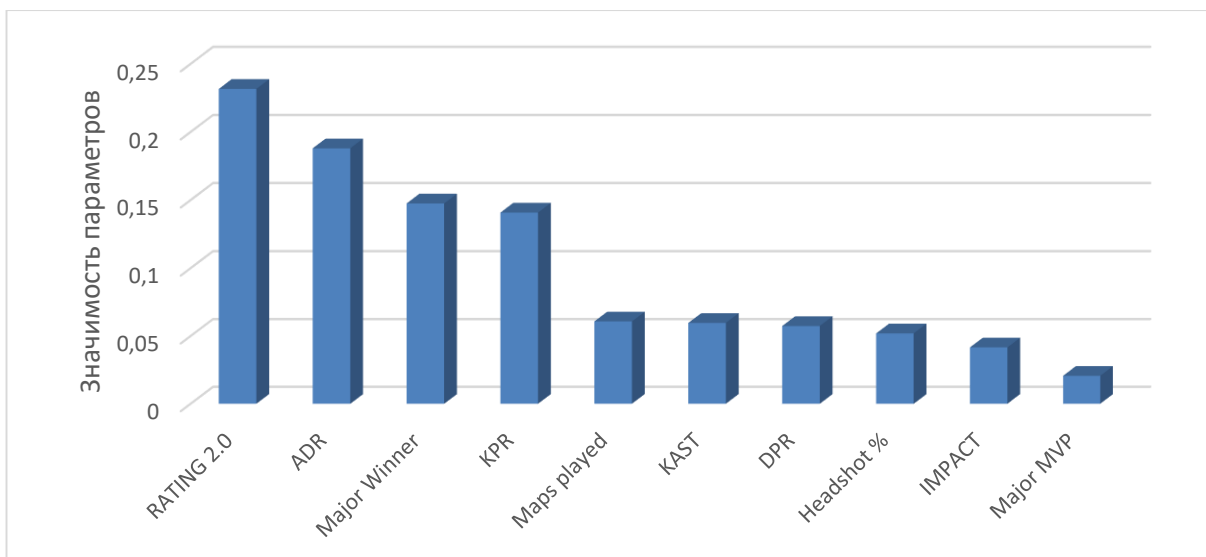


**Рисунок 1. Результат тестирования нейронной сети**

Из результатов, изображенных на рисунке 1, можно сделать вывод об вполне адекватной работе спроектированной нейронной сети.

Оценка значимости входных параметров была выполнена с помощью программы «Нейросимулятор 5», результат отображен на рисунке 2.

Как видно из рисунка 2, наиболее значимыми параметрами являются: рейтинг выступлений игрока, среднее количество урона и т.д. Как и ожидалось, наиболее влиятельным параметром является средний рейтинг эффективности игрока. Это можно объяснить тем, что рейтинг эффективности игрока в целом показывает на сколько игрок выступал лучше, чем остальные.

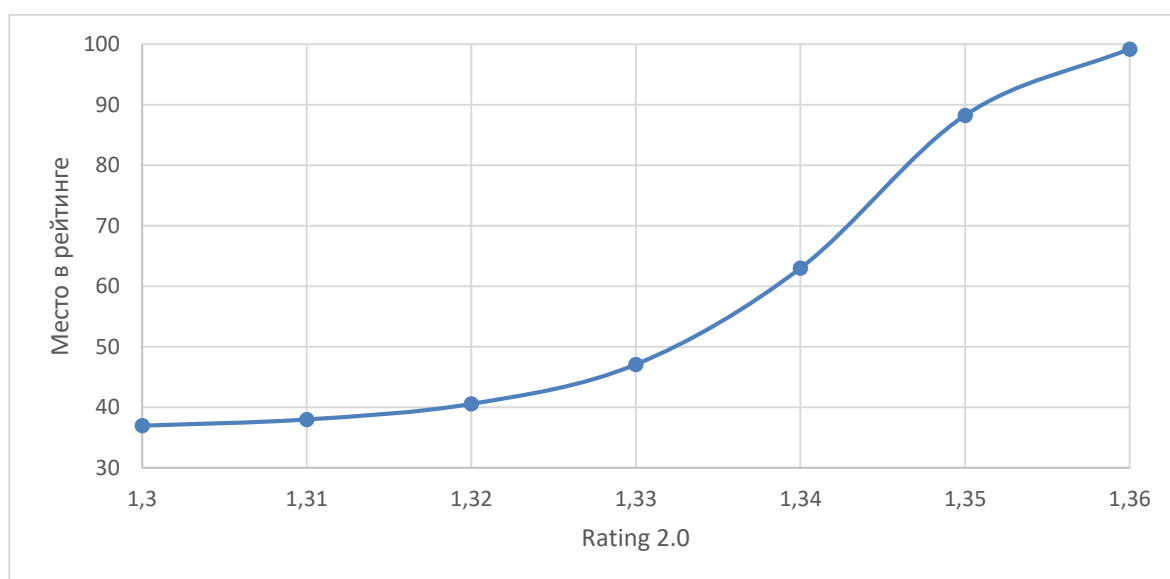


**Рисунок 2. Значимость входных параметров нейросетевой модели**

Далее было проведено исследование полученных зависимостей между входными параметрами и победителем голосования лучшего игрока. Исследование производилось с помощью метода «замораживания» [1-5, 8], суть которого заключается в варьировании значения одного параметра и фиксации значений всех других параметров. Для этой цели был взят прототип реального игрока с хорошей статистикой за 2021 год, но с некоторыми изменениями его параметров.

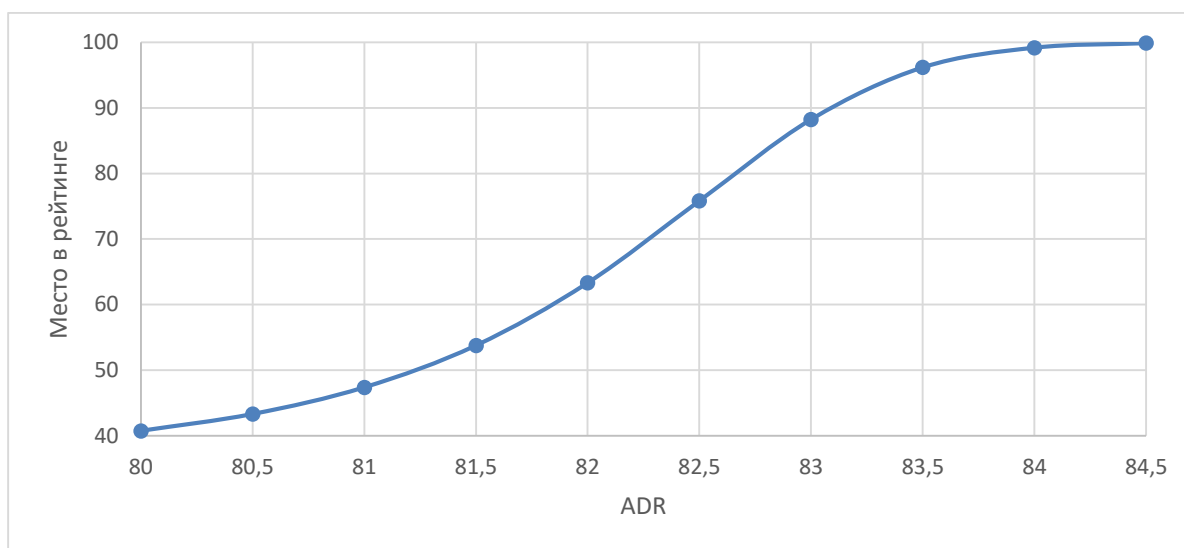
Параметры игрока:  $X_1 - 1.35$ ;  $X_2 - 0.6$ ;  $X_3 - 76$ ;  $X_4 - 1.43$ ;  $X_5 - 83$ ;  $X_6 - 0.88$ ;  $X_7 - 198$ ;  $X_8 - 38.2$ ;  $X_9 - 50$ ;  $X_{10} - 50$ .

На рисунке 3 продемонстрирована зависимость прогнозируемого победителя голосования от рейтинга выступлений игрока. Можно заметить, что чем выше рейтинг выступлений игрока, тем больше у игрока шансов на победу в голосовании.



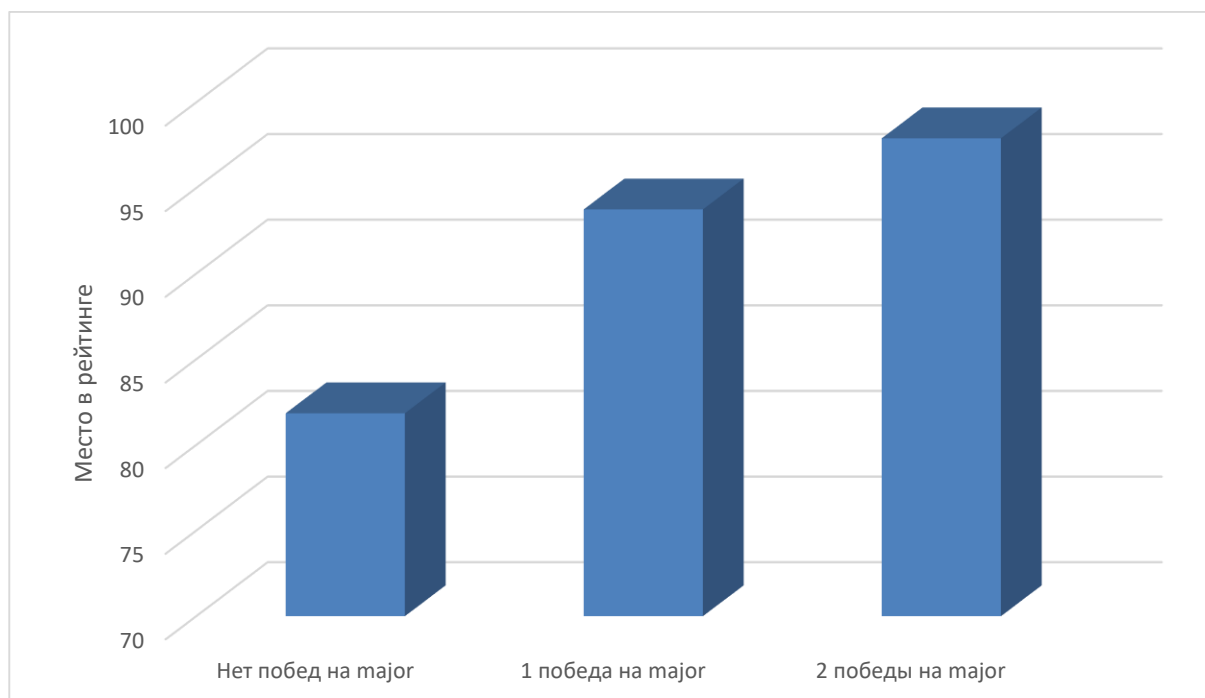
**Рисунок 3. Зависимость прогнозируемого победителя от рейтинга выступлений игрока**

На рисунке 4 показан график зависимости прогнозируемого победителя голосования от статистики среднего количества нанесенного урона. Как видно из графика, чем больше у игрока показатель ADR, тем больше у него шансов занять первое место в голосовании.



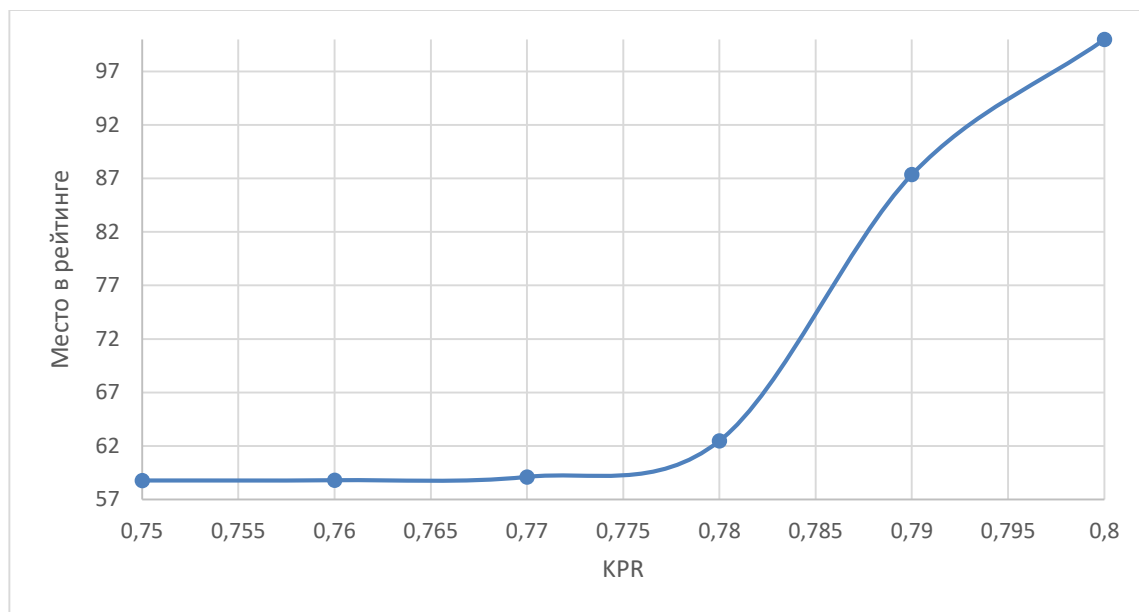
**Рисунок 4. Зависимость прогнозируемого победителя от статистики среднего урона**

На рисунке 5 продемонстрирована зависимость прогнозируемого победителя голосования от количества выигранных игроком главных чемпионатов. Как можно увидеть, чем больше игрок выиграл главных чемпионатов за сезон, тем выше у него шансы стать лучшим.



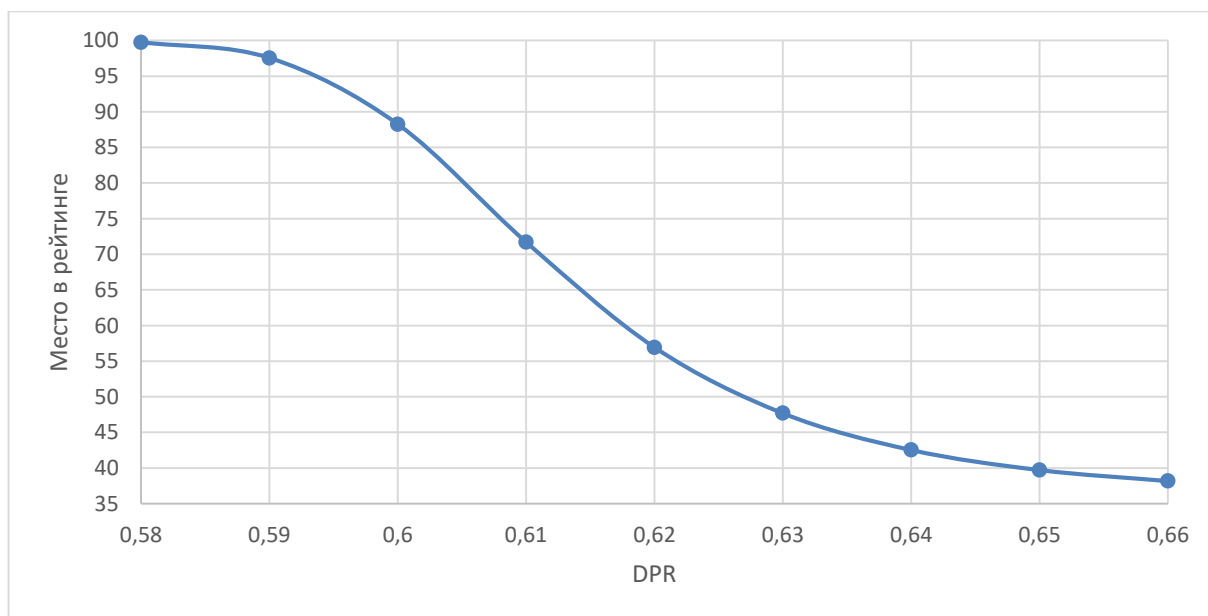
**Рисунок 5. Зависимость прогнозируемого победителя от побед на главных чемпионатах**

На рисунке 6 изображен график зависимости победителя от среднего количества убийств игроком. Как видно из графика, чем чаще игрок убивает противников, тем больше у него шансов на победу в голосовании.



**Рисунок 6. Зависимость прогнозируемого победителя от статистики убийств**

На рисунке 7 изображен график зависимости победителя от среднего количества смертей игрока. Как видно из графика, чем чаще игрок умирает, тем меньше у него шансов на победу в голосовании.



**Рисунок 7. Зависимость прогнозируемого победителя от среднего количества смертей игрока**

Полученные результаты исследований не противоречат реальности, что подтверждает, что спроектированную нейронную сеть вполне можно считать пригодной для прогнозирования результатов голосования лучшего игрока сезона.

**Заключение.** Построена система нейросетевого прогнозирования победителя голосования лучшего игрока сезона по компьютерной игре «CS GO». Спроектированная нейросетевая модель учитывает 10 параметров: рейтинг игрока; усредненное количество смертей; процент раундов, в которых игрок совершил убийство; влияние мульти убийств, открывающих убийств и клатчей; среднее количество урона; количество убийств; количество сыгранных за сезон карт; процент убийств в голову; количество наград лучшего игрока главных чемпионатов за сезон; количество выигранных главных чемпионатов за сезон; Методом сценарного прогнозирования построены графики зависимостей прогнозируемого победителя голосования от изменения входных параметров. Применение такого набора параметров в модели позволяет с высокой точностью прогнозировать победителя голосования. Данный набор параметров может быть изменен для прогнозирования результатов голосования, игроков полупрофессиональной сцены. В дополнение была реализована нейросетевая модель с помощью библиотеки Keras Python. По итогам ее реализации можно сказать, что полученная с помощью Keras нейросетевая модель уступает в качестве модели, полученной с помощью нейросимулятора 5.

### **Библиографический список**

1. Ясницкий Л.Н., Павлов И.В., Черепанов Ф.М. Прогнозирование результатов олимпийских игр 2014 года в неофициальном командном зачете методами искусственного интеллекта // Современные проблемы науки и образования. 2013. № 6. С. 985.
2. Ясницкий Л.Н., Vnukova O.V. Прогноз результатов олимпиады-2014 в мужском одиночном фигурном катании методами искусственного интеллекта // Современные проблемы науки и образования. 2014. № 1. С. 189.
3. Ясницкий Л.Н., Кировоса А.В., Ратегова А.В., Черепанов Ф.М. Прогноз результатов чемпионата мира-2015 по легкой атлетике методами искусственного интеллекта // Современные проблемы науки и образования. 2014. № 4. С. 624.
4. Ясницкий Л.Н., Абрамова Ю.С., Бабушкина С.Д. Возможности получения рекомендаций по улучшению результативности сборных команд, готовящихся к участию в чемпионате Европы по футболу Евро2016 методом нейросетевого моделирования // Вестник спортивной науки. 2015. № 5. С. 15-20.
5. Крутиков А.К. Прогнозирование результата хоккейного матча с использованием специализированного программного модуля на основе искусственной нейронной сети // Научное обозрение. Технические науки. No 2. 2019. С. 19-22.
6. HLTV.org – статистика игроков в CS GO. [Электронный ресурс]. Режим доступа: <https://www.hltv.org/>
7. Черепанов Ф.М., Ясницкий Л.Н. Нейросимулятор 4.0. Свидетельство о регистрации программы для ЭВМ RUS 2014612546. Заявка № 2014610341 от 15.01.2014.
8. Ясницкий Л.Н. Введение в искусственный интеллект. М.: Издательский центр «Академия», 2005. – 176 с.

# NEURAL NETWORK MODEL FOR PREDICTING A CYBER ATHLETE'S PLACE IN THE VOTING FOR BEST PLAYER OF THE YEAR

*Korobov Nikita A.*

National Research University "Higher School of Economics",  
Str. Studencheskaya 38, Perm, Russia, 614070, nik3a@mail.ru

This article describes the development of a neural network system for predicting the results of voting for the best player of the year among eSportsmen in the computer game "CS GO". In the course of the work the system was designed and trained using existing statistics on the results of eSportsmen, were analyzed the parameters that affect the voting place and determined their importance. With the help of the developed intelligent system, it is possible to generate useful recommendations for a cyber athlete.

**Keywords:** artificial intelligence, neural network technology, prediction, cybersport, cs go.

УДК 004.032.26

## НЕЙРОСЕТЕВАЯ СИСТЕМА ПРОГНОЗИРОВАНИЯ ПОБЕДИТЕЛЯ НА СКАЧКАХ

*Стрекаловская Елена Владимировна*

Национальный исследовательский университет «Высшая школа экономики»,  
614990, Россия, г. Пермь, ул. Бульвар Гагарина, 37,  
e.strekalovskaya@edu.hse.ru

В статье представлено описание разработки нейросетевой системы для прогнозирования результатов скачек. Система позволяет с большой точностью предсказать лошадь, которая может занять 1 место на скачках. С помощью разработанной интеллектуальной системы проведено исследование предметной области.

**Ключевые слова:** искусственный интеллект, нейросетевые технологии, прогнозирование, скачки, первое место.

**Введение.** Скачки — один из самых популярных видов спорта в целом ряде стран: в США, Австралии, Британии, ОАЭ. Соревнования проходят на разные дистанции, результаты зависят от веса лошади и жокея, также имеет значение покрытие ипподрома (трава, песок и т. д.) и погодные условия. И только это уже говорит о том, что, прежде чем делать прогнозы на победу одного из участников скачек, необходимо понять суть и нюансы данного вида спорта. Существует много стратегий вида ставок на лошадиные скачки, но



рассматриваться в данном исследовании будет самая простая – займет ли лошадь первое место [11]. В последнее время для выполнения прогнозов во многих предметных областях получает распространение один из наиболее эффективных разделов искусственного интеллекта – нейросетевые технологии. Правильно натренированная нейросетевая система может добиться успешности большей части прогнозов в спортивных соревнованиях. Как показал успешный опыт Пермской научной школы искусственного интеллекта, например, [1,2], нейронные сети могут помочь спрогнозировать результаты соревнований, [5], что в дальнейшем может дать возможность букмекерским сайтам и их пользователям извлечь больше прибыли.

На сегодняшний день существует очень мало исследований на тему прогнозирования результатов лошадиных скачек, в основном, исследователи пытаются разработать систему стратегий для ставок, например, система Датчи Шульца [9].

Основная цель настоящей работы заключается в сборе множества публичных данных о лошадях и жокеях, а также создание и обучение нейросетевой модели на основе имеющихся данных. В результате исследования должна работать нейросетевая система, способная прогнозировать победу лошади на скачке.

Для создания нейросетевой системы были выбраны следующие параметры: для сначала закодируем все данные в количественные:

X1 – разряд жокея. 5 разряд считается разрядом новичков – закодировано, как 1; 6 разряд – специалист, закодировано, как 2.

X2 – пол лошади. 1 – мужской пол, 2 – женский пол.

X3 – вес лошади

X4 – возраст лошади

X5 – количество призовых мест

X6 – дистанция

X7 – рейтинг конюшни

Выходной параметр – вероятность победить на скачке.

Обучающее множество было собрано с помощью программы-парсера с интернет-ресурсов, а также вручную. Перед переходом к проектированию и обучению нейросетевой модели, была выполнена очистка исходного множества от противоречивых примеров, выбросов, дубликатов. Изначально в выборке были данные с разных типов забегов, например, для новичков, мейденов, гандикапов и так далее [10]. Некорректными примерами считались те, где лошадь старшего возраста лучше бежит длинные дистанции, но были оставлены примеры с короткими дистанциями, ведь молодая лошадь лучше бежит до 2000 м. Таким образом, объем итогового множества с лошадьми-новичками на короткие дистанции включает в себя 70 примеров.

Данное множество было разделено на обучающее и тестирующее в соотношении 80% и 20% соответственно. Большая часть данных была собрана с интернет ресурсов [7-8].

Проектирование, обучение, тестирование нейронной сети выполнялись с помощью программы «Нейросимулятор 5» [3] по методике [8]. В результате оптимизации спроектированная нейронная сеть представляла собой пер-

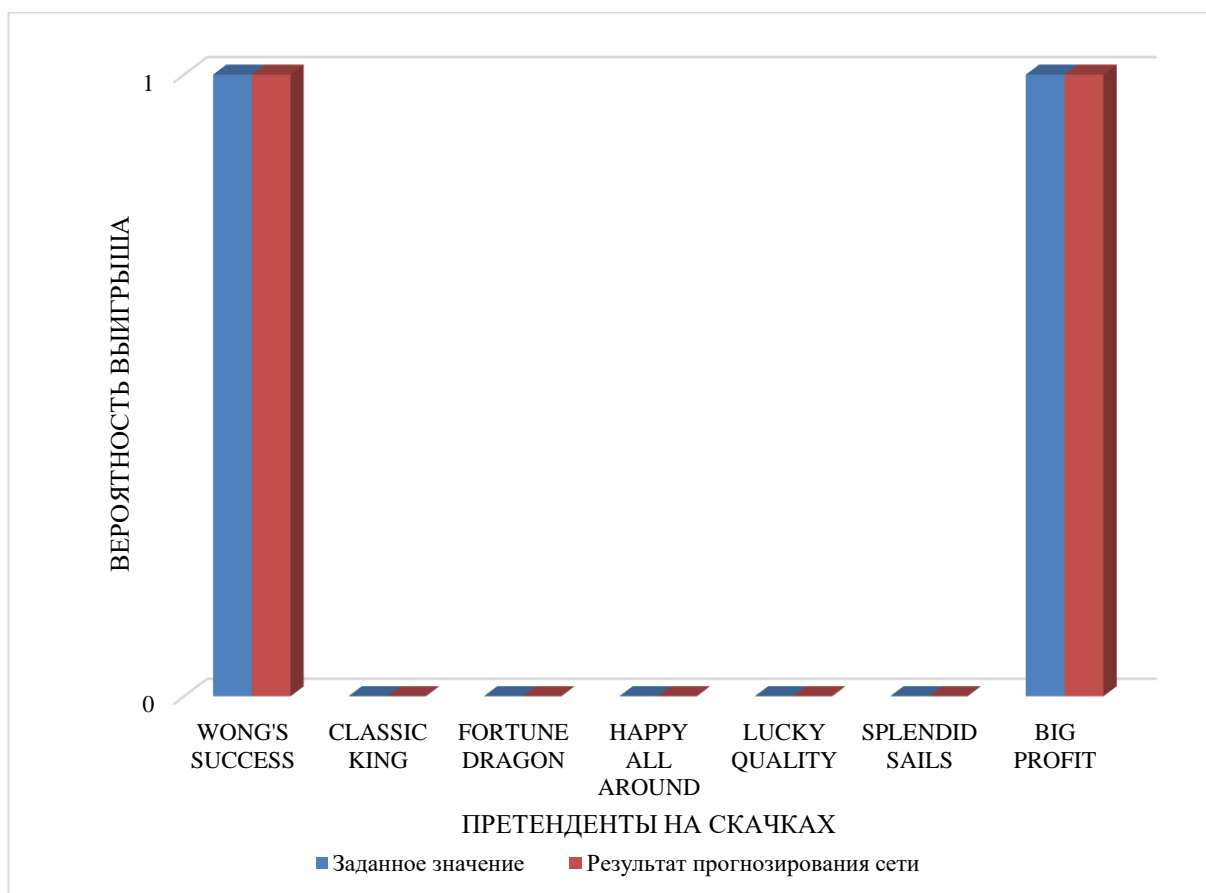
септрон, который имеет 7 входных нейронов, один выходной нейрон и один скрытый слой.

Для оценки корректной работы спроектированной нейронной сети использовалось тестирующее множество, состоящее из 30 примеров. Средняя относительная ошибка тестирования составила 0.1, что можно считать прекрасным результатом.

На рисунке 1 представлена гистограмма, демонстрирующая нулевую разницу между фактическим результатом победителями скачек и прогнозируемыми нейросетью.

Из результатов, изображенных на рисунке 1, можно сделать вывод об исправной работе спроектированной нейронной сети, прогноз полностью совпал с ожидаемыми результатами.

Оценка значимости входных параметров была выполнена с помощью программы «Нейросимулятор 5», результат отображен на рисунке 2.

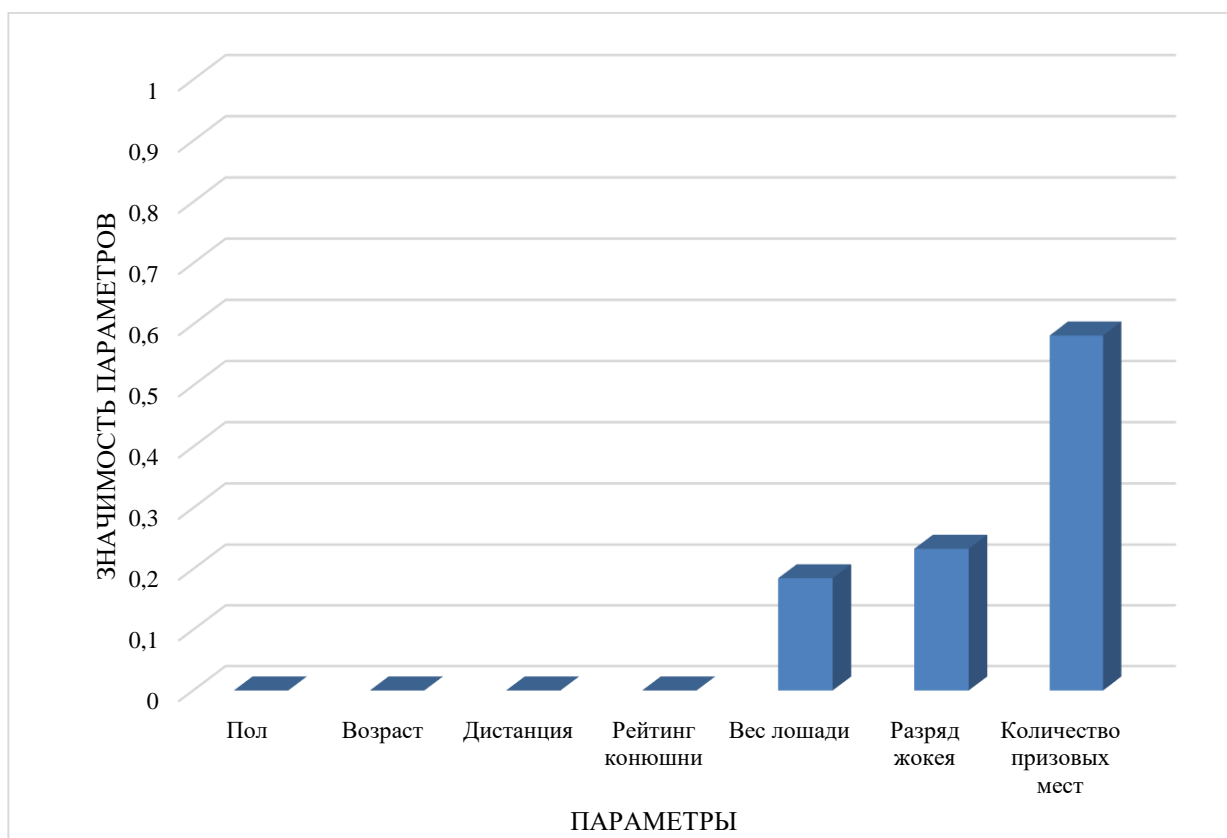


**Рисунок 1. Результат тестирования нейронной сети**

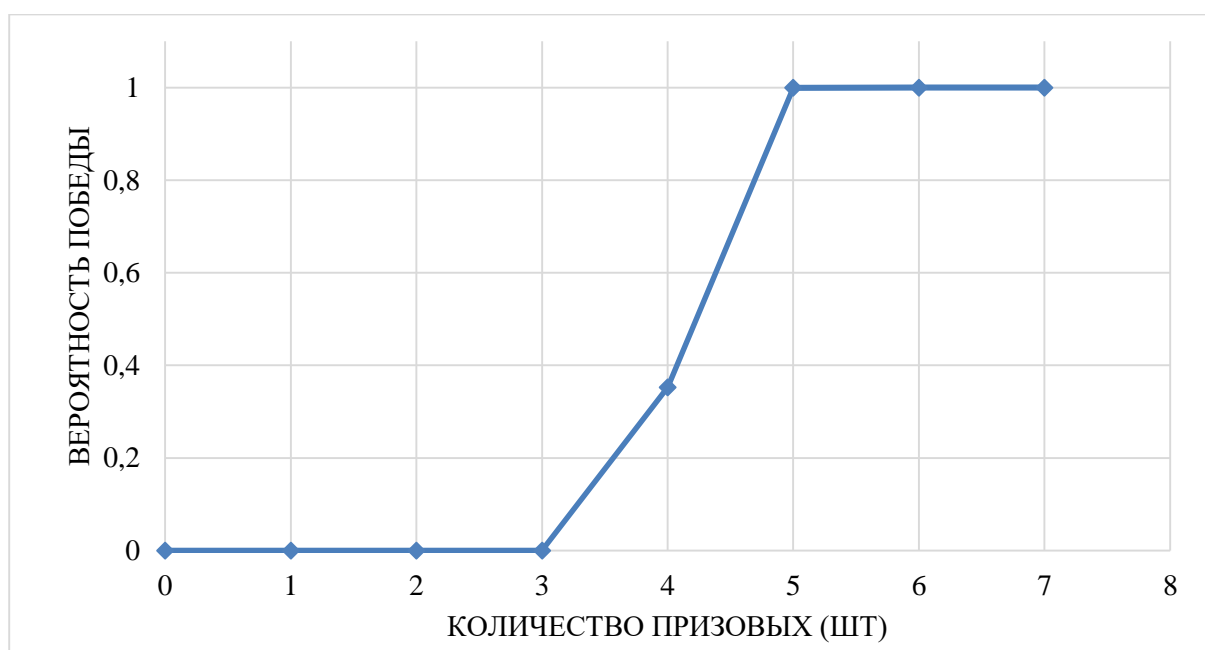
Как видно из рисунка 2, наиболее значимыми является параметр с количеством призовых мест, далее – вес лошади и разряд жокея. Как и следовало ожидать, наиболее влиятельным параметром является количество призовых мест за всю карьеру скакуна.

Далее было проведено исследование полученных зависимостей между входными параметрами и лошадьми-победителями. Исследование производилось с помощью метода «замораживания» [11], суть которого заключается

в варьировании значения одного параметра и фиксировании значений всех других параметров. Для этой цели был отобран «нейтральный пример», про который нейросеть не может с уверенностью сказать, у кого из скакунов значительное преимущество.



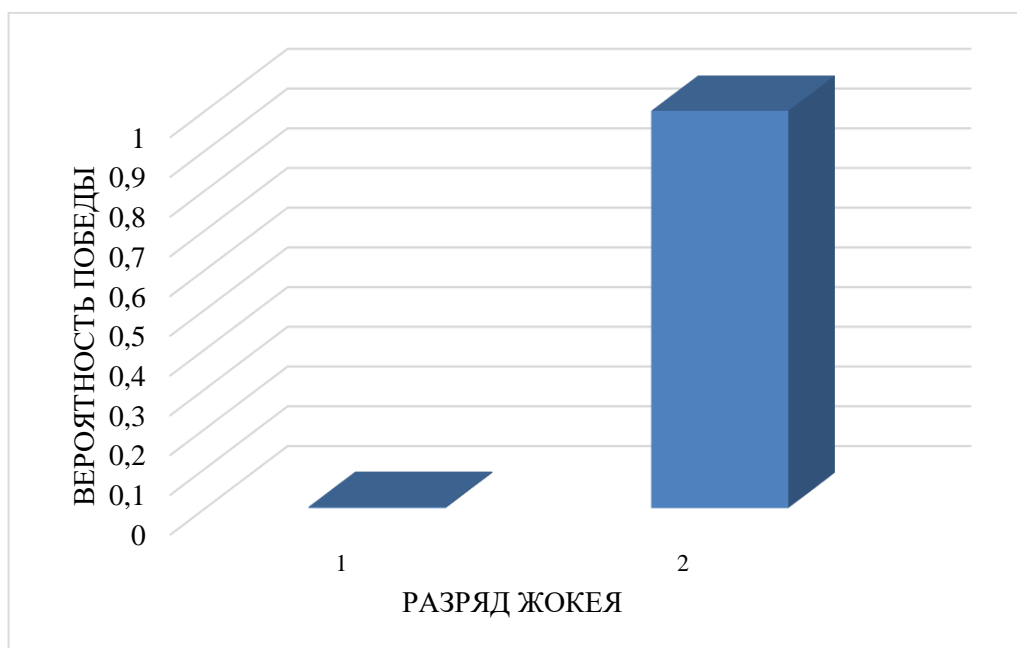
**Рисунок 2. Значимость входных параметров нейросетевой модели**



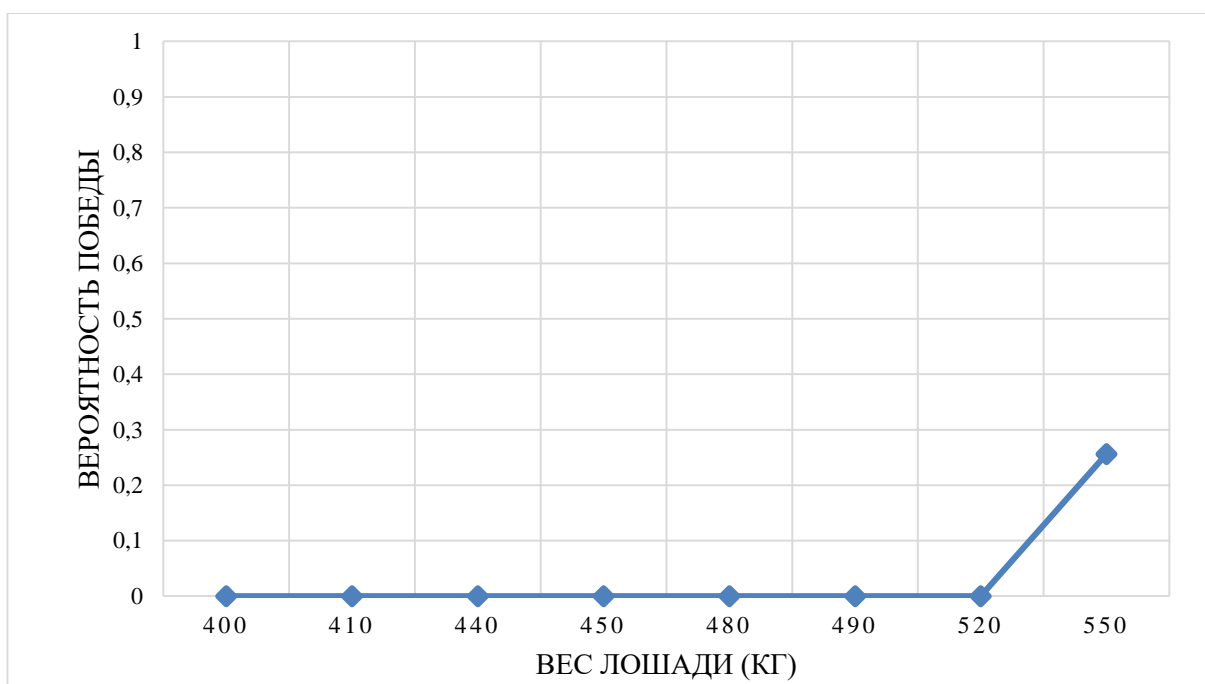
**Рисунок 3. Зависимость прогнозируемого победителя скачки от количества призовых мест**

На рисунке 3 показан график зависимости прогнозируемой лошади-победителя от результата изменения количества первых мест в предыдущих скачках. В том случае, когда у скакуна большое количество призовых мест, нейросеть прогнозирует значительное увеличение шанса на победу.

На рисунке 4 продемонстрирована зависимость прогнозируемого победителя скачки от разряда жокея. Можно заметить, что при разряде жокея 1, другими словами, жокей-новичок, вероятность победы снижается.



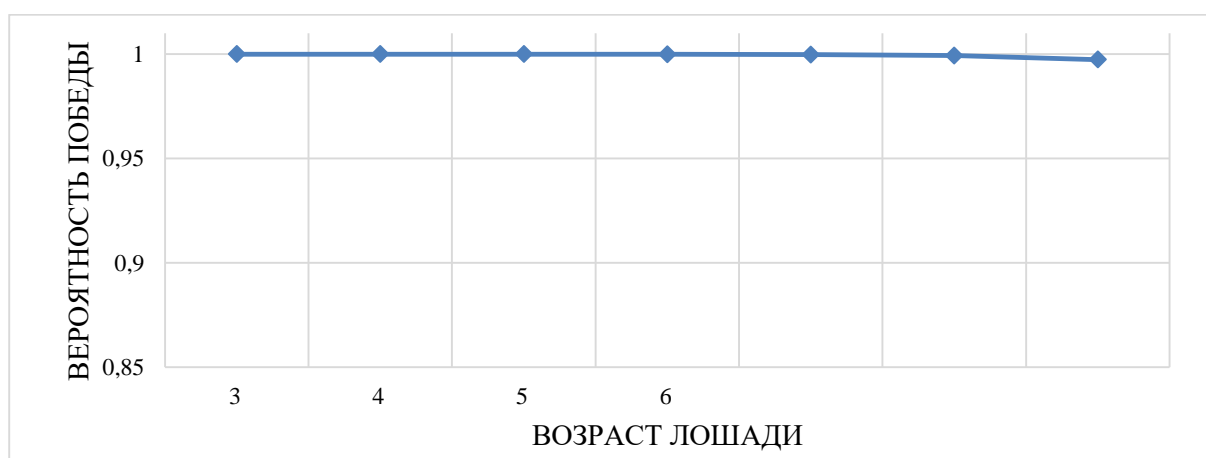
**Рисунок 4. Зависимость прогнозируемого победителя скачки от разряда жокея**



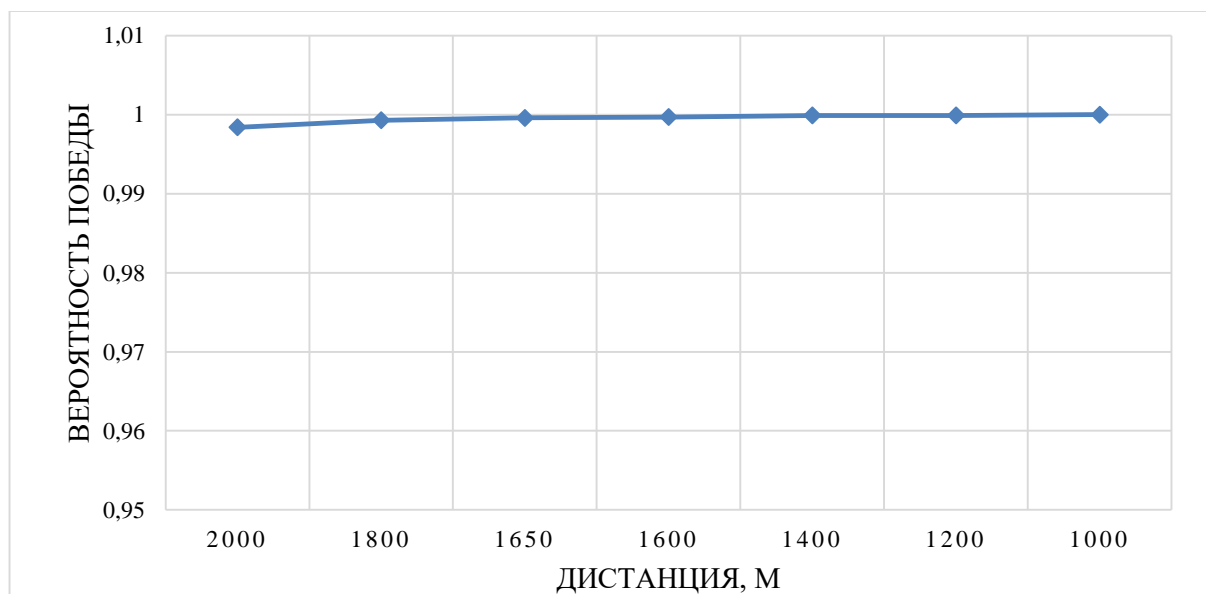
**Рисунок 5. Зависимость прогнозируемой победы лошади от её веса**

На рисунке 5 изображен график зависимости победы лошади от её веса. Вес лошадей варьируется от 410 до 550 кг, и даже больше, но стоит учитывать, что вес жокея при этом тоже играет роль, к сожалению, в собранных данных нет информации о весе жокея, но по имеющимся результатам веса лошадей видно, что вероятность победы лошади с весом более 500 кг более высока.

На рисунке 6 продемонстрирована зависимость прогнозируемой победы лошади от её возраста. Как видно из графика, чем меньше возраст лошади – тем больше шанс на победу. Разница по возрасту на короткие забеги не особо важна, так как предыдущие три фактора оказывают большее влияние, но на короткие дистанции молодые скакуны, как уже было сказано ранее, бегут гораздо лучше, если сравнивать с лошадьми-старичками. Вероятность варьируется от 0,9 до 1.



**Рисунок 6. Зависимость прогнозируемой победы лошади от её возраста**

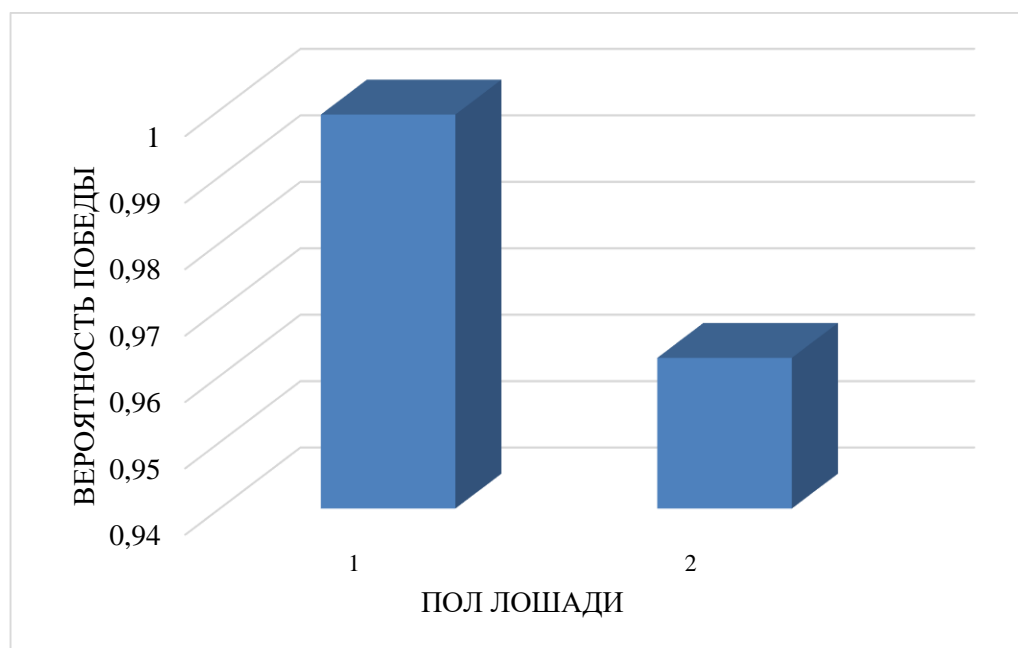


**Рисунок 7. Зависимость прогнозируемой победы лошади от дистанции**

На рисунке 7 продемонстрирована зависимость прогнозируемого прогнозируемой победы лошади от дистанции забега. В дальнейшем можно собрать данные об ипподроме, где проходит забег и о погодных условиях. Как

можно заметить, вероятность победы варьируется от 0,9984 до 1, а это значит, что дистанция не так сильно влияет на результат.

На рисунке 8 продемонстрирована зависимость прогнозируемой победы лошади от её пола. Мужской пол закодирован, как 1, женский – как 2. Нетрудно заметить, что пол лошади также не влияет на результат выигрыша, вероятность варьируется от 0,94 до 1. Значит параметр по половому признаку не имеет значения для победителя.



**Рисунок 8. Зависимость прогнозируемой победы лошади от пола**

Полученные результаты исследований не противоречат реальности, что подтверждает, что спроектированную нейронную сеть можно считать пригодной для прогнозирования результатов скачек в разных странах.

**Заключение.** Построена система нейросетевого прогнозирования победителей скачек. Спроектированная нейросетевая модель учитывает 7 параметров: разряд жокея (новичок или специалист), пол лошади, её вес, возраст, количество призовых мест, дистанцию забега и рейтинг конюшни. Методом сценарного прогнозирования построены графики зависимостей прогнозируемой победы лошади от изменения входных параметров. Применение такого набора параметров в модели позволяет с прогнозировать победителя спортивных скачек. Данный набор параметров может быть доработан, например, если добавить информацию о выносливости лошади, дате последней скачки и о погодных условиях.

### **Библиографический список**

1. Ясницкий Л.Н., Павлов И.В., Черепанов Ф.М. Прогнозирование результатов олимпийских игр 2014 года в неофициальном командном зачете методами искусственного интеллекта // Современные проблемы науки и образования. 2013. № 6. С. 985.

2. Ясницкий Л.Н., Кировоса А.В., Ратегова А.В., Черепанов Ф.М. Прогноз результатов чемпионата мира-2015 по легкой атлетике методами искусственного интеллекта // Современные проблемы науки и образования. 2014. № 4. С. 624.
3. Ясницкий Л.Н. Черепанов Ф.М. Нейросимулятор 4.0. Свидетельство о регистрации программы для ЭВМ RUS 2014612546. Заявка № 2014610341 от 15.01.2014.
4. Ясницкий Л.Н. Черепанов Ф.М. Искусственный интеллект. Методическое пособие. – М.: БИНОМ. Лаборатория знаний. 2012. – 216 с.
5. Ясницкий Л.Н., Внукова О.В., Черепанов Ф.М. Прогноз результатов олимпиады-2014 в мужском одиночном фигурном катании методами искусственного интеллекта // Современные проблемы науки и образования. 2014. URL: <https://science-education.ru/ru/article/view?id=11339> (дата обращения: 08.03.2022).
6. Yasnitsky L.N. Algorithm for searching and analyzing abnormal observations of statistical information based on the Arnold – Kolmogorov – Hecht-Nielsen theorem // International Journal of Advanced Trends in Computer Science and Engineering. 2020. 9(2). Pp. 1814-1819. DOI: 10.30534/ijatcse/2020/139922020 <http://www.warse.org/IJATCSE/static/pdf/file/ijatcse139922020.pdf>
7. Kaggle.com – статистика скачек в Гонконге. [Электронный ресурс]. Режим доступа: <https://www.kaggle.com/lantanacamara/hong-kong-horse-racing> (дата обращения: 20.02.2022).
8. Kaggle.com – Данные о скачках с 1990. [Электронный ресурс]. Режим доступа: <https://www.kaggle.com/hwaitt/horse-racing> (дата обращения: 20.02.2022).
9. Legalbet.ru – Ставки на скачки: главные факторы и популярные стратегии – прогноз результатов спортивных матчей. [Электронный ресурс]. Режим доступа: <https://legalbet.ru/shkola-bettinga/stavki-na-skachki-glavnie-faktori-i-populyarnie-strategii/> (дата обращения: 08.03.2022).
10. СадОгород – Стратегия ставок на скачки. [Электронный ресурс]. Режим доступа: <https://vikings-warofclans.ru/kroliki/loshad-pretendent-1-na-pobedu-v-skachkah.html> (дата обращения: 08.03.2022).
11. РБ-Знания – Лошадиные скачки: что нужно знать, прежде чем начинать делать ставки?. [Электронный ресурс]. Режим доступа: <https://bookmaker-ratings.ru/wiki/loshadinyie-skachki-cto-nuzhno-znat-prezhde-chem-nachinat-delat-stavki/> (дата обращения: 08.03.2022).
12. Ясницкий Л.Н. Введение в искусственный интеллект. М.: Издательский центр «Академия», 2005. – 176 с.

# NEURAL NETWORK SYSTEM FOR PREDICTING THE WINNER OF HORSERACING

*Strekalovskaya Elena V.*

National Research University “Higher School of Economics”  
Studencheskaya st., 38, Perm, Russia, 614990, e.strekalovskaya@edu.hse.ru

The article describes the development of a neural network system for predicting the results of horseracing. The system allows you to predict with great accuracy the horse that can take 1st place in the race.

**Keywords:** artificial intelligence, neural network technologies, forecasting, horseracing, first place.

УДК 004.032.26

## ПРОГНОЗИРОВАНИЕ ИСХОДА БОЕВ СМЕШАНЫХ ЕДИНОБОРСТВ МЕТОДОМ НЕЙРОСЕТЕВОГО МОДЕЛИРОВАНИЯ

*Верхоланцев Филипп Андреевич*

Пермский государственный национальный исследовательский университет,  
БАС. 614990, Россия, г. Пермь, ул. Букирева, 15  
filipp.verh.olantsev@yandex.ru

В статье представлено сравнение трех нейросетевых моделей, прогнозирующих исходы боев смешанных единоборств. Выявляется лучшая модель, и с помощью нее оценивается результат главного боя турнира UFC 269 11 декабря 2021. Модель учитывает исторические данные боев и физические данные бойца. Модель основана на данных организации UFC, но может быть использована для прогнозирования боев других организаций.

**Ключевые слова:** нейросетевая модель, нейросеть, метод замораживания, UFC, прогнозирование результатов, обучение сети.

**Введение.** С каждым днем спорт становится все более популярным. Тренд на здоровый образ жизни охватывает большое количество людей. Спорт развивается, появляются новые виды, стремительно набирают новых участников и фанатов классические спортивные состязания. Одним из самых популярных видов спорта являются бои без правил.

К спортивным организациям, базирующимся на данном виде спорта относятся UFC, Bellator, One и др. Самой популярной является UFC [5]. Ближайший турнир этой организации пройдет 11 декабря. Главный бой будут представлять Дастин Порье, претендент на пояс в легком весе и Чарльз Оливейра, который будет защищать пояс. В данной работе планируется создать нейросетевую модель, которая сможет предсказать исход этого боя и выявить, сможет ли Оливейра защитить свой пояс и остаться чемпионом [1].



На текущий момент существуют большое количество работ в спортивной науке, которые используют нейросетевые технологии, к примеру, прогнозирование результатов спортивных состязаний олимпийских игр 2014 [13,15], предсказание результатов чемпионата мира по легкой атлетике 2015 [14], получение рекомендаций футбольным командам с помощью нейросетевой модели [12] и др. Данное же исследование основывается на сравнении трех нейросетевых моделей, прогнозирующих исход боев смешанных боевых искусств, и выявлении лучшей из них.

Основной целью данной работы будет создание нейросетевой модели, которая сможет предсказывать исход боя, путем сбора данных и обучения сети на этих данных.

Для реализации поставленной цели, в результате анализа готовых датасетов и разных источников [1-6], было разработано и проанализировано три нейросетевых моделей.

## 1. Описание моделей

### Модель 1

Первая нейросетевая модель состояла из 23 входных параметров (таблица 1). Выходной параметр сети D1 – вероятность победы 1 бойца.

Таблица 1

Входные параметры модели №1

X1	X2, X13	X3, X14	X4, X15	X5, X16	X6, X17	X7, X18	X8, X19	X9, X20	X10, X21	X11, X22	X12, X23
Пол бойцов	Стойка 1 и 2 бойцов соответственно	Ударные навыки бойцов	Навыки борьбы бойцов	Количество поражений	Количество побед решением	Количество нокаутов	Количество побед сдачей	Рост	Размах рук	Вес	Возраст

Множество для обучения было собранно с интернет-ресурса Kaggle [6], содержащего большое количество дата-сетов. Перед тем как предоставить его нейросети, оно было очищено от выбросов с помощью алгоритма поиска статических выбросов [7]. Получившееся множество было разбито случайным образом на обучающее, состоящее из 1875 примеров и тестовое, включающее 210 примеров. Обучение нейросети было выполнено с помощью программы «Нейросимулятор 5» [9].

### Модель 2

Вторая нейросетевая модель состояла из двенадцати входных параметров. По сравнению с предыдущей моделью была взята разность соответствующих параметров (кроме параметра пол бойцов), убран параметр стойка бойцов и добавлен параметр количество попыток выполнить болевой прием (среднее значение за 15 мин).

Выход нейросети был оставлен без изменений.

Исходное множество [6] было приведено к нужному виду и заново очищено от выбросов с помощью новой модели. Также было решено увели-

чить множество, поменяв местами бойцов, соответствующие параметры и выходной параметр, таким образом сразу показать нейросети, что вероятность второго бойца победить равна (1–вероятность победить первого бойца). Такое не было учтено в первой модели и нейросети было труднее выявить зависимость. Обучающее множество состояло из 4954 примеров. Тестовое множество включало в себя 560 примеров.

### Модель 3

Прежде чем переходить к прогнозированию была составлена еще одна модель, в которой были исключены параметры: пол бойцов и разница в количестве попыток выполнить болевой прием. Данная модель обучалась на данных UFC боев с 1993 по 2021 [4]. Данные были приведены к нужному виду и очищены от выбросов. В модель были добавлены данные о «ничьих», которые не рассматривались в первых моделях. Обучающее множество включало в себя 3878 примера, тестовое – 440.

Выходным параметром также является вероятность победить первого бойца. Ничьей в выходных данных будет 0,5.

Все три нейросетевых модели работают адекватно, действительный результат отклонялся от желаемого в допустимых пределах (рисунок 1.1 – 1.3).

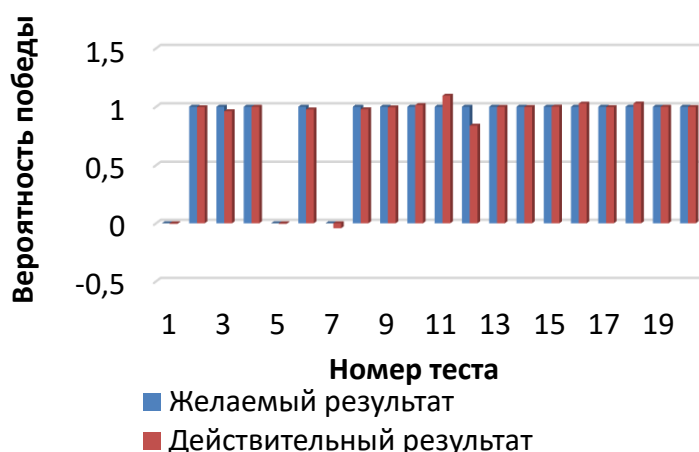


Рисунок 1.1. Адекватность модели №1

С помощью программы «Нейросимулятор 5» [9] была получена значимость параметров (для примера, на рисунке 2 показана значимость параметров третьей модели).

Для первой модели выявилось, что значимость параметров, относящихся к первому бойцу выше, чем значимость тех же параметров второго бойца. Так происходит и с другими одинаковыми параметрами двух бойцов. Таким образом первая модель может ошибаться в предсказаниях, так как не учитывает шансы бойцов адекватно. Для второй модели самыми значимыми параметрами оказались разница в точности борьбы, в обхвате рук и разница в поражениях. Пол бойцов оказался самым незначимым параметром, как и у первой нейросети. Наиболее значимым параметром в третьей модели является разница в возрасте, а разница в точности борьбы имеет самую низкую значимость в отличие от модели №2.



Рисунок 1.2. Адекватность модели №2

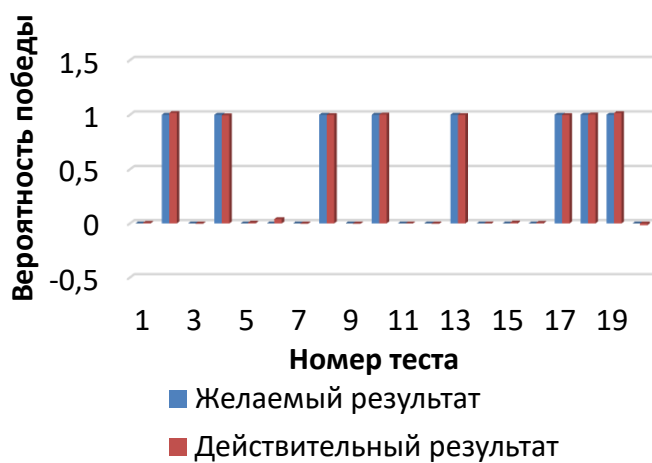


Рисунок 1.3. Адекватность модели №3

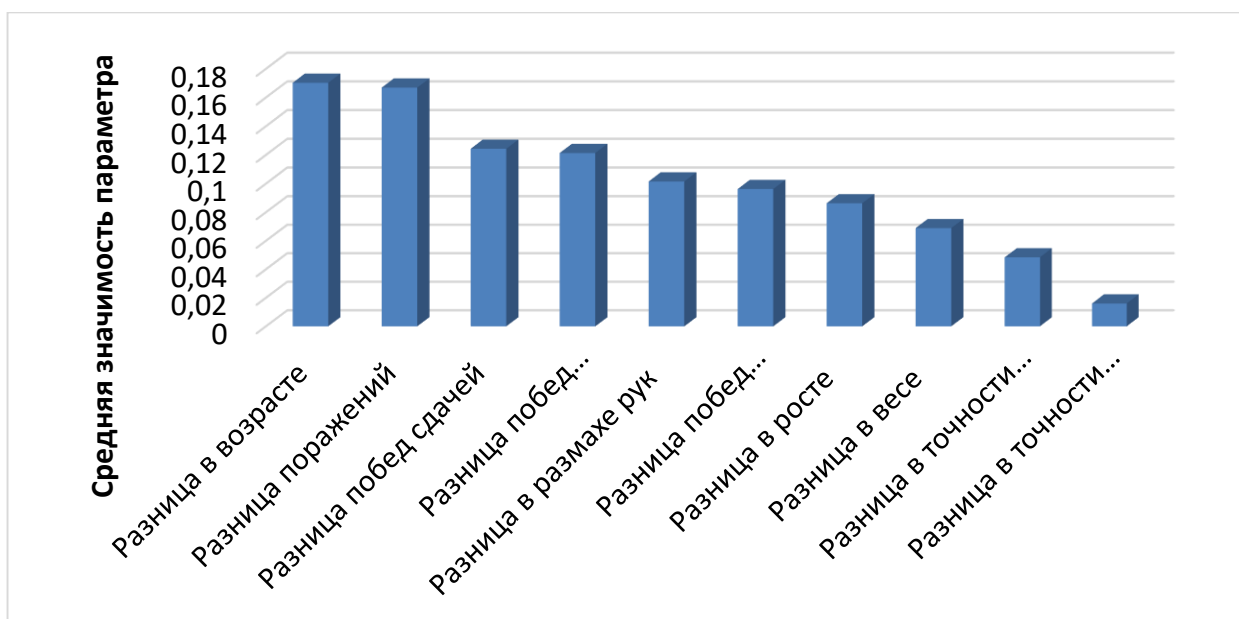


Рисунок 2. Средняя значимость параметра модели №3

Все нейросети имеют разную значимость тех или иных параметров. Модель №2 и модель №3 учитывают показатели бойцов совместно, первая же нейросеть дает разную значимость того или иного показателя в зависимости от номера бойца, что может давать неверные результаты при смене порядка бойцов.

## 2. Сравнение моделей на основе главного боя UFC 269.

Для прогноза был взят бой, который пройдет 11 декабря на турнире UFC 269, за титул чемпиона между Дастином Порье и Чарльзом Оливейрой. Различные букмекерские компании, к примеру, WinLine, 1xСтавка, прогнозируют победу Порье [8].

Результаты прогнозов построенных нейросетевых моделей, обученных на исторических данных представлены в таблице 2.

**Таблица 2**

### Результаты прогнозов

Номер модели	Вероятность победы Дастина Порье	Вероятность победы Чарльза Оливейры
1	0,0139	0,9453
2	0,0023	0,9975
3	-0,045	1,053

Как видно из таблицы, все три нейросетевые модели предсказали победителем Чарльза Оливейру.

Далее было проведено исследование нейросетевых моделей методом замораживания, который состоит в том, чтобы изменяя один параметр и оставляя без изменений другие параметры, выявлять зависимость между изменяемым параметром и выходом сети [10, 11].

Но первая нейросетевая модель при использовании этого метода давала довольно нереалистичные результаты. Поэтому метод замораживания был применен только ко вторым двум моделям. Первая модель оказалась непригодна для прогнозирования исхода боев.

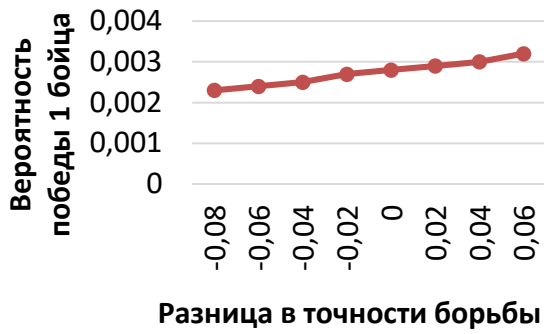
Начальные входные параметры для второй и третьей модели представлены в таблице 3 [1, 3, 5] (боец 1 – Д. Порье, боец 2 – Ч. Оливейра).

**Таблица 3**

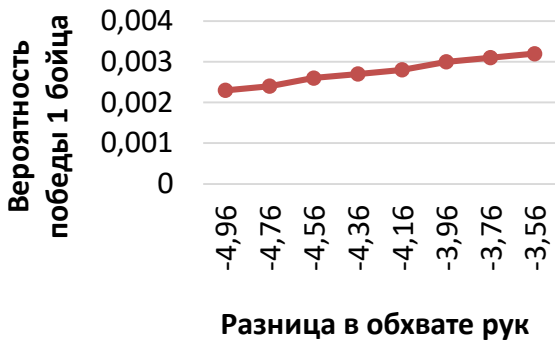
### Начальные входные параметры для моделей

Пол (только для модели 2)	Разница поражений первого и второго бойца	Разница побед решением	Разница побед нокаутом	Разница побед сдачей	Разница в росте	Разница в обхвате рук	Разница в весе	Разница в возрасте	Разница в точности ударов	Разница в попытках выложить болевой (только для модели 2)	Разница в точности борьбы
1	-2	4	5	12	2,54	4,96	0	0	-0,02	-1,49	-0,08

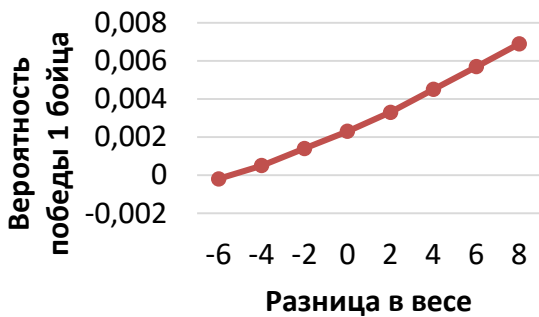
На рисунках 3.1-3.8 представлено сравнение двух моделей с помощью метода замораживания.



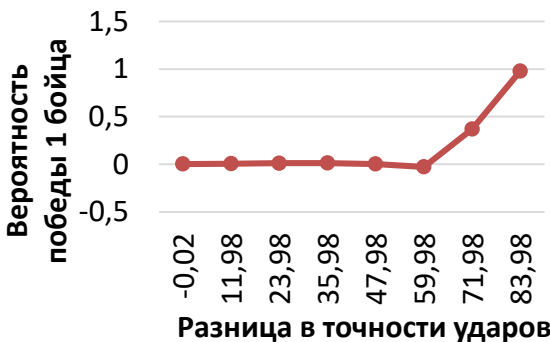
**Рисунок 3.1. Разница в точности борьбы, модель 2**



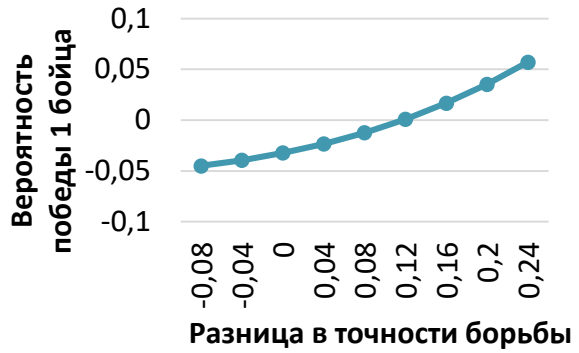
**Рисунок 3.3. Разница в объёме рук, модель 2**



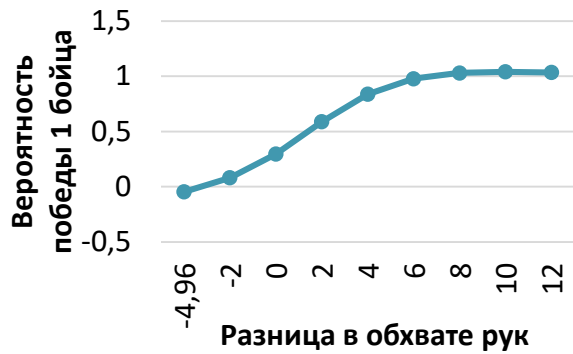
**Рисунок 3.5. Разница в весе, модель 2**



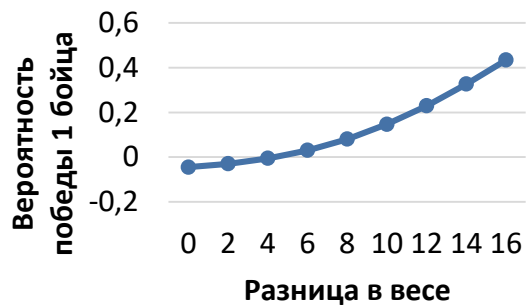
**Рисунок 3.7. Разница в точности ударов, модель 2**



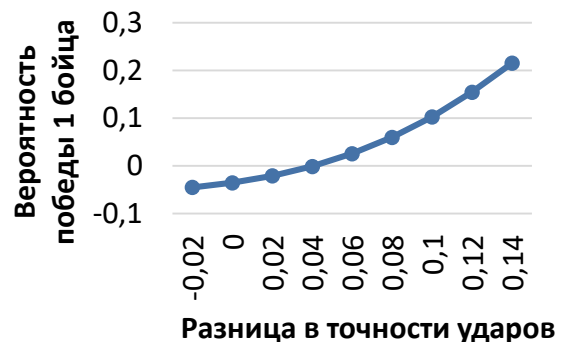
**Рисунок 3.2. Разница в точности борьбы, модель 3**



**Рисунок 3.4. Разница объёме рук, модель 3**



**Рисунок 3.6. Разница в весе, модель 3**



**Рисунок 3.8. Разница в точности ударов, модель 3**

По графикам видно, что модель 3 ведет себя более адекватно и результаты третьей модели больше соответствуют реальности, чем модели два. Третья модель при увеличении разности какого-то параметра увеличивает шансы победы соразмерно, чего нельзя сказать о модели 2, которая, к примеру, при увеличении разности в объёме рук или в точности борьбы увеличивает шансы на тысячные доли.

Такое различие нейросетей, возможно, обусловлено тем, что они обучались на разных данных и вполне вероятно, что для второй сети либо обучающие данные содержали противоречивые примеры, либо анализируемый бой непохож ни на один из тех, на которых обучалась сеть.

Результаты, полученные с помощью нейросетевой модели 3 больше соответствуют реальности, чем результаты метода замораживания для первых двух моделей, поэтому третью модель можно использовать для прогнозирования исходов боев.

Таким образом, исходя из графиков для модели 3 (рисунок 3), чтобы Дастина Порье увеличить шансы на победу, ему нужно повышать ударные навыки, а также можно набрать вес в пределах весовой категории. Чтобы повысить шансы ему также можно улучшать борцовские навыки, но навыки в ударной технике имеют большую значимость.

**Заключение.** В ходе работы было разработано и проанализировано три нейросетевых модели. С помощью метода замораживания была определена лучшая из трех моделей.

Данные исследования можно применить для прогноза на других площадках, потому как входными параметрами нейросети являются показатели и статистики бойцов, но возможно отклонения будут больше, потому как сеть обучалась на данных только одной площадки.

Также в процессе работы был предсказан результат главного боя турнира UFC 269 за титул в легком весе. Впоследствии оказалось, что прогноз верный, Чарльз Оливейра сумел защитить пояс.

### Библиографический список

1. FightTime – Все новости MMA, UFC, Bellator, M-1 Global, K1 [Электронный ресурс]. – URL: <https://fighttime.ru/> (дата обращения: 05.02.2021).
2. MMALegeng – портал о последних событиях в мире MMA [Электронный ресурс]. – URL: <https://mmalegend.ru/dustin-poirier/> (дата обращения: 05.02.2021).
3. Stats | UFC [Электронный ресурс]. – URL: <http://www.ufcstats.com/statistics/events/completed> (дата обращения: 05.02.2021).
4. UFC Historical Data from 1993 to 2021 | Kaggle [Электронный ресурс]. - URL: <https://www.kaggle.com/rajeevw/ufcdata> (дата обращения: 02.02.2021).
5. UFC Russia – трансляции, бойцы, видео, рейтинги [Электронный ресурс]. – URL: <https://ufc.ru/> (дата обращения: 05.02.2021).
6. Ultimate UFC Dataset | Kaggle [Электронный ресурс]. URL: <https://www.kaggle.com/mdabbert/ultimate-ufc-dataset> (дата обращения: 02.02.2021).

7. Yasnitsky L.N. Algorithm for searching and analyzing abnormal observations of statistical information based on the Arnold – Kolmogorov – Hecht-Nielsen theorem // International Journal of Advanced Trends in Computer Science and Engineering. 2020. 9(2). Pp. 1814-1819.

8. Чарльз Оливейра – Дастин Порье: коэффициенты и ставки на бой UFC 269 11 декабря 2021 [Электронный ресурс]. – URL: <https://legalbet.ru/best-posts/pore-olivejra-stavki-i-koeffitcienti-na-boj/> (дата обращения: 04.02.2021).

9. Черепанов Ф.М., Ясницкий Л.Н. Нейросимулятор 4.0. Свидетельство о регистрации программы для ЭВМ RUS 2014612546. Заявка № 2014610341 от 15.01.2014.

10. Ясницкий Л.Н. Введение в искусственный интеллект. М.: Издательский центр «Академия», 2005. – 176 с.

11. Ясницкий Л.Н. Интеллектуальные системы. – М.: Лаборатория знаний, 2016. – 221 с.

12. Ясницкий Л.Н., Абрамова Ю.С., Бабушкина С.Д. Возможности получения рекомендаций по улучшению результативности сборных команд, готовящихся к участию в чемпионате Европы по футболу Евро-2016 методом нейросетевого моделирования // Вестник спортивной науки. 2015. № 5. С. 15-20.

13. Ясницкий Л.Н., Внукова О.В., Черепанов Ф.М. Прогноз результатов олимпиады-2014 в мужском одиночном фигурном катании методами искусственного интеллекта // Современные проблемы науки и образования. – 2014 – № 1 – С. 189.

14. Ясницкий Л.Н., Кировоса А.В., Ратегова А.В., Черепанов Ф.М. Прогноз результатов чемпионата мира-2015 по легкой атлетике методами искусственного интеллекта // Современные проблемы науки и образования. – 2014 – № 4; URL: [www.science-education.ru/118-14423](http://www.science-education.ru/118-14423) (дата обращения: 30.01.2021).

15. Ясницкий Л.Н., Павлов И.В., Черепанов Ф.М. Прогнозирование результатов олимпийских игр 2014 года в неофициальном командном зачете методами искусственного интеллекта // Современные проблемы науки и образования. 2013. № 6. С. 985.

## **PREDICTION OF THE MIXED MARTIAL ARTS FIGHTS' RESULTS BY NEUTRAL NETWORK MODELING**

*Verholantsev Philipp A.*

Perm State University,

614990, Russia, Perm, st. Bukireva, 15, [filipp.verh.olantsev@yandex.ru](mailto:filipp.verh.olantsev@yandex.ru)

The article describes comparisons of three neural network models. Models predict the results of mixed martial arts fights. The best model had been defined and used to evaluate the result of the UFC 269 title fight on December 11, 2021. The model is based on the physical characteristics of the fighters and on histori-

cal data of fights. The model is based on data from the UFC, but can be used to predict the fights' results of other organizations.

**Keywords:** neural network model, neural network, freezing method, UFC, predicting results, neural network training.

УДК 004.032.26

## НЕЙРОСЕТЕВАЯ СИСТЕМА ПРОГНОЗИРОВАНИЯ РЕЗУЛЬТАТОВ В ЖЕНСКОМ ОДИНОЧНОМ ФИГУРНОМ КАТАНИИ НА ОЛИМПИАДЕ 2026

*Мицкевич Анфиса Дмитриевна*

Национальный исследовательский университет «Высшая школа экономики»  
614107, Россия, г. Пермь, ул. Бульвар Гагарина, 37а,  
e-mail: amits2000@mail.ru

Статья представляет из себя процесс разработки нейросетевой системы для прогнозирования результатов зимней олимпиады в женском одиночном фигурном катании 2026 года. Данная система, обученная на результатах предыдущих соревнований по фигурному катанию: олимпиады предыдущих лет, чемпионаты мира и Европы, этапы Гран-при, с большой точностью предсказывает шансы на победу в следующей зимней олимпиаде среди претендентов от России. Система позволяет оценить влияние определенных параметров, которые характеризуют спортсменов, на их результаты, а также подобрать идеальное сочетание параметров для каждого из претендентов. Путем исследования получены рекомендации для следующих претендентов от России: Софья Акатьева, Софья Самоделкина, Аделия Петросян, Вероника Жилина, направленные на улучшение их спортивных результатов.

**Ключевые слова:** Олимпиада 2026, искусственный интеллект, прогноз результатов, фигурное катание, Акатьева, Самоделкина, Петросян, Жилина.

**Введение.** Результаты спортивных соревнований давно начали предсказывать, строить прогнозы, предугадывать победителей определенного вида спорта. Как показывает опыт результаты спортсменов во многом зависят от их показателей, характеристик, что позволяет более точно определить сочетание показателей спортсмена для его успешного выступления на соревнованиях. Перспективным инструментом для построения моделей прогнозирования являются методы искусственного интеллекта – нейросетевые технологии [1-3].

При анализе литературных источников было выяснено, что работы с нейросетевыми технологиями ведутся довольно давно. Большой вклад в развитие технологий был внесен Пермской научной школой искусственного интеллекта. Фигурное катание не стало исключением при исследовании при



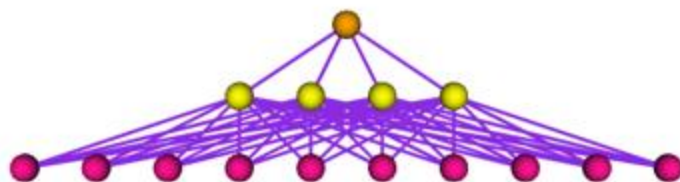
помощи нейросетевой системы. Как известно, работу на данную тематику уже проводили Ясницкий Л. Н., Внукова О. В., Черепанов Ф. М. [4]. В своей работе они на основании анализа предыдущих соревнований по фигурному катанию выявили параметры, которые способствовали прогнозированию результатов среди мужчин одиночников. Из представленной работы следует, что на успешность выступления претендентов влияют возраст, вес и рост.

Основной целью настоящей работы является составление рекомендаций по сочетанию значимых параметров для спортсменок из России, которые могут претендовать на призовые места на зимних олимпийских играх в 2026 году в одиночном фигурном катании.

**Методика прогнозирования.** Для построения нейросетевой модели в прогнозировании результатов в женском одиночном фигурном катании на Олимпиаде в 2026 году были проанализированы результаты соревнований по фигурному катанию международного стандарта ISU за предыдущие годы: чемпионаты мира и Европы, этапы Гран-при, Олимпиада 2022 года. В результате было сформировано множество примеров для обучения системы. Входные параметры модели, по которым обучалась система, были взяты из доступных источников, протоколов соревнований [5, 6]. В результате были отобраны следующие параметры:  $X_1$  – рейтинг страны (проанализировав результаты, был составлен рейтинг стран),  $X_2$  – возраст,  $X_3$  – рост,  $X_4$  – вес,  $X_5$  – возраст начала тренировок,  $X_6$  – лучший результат за элементы,  $X_7$  – лучший результат за компоненты,  $X_8$  – наличие травм в текущем сезоне,  $X_9$  – наличие мирового рекорда,  $X_{10}$  – количество медалей (значимые соревнования в мире фигурного катания). Выходной параметр модели  $Y_1$  – место на соревнованиях. Было принято решение исследовать первые 5 мест каждого из соревнований.

Множество примеров было составлено вручную путем сбора информации о спортсменках из различных источников. В результате было сформировано 60 примеров для построения модели. Данное множество было разделено на обучающее и тестирующее в соотношении 80% и 20% соответственно. Следовательно обучающее множество состоит из 48 примеров, а тестирующее из 12.

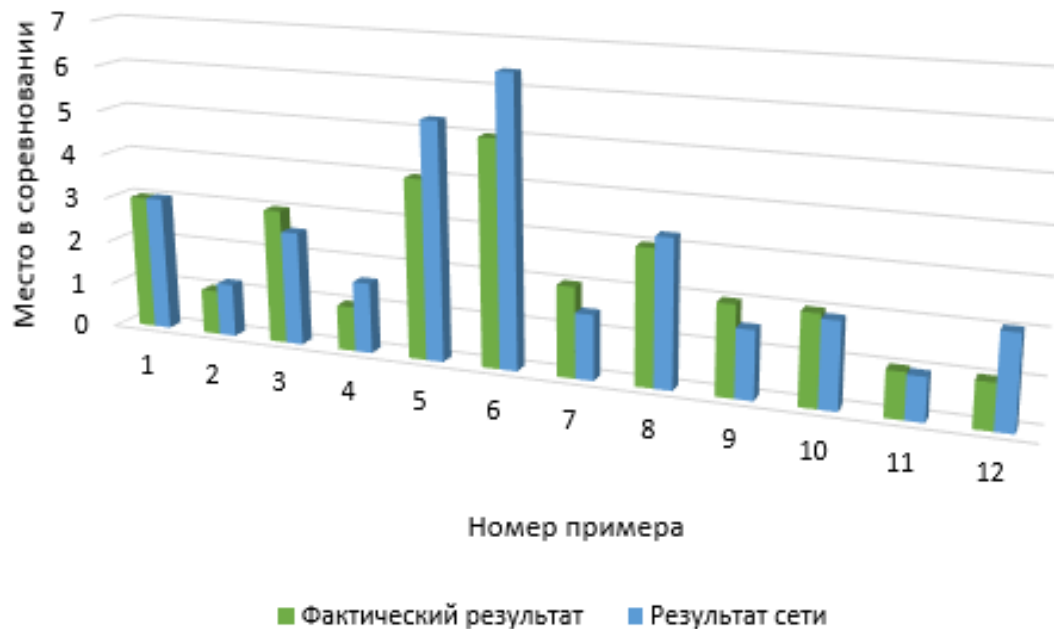
Проектирование, обучение, тестирование нейронной сети выполнялись с помощью программы «Нейросимулятор 5» [7] по методике [2]. В результате оптимизации спроектированная нейронная сеть представляла собой персептрон, который имеет десять входных нейронов, один выходной и один скрытый слой с четырьмя нейронами (рисунок 1).



**Рисунок 1. Структура нейронной сети**

Для оценки корректной работы спроектированной нейронной сети использовалось тестирующее множество, состоящее из 12 примеров. Результаты проверки (рисунок 2) показывают, что полученные с помощью нейросети

прогнозные значения отличаются от фактических не более чем на 13,3%, что свидетельствует о том, что нейронная сеть, хоть и с небольшой погрешностью, но усвоила закономерности моделируемых процессов и что теперь эти закономерности можно изучать путем исследования полученной математической модели.



**Рисунок 2. Результаты тестирования системы**

Оценка значимости входных параметров была выполнена с помощью программы «Нейросимулятор 5» (рисунок 3).

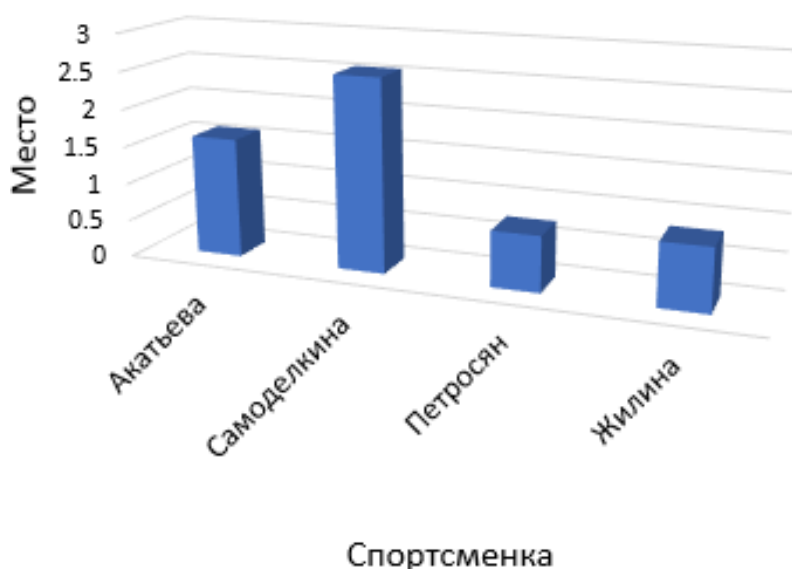


**Рисунок 3. Значимость входных параметров**

Как видно наиболее значимым параметром является лучший результат за элементы в программе, вес и рост спортсменки. Как и ожидалось, что

наиболее влиятельным параметром окажется лучший результат за элементы в программе, так как именно от набранных баллов и сложность выполняемых элементов (прыжков, дорожек, вращений) зависит итоговое место на соревнованиях. Однако необходимо понимать, что и рост, и вес спортсменки во многом определяют ее результативность и успешность.

**Результаты экспериментов.** Перейдем к прогнозированию результатов претенденток от России на зимней олимпиаде в 2026 году. К ним по существующим данным [8] относятся Софья Акатьева, Софья Самоделкина, Аделия Петросян, Вероника Жилина. Исследуем их текущие показатели, чтобы определить их шансы на победу (рисунок 4).

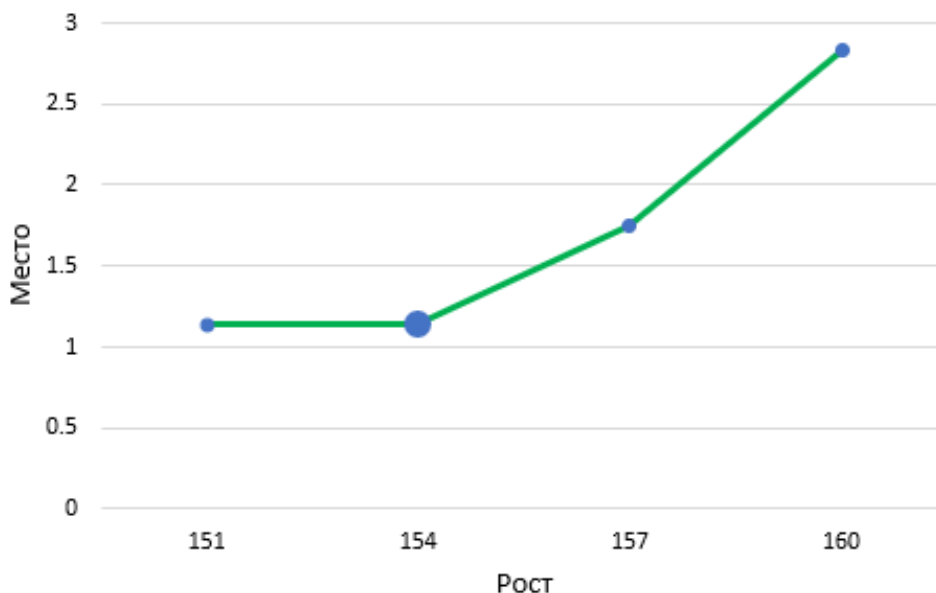


**Рисунок 4. Результаты спортсменок по текущим параметрам**

Можно заметить, что у Петросян и Жилиной есть все шансы с текущими показателями занять первое место. Спортсменка Акатьева уступает своей соперницам и по прогнозам может занять как первое, так и второе место. Самоделкина будет бороться с ней за второе место. Однако необходимо точно понимать, что у претенденток изменится возраст к моменту соревнований, а также поменяются параметры роста и веса.

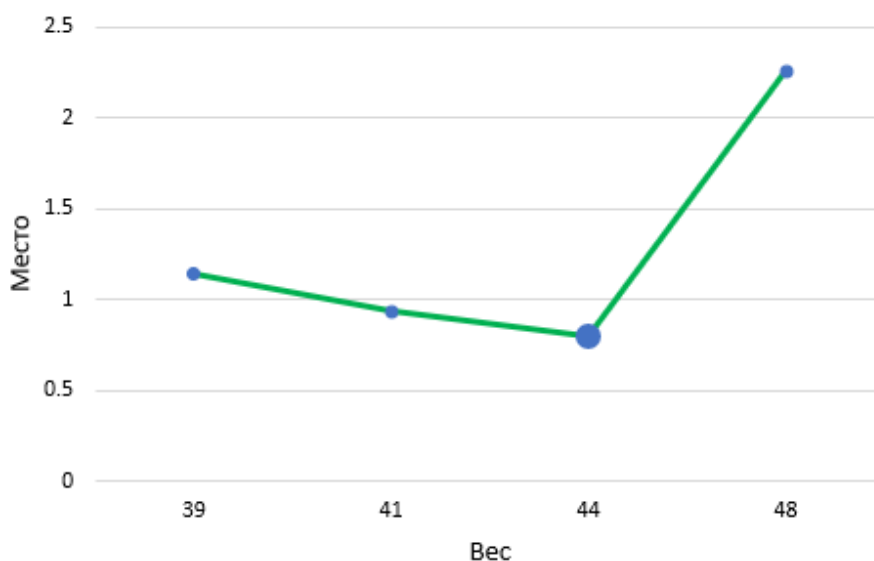
Как было выяснено ранее на результативность спортсменки влияют определенные параметры. Проведем исследование, где будем менять параметры спортсменок, чтобы понимать в какой комбинации шансы на победу будут выше. На графиках жирный маркер будет обозначать самый эффективный показатель для спортсменки.

Начнем со спортсменки Софьи Акатьевой. На период следующих олимпийских игр ее возраст будет 18 лет. Поэтому будем редактировать параметры роста и веса. Начнем с роста спортсменки. Предположим, что у Софьи будет меняться только рост, тогда мы получим результаты, при которых наиболее успешное выступление будет при текущем росте в 154 см, а чем выше, тем шанс на победу будет ниже (рисунок 5).



**Рисунок 5. Изменение роста Акатъевой**

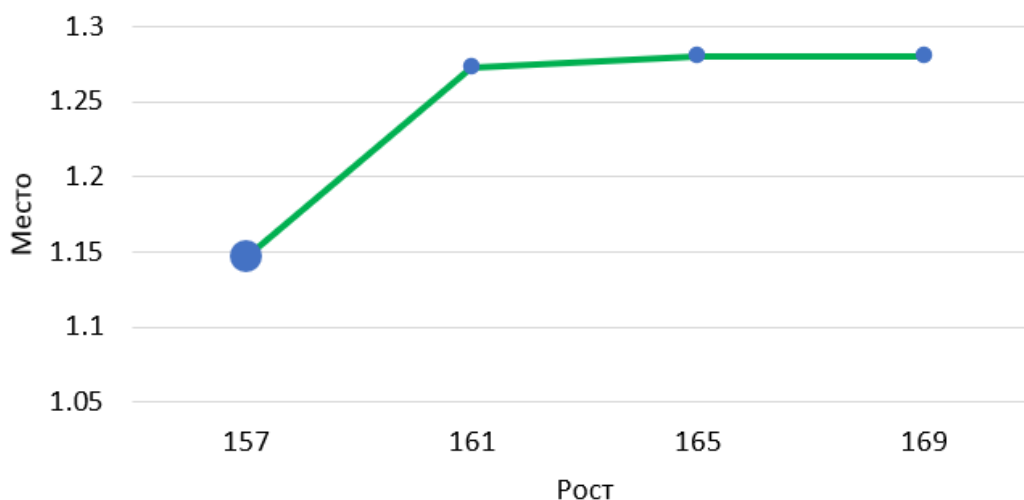
Тогда попробуем корректировать вес спортсменки, оставляя остальные параметры неизменными (рисунок 6). По результатам видно, что с текущими параметрами Софье необходимо набрать до 5 кг веса для более результативного выступления.



**Рисунок 6. Изменение веса Акатъевой**

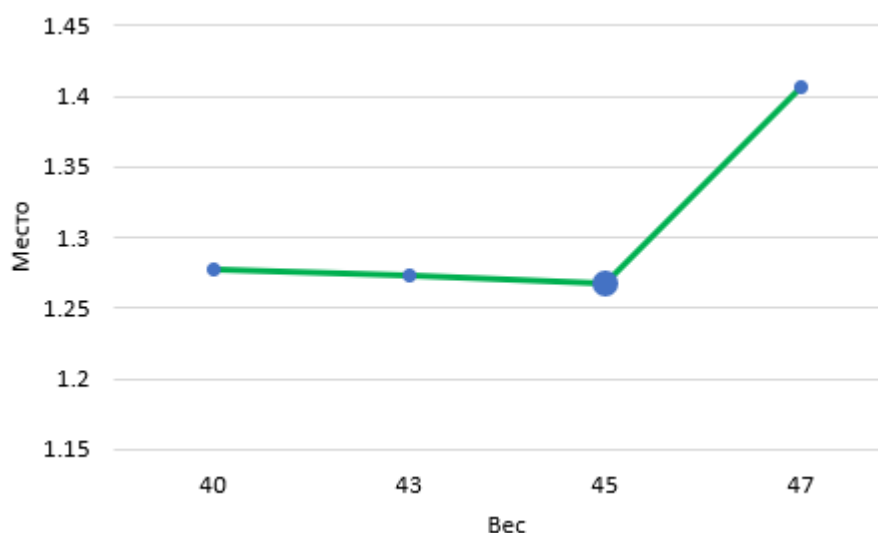
Таким образом, можно сделать выводы, что Софье Акатъевой в идеале было бы лучше прийти в олимпийский сезон с весом до 44 кг.

Далее посмотрим на параметры Софьи Самodelкиной. У Софьи уже видно, что рост выше, что средний среди фигуристок – 161 см. Попробуем откорректировать данный параметр (рисунок 7). Результативнее Софья была бы с более низким ростом, однако необходимо понимать, что спортсменка сможет уже только вырасти. Поэтому основываясь на полученных результатах, видно, что вероятность занять первое место у нее выше, если она больше не будет расти.



**Рисунок 7. Изменение роста Самоделкиной**

Рассмотрим варианты, когда вес спортсменки будет меняться (рисунок 8). В случае с данной спортсменкой также необходимо набрать 2 кг веса. А при потере веса результаты могут еще ухудшиться.



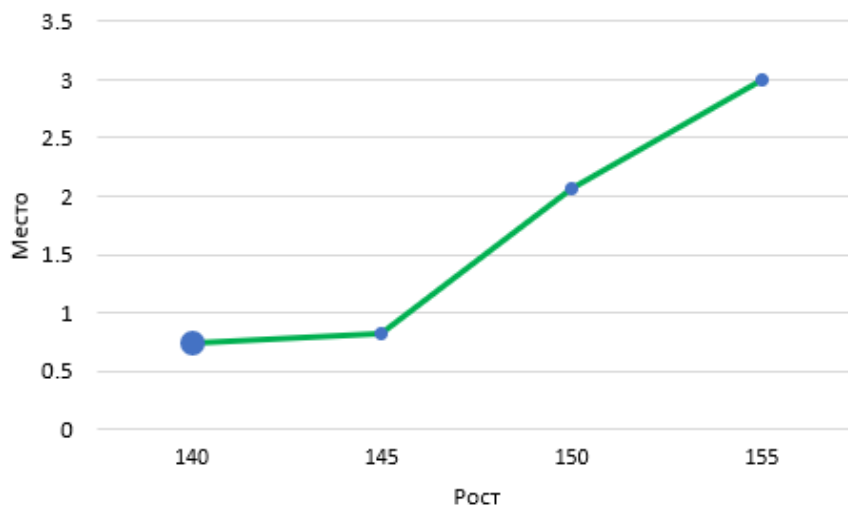
**Рисунок 8. Изменение веса Самоделкиной**

Далее будем рассматривать корректировку параметров Аделии Петросян. Аналогично необходимо не забыть, что в олимпийский сезон у спортсменки изменится возраст. На данный момент рост спортсменки 140 см, будем его корректировать и отслеживать результативность (рисунок 9). Делаем выводы, что Аделия сейчас находится в идеальном росте, однако даже если она вырастет на 5 см на результатах это сильно не отразится.

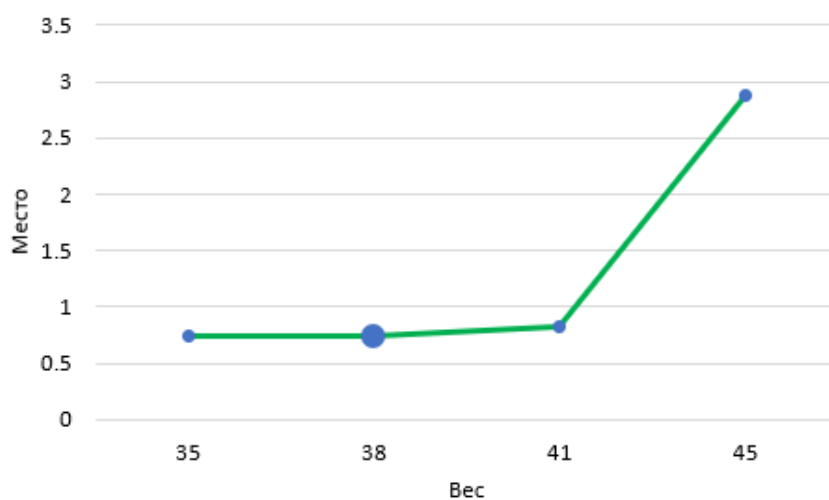
Посмотрим, как изменение веса будет сказываться на успешности спортсменки (рисунок 10). При сохранении параметров и изменении веса Аделии до 41 кг результаты будут по-прежнему оставаться успешными, однако при наборе веса уже в 45 кг шансы на победу резко сократятся.

Последней претенденткой на победу является Вероника Жилина. Текущие показатели спортсменки уже очень хорошие для победы. Однако посмотри на изменение результатов при корректировке параметров. Аналогич-

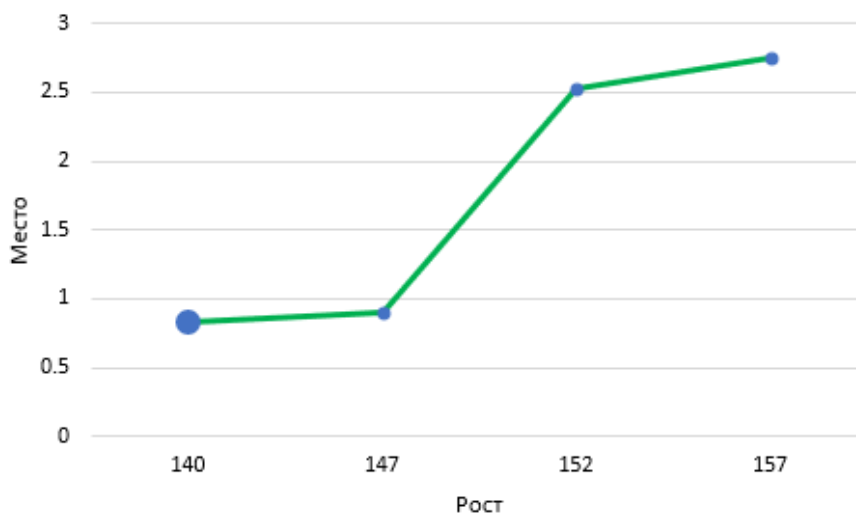
но начнем с роста спортсменки (рисунок 11), который на данный момент 140 см. При увеличении роста показатели будут также эффективны до 147 см, однако после вероятность победы после будет гораздо меньше.



**Рисунок 9. Изменение роста Петросян**

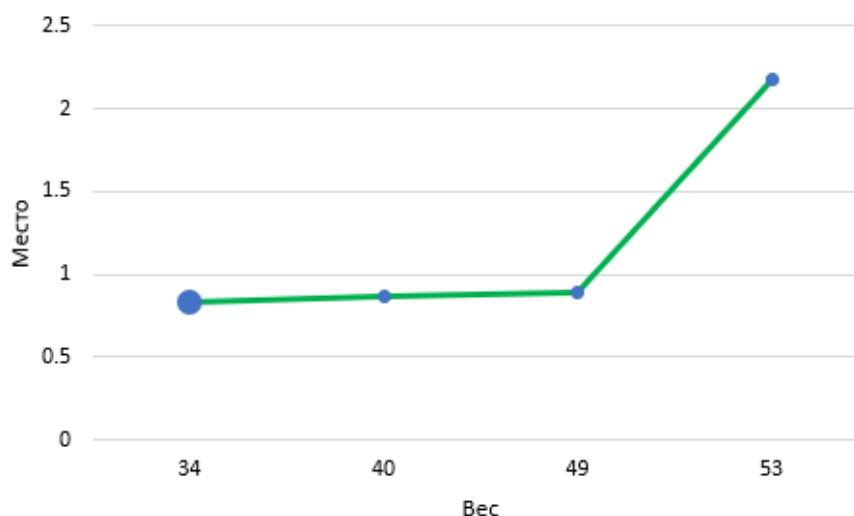


**Рисунок 10. Изменение веса Петросян**



**Рисунок 11. Изменение роста Жилиной**

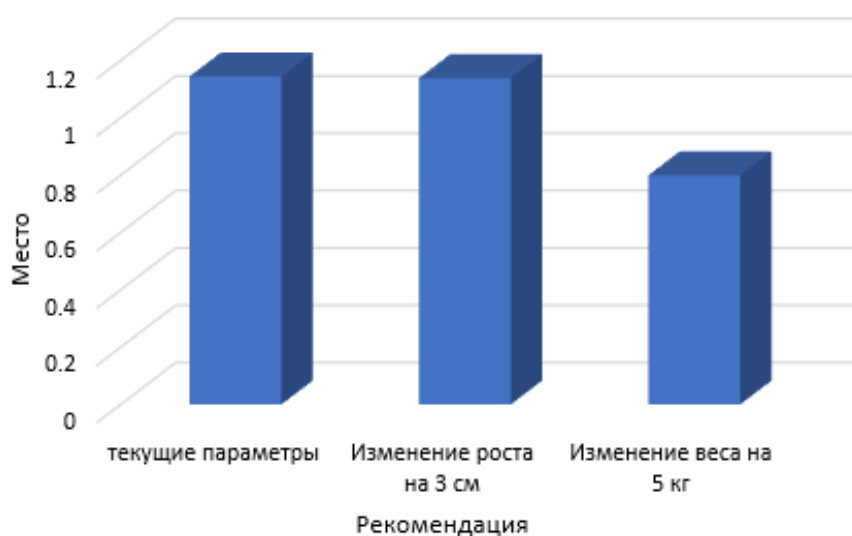
И оценим, насколько корректировка параметра веса отразится на показателях итогов (рисунок 12). С текущего веса в 34 кг спортсменка может прибавить в весе до 49 кг и все равно иметь все шансы на победу, но после достигнутого веса успешность будет снижаться.



**Рисунок 12. Изменение веса Жилиной**

**Рекомендации.** На основании полученных результатов и исследований приведем рекомендации для спортсменок. Отследим, при изменении каких параметров спортсменка будет достигать наилучших результатов. Помимо измененных роста и веса спортсменки будем также сравнивать с текущими показателями.

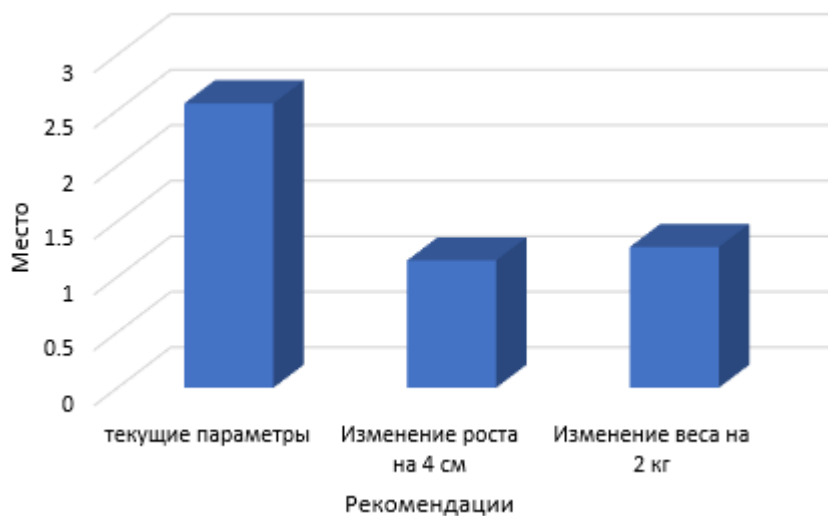
Спортсменке Софье Акатьевой рекомендуется либо набрать веса до 44 кг, либо вырасти на 3 см (рисунок 13). Эффективнее выступление будет при наборе веса.



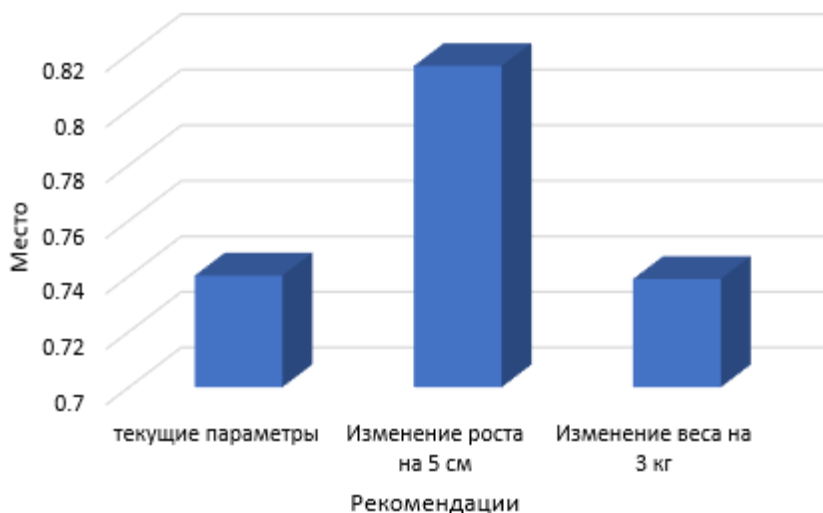
**Рисунок 13. Рекомендации для Акатьевой**

Софья Самоделкина имеет рост выше среднего для фигуристки, отчего более результативной спортсменка была с ростом ниже. Поскольку мы по-

нимаем, что в обратную сторону Софья расти не сможет, наилучшая рекомендация для нее – набрать в весе 2 кг (рисунок 14).



**Рисунок 14. Рекомендации для Самоделкиной**



**Рисунок 15. Рекомендации для Петросян**



**Рисунок 16. Рекомендации для Жилиной**



Фигуристка Аделия Петросян уже обладает очень хорошими данными для победы в соревнованиях (рисунок 15). Можно заметить, что наилучших результатов спортсменка будет достигать и с текущими параметрами, и с изменением роста и веса до определенных показателей.

Аналогично спортсменка Вероника Жилина уже обладает всеми данными для победы в турнире, однако даже при корректировке параметров фигуристка будет способна побеждать. Проанализировав ее показатели, можно выявить следующие рекомендации (рисунок 16).

**Заключение.** Разработана нейросетевая модель, с помощью которой выполнен прогноз результатов олимпийских игр 2026 года по фигурному катанию среди женщин одиночек – претенденток от России. Исследовано влияние на рейтинг некоторых параметров спортсменок и сделана попытка разработки рекомендаций по улучшению их спортивных результатов.

### Библиографический список

1. Ясницкий Л. Н., Богданов К. В., Черепанов Ф. М. Технология нейросетевого моделирования и обзор работ пермской научной школы искусственного интеллекта // *Фундаментальные исследования*. 2013. № 1–3. С. 736–740.

2. Ясницкий Л. Н., Бондарь В. В., Бурдин С.Н. и др. Пермская научная школа искусственного интеллекта и ее инновационные проекты. 2-е изд. / НИЦ "Регулярная и хаотическая динамика". М.; Ижевск, 2008. 75 с.

3. Ясницкий Л.Н. Введение в искусственный интеллект. М.: Изд. центр "Академия", 2010. 176 с.

4. Ясницкий Л. Н., Внукова О. В., Черепанов Ф. М. Прогноз результатов олимпиады-2014 в мужском одиночном фигурном катании методами искусственного интеллекта // *Современные проблемы науки и образования*. – 2014. – № 1. – С. 189.

5. Результаты чемпионатов по фигурному катанию, расположенные на сайте Чемпионат. – Режим доступа: <https://www.championat.com/figure-skating/news-4464001-figurnoe-katanie-raspisanie-sezona-2021-2022.html> – Дата обращения: 13.02.2022.

6. Результаты чемпионатов по фигурному катанию, расположенные на сайте ISU. – Режим доступа: <https://www.isu.org/figure-skating> – Дата обращения: 13.02.2022.

7. Черепанов Ф. М., Ясницкий Л. Н. Нейросимулятор 4.0. Свидетельство о регистрации программы для ЭВМ RUS 2014612546. Заявка № 2014610341 от 15.01.2014.

8. На кого мы можем рассчитывать через четыре года в Милане. – Режим доступа: <https://www.sport-express.ru/figure-skating/reviews/figurnoe-katanie-olimpiada-2026-kto-iz-odinochnic-mozhet-vystupit-za-sbornuyu-rossii-v-milane-akateva-samodelkina-muraveva-mnenie-1895319/> – Дата обращения: 20.02.2022.

# NEURAL NETWORK SYSTEM FOR PREDICTING RESULTS IN WOMEN'S FIGURE SKATING AT THE 2026 OLYMPICS

*Mitskevich Anfisa D.*

National Research University "Higher School of Economics",  
Str. Bulvar Gagarina, 37a, Perm, Russia, 614107, amits2000@mail.ru

This article presents the process of developing a neural network system to predict the results of the 2026 Winter Olympics women's singles figure skating. This system, trained on the results of previous figure skating competitions: the Olympics of previous years, World and European Championships, Grand Prix stages, predicts with great accuracy the chances of winning the next Winter Olympics among contenders from Russia. The system makes it possible to assess the impact of certain parameters that characterise athletes on their results, as well as to select the ideal combination of parameters for each of the contenders. The study made recommendations for the following Russian candidates: Sofia Akatieva, Sofia Samodelkina, Adelia Petrosyan and Veronika Zhilina to improve their performance.

**Keywords:** 2026 Olympics, artificial intelligence, performance prediction, figure skating, Akatieva, Samodelkina, Petrosyan, Zhilina.

УДК 004.032.26

## НЕЙРОСЕТЕВАЯ СИСТЕМА ПРОГНОЗИРОВАНИЯ ШАНСОВ ПОБЕДЫ СБОРНЫХ НА ЧЕМПИОНАТЕ МИРА ПО ФУТБОЛУ

*Павлов Денис Андреевич*

Национальный исследовательский университет «Высшая школа экономики»  
614990, Россия, г. Пермь, ул. Бульвар Гагарина, 37а, deniis.pavlov@mail.ru

В статье представлено описание разработки нейросетевой системы для прогнозирования шансов сборных на победу на чемпионате мира. Система позволяет с большой точностью предсказать шансы сборных на победу на чемпионате мира на основании данных предыдущих выступлений команд на предыдущих чемпионатах мира. С помощью разработанной интеллектуальной системы проведено исследование предметной области, выявлены закономерности, имеющие практической значение.

**Ключевые слова:** искусственный интеллект, нейросетевые технологии, прогнозирование, футбол, чемпионат мира по футболу.

**Введение.** Прогнозирование спортивных результатов является крайне сложной задачей, и футбол среди всех видов спорта представляет наибольшую сложность. Тем не менее, опыт Пермской научной школы искусствен-

ного интеллекта, например, [1-3], убедительно показывает, что правильно натренированная нейросетевая система может добиться успешности значительной части прогнозов. Главным News футбольным мировым турниром является чемпионат мира по футболу, он привлекает наибольшее количество фанатов по всему миру. Анализ и использование шансов сборных на победу на чемпионате мира может быть полезно для частных лиц, букмекерских компаний для увеличения прибыли компании.

Изучая литературные источники, выяснилось, что работ на тему прогнозирования шансов сборных на чемпионате мира достаточно мало, и результаты даже этих исследований с трудом позволяют говорить о точности их критериев. Так, BBC News в своей работе [4] среди критериев выделяли «попадание в первую корзину», «не пропустить больше 4 голов за 7 матчей», «быть европейской страной», «иметь большое количество матчей у игроков на международной арене», «иметь в своём составе лучшего вратаря». В данной работе автор использует результаты всех сборных чемпионата мира 2002, 2006, 2010, 2014, 2018 годов и всех призеров, начиная с 1994 года и использует большое количество параметров, для того, чтобы увеличить процент верных прогнозов.

Основная цель настоящей работы заключается в сборе множества публичных данных о результатах сборных на предыдущих чемпионатах мира на основе их статистических показателей, а также создание и обучение нейросетевой модели на этих данных. Конечным результатом должна быть нейросетевая система, способная прогнозировать шансы сборной на победу на чемпионате мира в больше, чем 83% случаев.

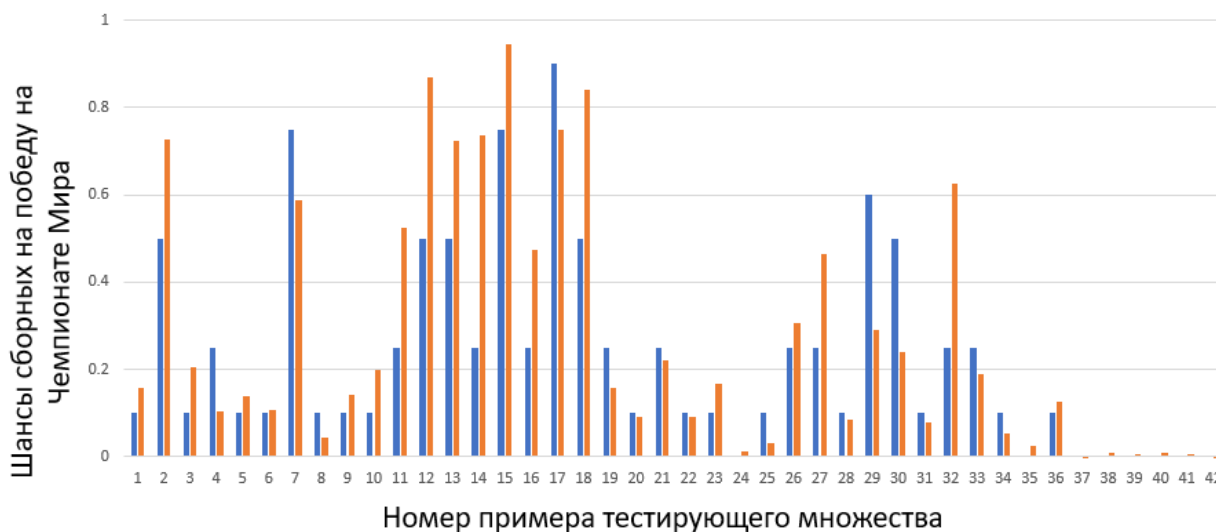
Для создания нейросетевой системы были выбраны следующие параметры: X1 – попадание сборной в первую корзину (0 или 1), X2 – количество забитых мячей/количество матчей, X3 – количество пропущенных мячей/количество матчей, X4 – наличие лучшего игрока в мире (0 или 1), X5 – место на предыдущем чемпионате Европы/кубке Америки по футболу, X6 – место на предыдущем чемпионате мира по футболу, X7 – количество лидеров сборной (8) играет в ТОП-5 лигах в лучших клубах (0, 0.5 или 1), X8 – ожидаемый прогноз по очкам в группе, X9 – является ли сборная хозяйкой чемпионата мира (0 или 1), X10 – является ли сборная лидером или аутсайдером (0 или 1), X11 – средний возраст состав сборной, X12 – свой тренер или чужой. Выходной параметр – шансы сборной победу на чемпионате мира по футболу.

Обучающее множество было собрано вручную. Перед переходом к проектированию и обучению нейросетевой модели, была выполнена очистка исходного множества от дубликатов. Таким образом, объем итогового множества включает в себя 150 примеров. Данное множество было разделено на обучающее и тестирующее в соотношении 66% и 34% соответственно. Большая часть данных была собрана с интернет-ресурсов [5-6].

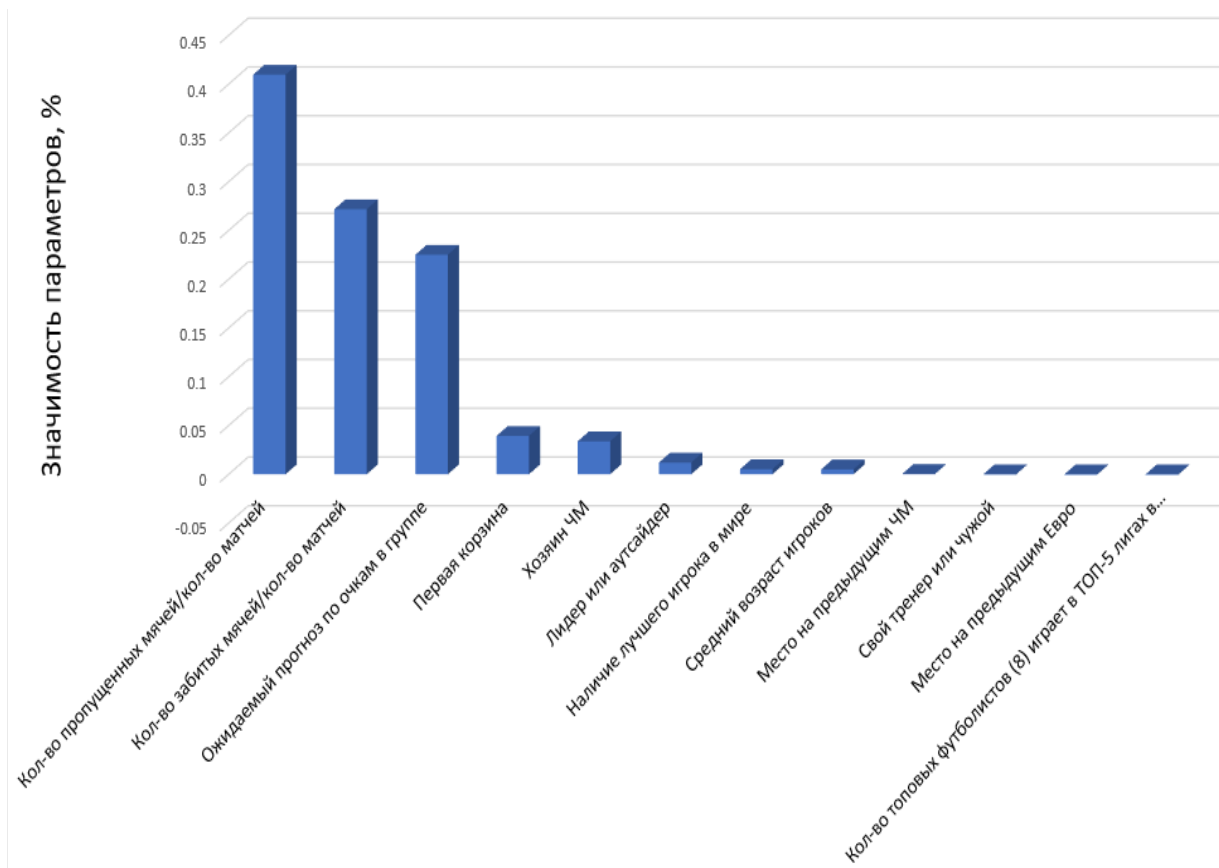
Проектирование, обучение, тестирование нейронной сети выполнялись с помощью программы «Нейросимулятор 5» [7] по методике [8]. В результате оптимизации спроектированная нейронная сеть представляла собой пер-

септрон, который имеет двенадцать входных нейронов, один выходной и один скрытый слой с одним нейроном.

Для оценки корректной работы спроектированной нейронной сети использовалось тестирующее множество, состоящее из 50 примеров.



**Рисунок 1. Результат тестирования нейронной сети**



**Рисунок 2. Значимость входных параметров нейросетевой модели**

Средняя относительная ошибка тестирования составила 13.05%, что можно считать приемлемым результатом. На рисунке 1 представлена гисто-

грамма, демонстрирующая разницу между фактическим и прогнозируемым нейросетью победителями матчей

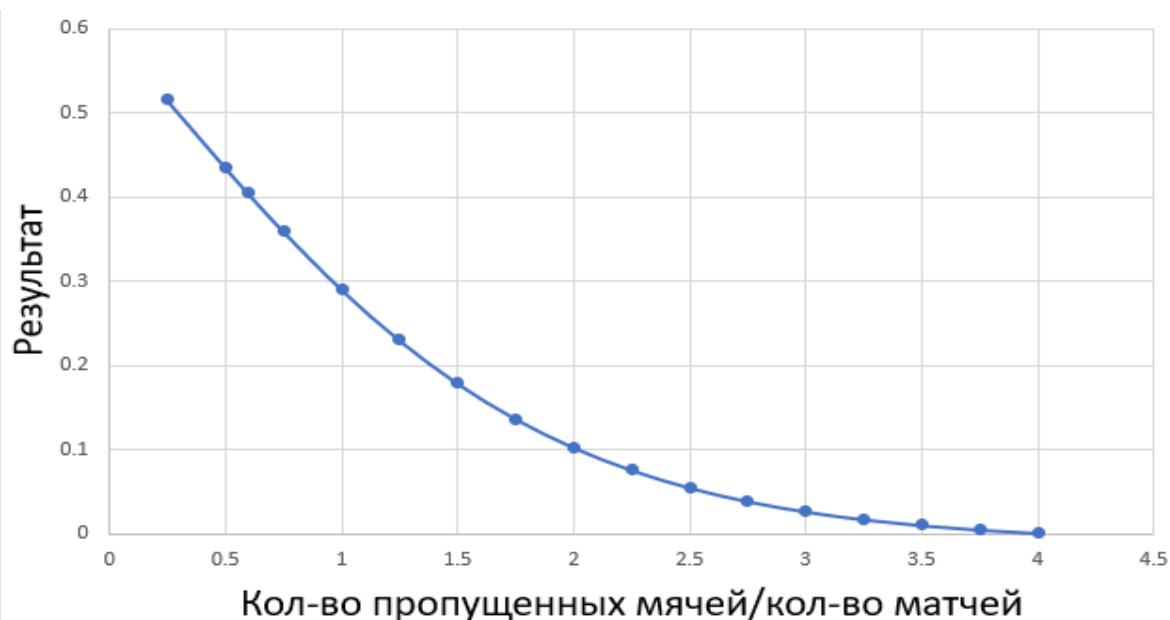
Из результатов, изображенных на рисунке 1, можно сделать вывод об адекватной работе спроектированной нейронной сети.

Оценка значимости входных параметров была выполнена с помощью программы «Нейросимулятор 5», результат отображен на рисунке 2.

Как видно из рисунка 2, наиболее значимыми являются количество пропущенных мячей/количество матчей, количество забитых мячей/количество матчей, ожидаемый прогноз по очкам в группе и т.д. Как и ожидалось, наиболее влиятельным параметром является количество пропущенных мячей/количество матчей. Это объясняется тем, что футбол крайне мало результативная игра (в среднем забивает от 0 до 2 мячей), и пропущенный гол означает большой задел для одной из команд.

Далее было проведено исследование полученных зависимостей между входными параметрами и шансами сборной на победу на чемпионате мира. Исследование производилось с помощью метода «замораживания» [1-3, 8], суть которого заключается в варьировании значения одного параметра и фиксации значений всех других параметров.

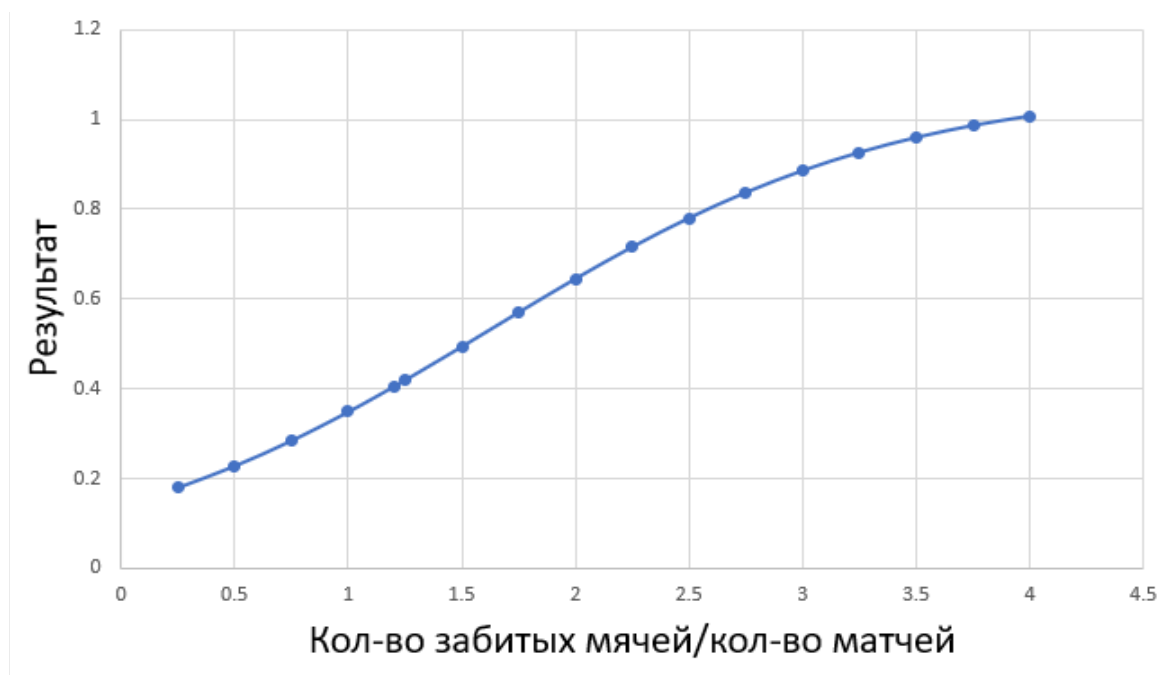
На рисунке 3 показан график зависимости шансов сборной на победу на чемпионате мира от количества пропущенных мячей в среднем за матч. В том случае, когда команда меньше пропускает, нейросеть прогнозирует ей увеличение шанса на победу.



**Рисунок 3. Зависимость шансов сборных на победу на чемпионате мира от количества пропущенных мячей/количество матчей**

На рисунке 4 продемонстрирована зависимость шансов сборных на победу на чемпионате мира от количества забитых мячей/количество матчей. Можно заметить, что в среднем шансы команды в значительной степени по-

вышаются, когда команда больше забивает, и если команда забивает 4 мяча за матч в среднем, то это практически гарантирует ей победу на турнире.



**Рисунок 4. Зависимость шансов сборных на победу на чемпионате мира от количества забитых мячей/количество матчей**

На рисунке 5 изображен график шансов сборных на победу на чемпионате мира от ожидаемых очков в группе. Как видно из графика, график хорошо отражает вероятность победы команд, хорошо выступавших на групповом этапе, чем больше очков команда набирает в группе, тем выше её шансы победить на чемпионате мира по футболу.



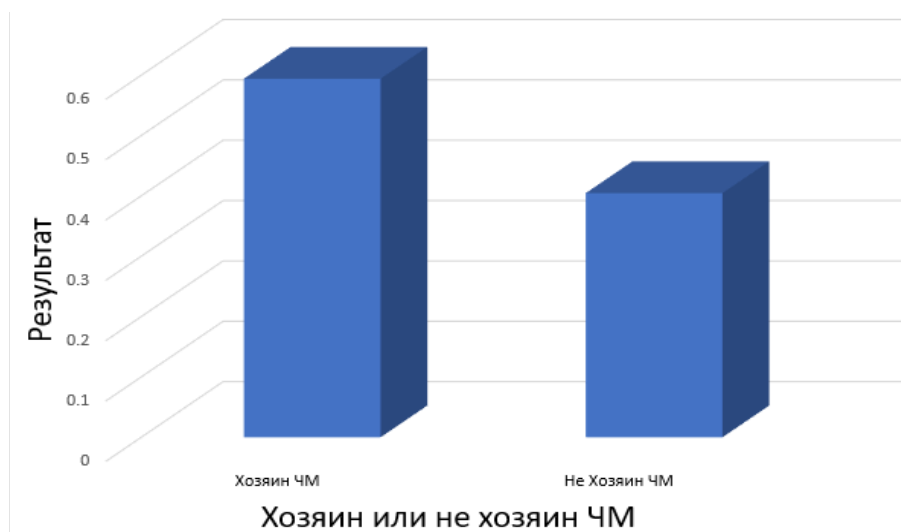
**Рисунок 5. Зависимость шансов сборных на победу на чемпионате мира от ожидаемых очков в группе**

На рисунке 6 продемонстрирована зависимость шансов сборных на победу на чемпионате мира от попадания или непопадания команды в первую корзину. Как видно из графика, при попадании команды в первую корзину её шансы добиться больших высот вырастают, по сравнению с командами в неё не попавшими.



**Рисунок 6. Зависимость шансов сборных на победу на чемпионате мира от попадания или непопадания команды в первую корзину**

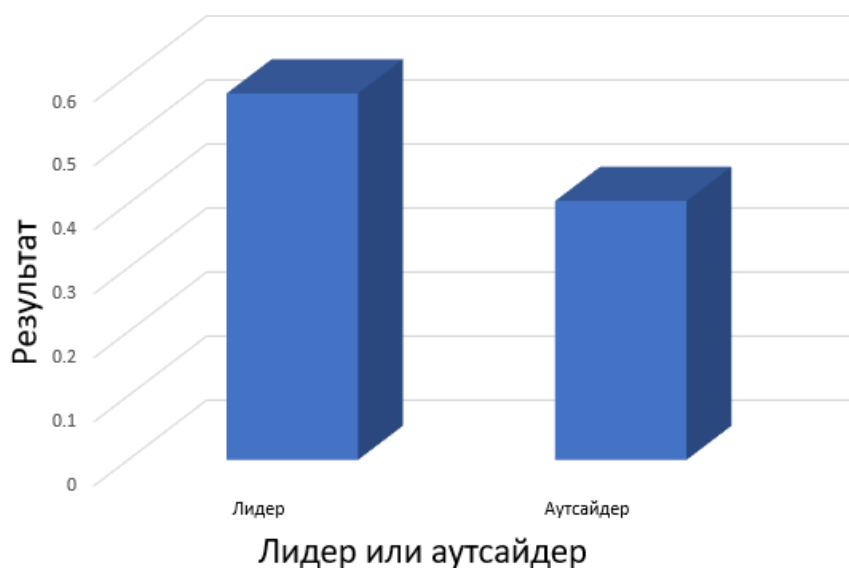
На рисунке 7 продемонстрирована зависимость шансов сборных на победу на чемпионате мира от того, является ли сборная хозяином ЧМ или нет. Этот пример свидетельствует, что в домашних условиях сборные лучше выступают на международных турнирах.



**Рисунок 7. Зависимость шансов сборных на победу на чемпионате мира от того, является ли сборная хозяином ЧМ или нет**

На рисунке 8 продемонстрирована зависимость шансов сборных на победу на чемпионате мира от того является ли команда лидером или

аутсайдером. Из результатов можно сделать вывод, что лидеры лучше выступают на чемпионате мира по футболу.



**Рисунок 8. Зависимость шансов сборных на победу на чемпионате мира от того является ли команда лидером или аутсайдером**

Полученные результаты исследований не противоречат реальности, что подтверждает, что спроектированную нейронную сеть можно считать пригодной для прогнозирования шансов сборных на победу на чемпионате мира по футболу.

**Заключение.** Построена система нейросетевого прогнозирования

Шансов сборной на победу на чемпионате мира по футболу. Спроектированная нейросетевая модель учитывает 12 параметров: попадание сборной в первую корзину (0 или 1), количество забитых мячей/количество матчей, количество пропущенных мячей/количество матчей, наличие лучшего игрока в мире (0 или 1), место на предыдущем чемпионате Европы/кубке Америки по футболу, место на предыдущем чемпионате мира по футболу, количество лидеров сборной (8) играет в ТОП-5 лигах в лучших клубах (0, 0.5 или 1), ожидаемый прогноз по очкам в группе, является ли сборная хозяйкой чемпионата мира (0 или 1), является ли сборная лидером или аутсайдером (0 или 1), средний возраст состав сборной, свой тренер или чужой.

Методом сценарного прогнозирования построены графики зависимостей прогнозируемых шансов победителя чемпионата мира от изменения входных параметров. Применение такого набора параметров в модели позволяет с высокой точностью прогнозировать победителя будущего чемпионата мира по футболу. Данный набор параметров может быть изменен для прогнозирования шансов клубов в других футбольных соревнованиях.

Также параллельно была построена данная модель на Python. В ходе выполнения работы была сформирована модель, разделена на составляющие, на основе которых проектировалась модель, которая затем была обучена и проведена тестовая проверка, после чего было проверено влияние отдельных показателей, их значимость на работу модели. В ходе опытов изменялись по-



казатели, влияющие на качество работы модели, однако существенного различия в результатах в лучшую сторону они не показатели. Результат модели на Python немного менее точен и не достиг подобного результата, который показал нейросимулятор.

В результате проделанной работе можно сделать вывод, что использование данного способа можно сравнить с использованием нейросимулятора, за исключением того, что нейросимулятор интуитивно проще и не требует сложного и глубокого погружения в материал.

### Библиографический список

1. Ясницкий Л.Н., Павлов И.В., Черепанов Ф.М. Прогнозирование результатов олимпийских игр 2014 года в неофициальном командном зачете методами искусственного интеллекта // Современные проблемы науки и образования. 2013. № 6. С. 985.
2. Ясницкий Л.Н., Vnukova O.V. Прогноз результатов олимпиады-2014 в мужском одиночном фигурном катании методами искусственного интеллекта // Современные проблемы науки и образования. 2014. № 1. С. 189.
3. Ясницкий Л.Н., Кировоса А.В., Ратегова А.В., Черепанов Ф.М. Прогноз результатов чемпионата мира-2015 по легкой атлетике методами искусственного интеллекта // Современные проблемы науки и образования. 2014. № 4. С. 624.
4. BBC News – Кто должен выиграть чемпионат мира? Отвечает статистика. [Электронный ресурс]. Режим доступа: <https://www.bbc.com/russian/features-44468736/>
5. Statista – Predicting the Unpredictable. [Электронный ресурс]. Режим доступа: <https://www.statista.com/chart/14251/world-cup-predictions/>
6. Gracernote.com – FIFA World Cup Predictions 2018. [Электронный ресурс]. Режим доступа: <https://www.gracernote.com/sports/fifa-world-cup-predictions-2018/>
7. Черепанов Ф.М., Ясницкий Л.Н. Нейросимулятор 4.0. Свидетельство о регистрации программы для ЭВМ RUS 2014612546. Заявка № 2014610341 от 15.01.2014.
8. Ясницкий Л.Н. Интеллектуальные системы: учебник. М.: Лаборатория знаний, 2016. 221 с.
9. Statista – The Quarter Final Winners (According to Google). [Электронный ресурс]. Режим доступа: <https://www.statista.com/chart/14550/world-cup-winners-according-to-google/>
10. Towards Data Science – Simulating the FIFA World Cup 2022. [Электронный ресурс]. Режим доступа: <https://towardsdatascience.com/simulating-the-fifa-world-cup-2022-d363fad7da22/>

# NEURAL NETWORK SYSTEM FOR PREDICTING CHANCE OF WORLD CUP WINNERS

*Pavlov Denis A.*

National Research University “Higher School of Economics”  
Str. Gagarin Boulevard, 37a, Perm, Russia, 614990, deniis.pavlov@mail.ru

This article describes the development of a neural network system for predicting the chances of national teams to win the World Cup. The system allows predicting the chances of national teams to win the World Championships with great accuracy, based on the previous performances of teams at previous world championships. With the help of the developed intelligent system the study of the subject area has been carried out and the regularities which have practical importance have been revealed.

**Keywords:** artificial intelligence, neural network technology, forecasting, football, FIFA World Cup.

# **ИНСТРУМЕНТЫ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА И МАТЕМАТИКА**

## ПРОГНОЗ НЕОБХОДИМОГО ЧИСЛА БИНАРНЫХ НЕЙРОНОВ ПРИ ПОВЕРКЕ ГИПОТЕЗЫ НЕЗАВИСИМОСТИ ДАННЫХ

*Золотарева Татьяна Александровна, Иванов Александр Иванович*  
Липецкий государственный педагогический университет  
имени П.П. Семенова-Тян-Шанского,  
398020, Россия, г. Липецк, улица Ленина, 42,  
zolotarevatatyana2016@yandex.ru

Оценивается приблизительное число статистических критериев проверки гипотезы независимости двух малых выборок в 16 опытов для принятия решения. Показано, что для обобщения нескольких статистических критериев, каждый из критериев следует представить эквивалентным искусственным нейроном.

**Ключевые слова:** статистический критерий, малые выборки, проверка гипотезы независимости данных.

**Введение.** В различных сферах деятельности сложно получить большие выборки реальных данных. При исследованиях приходится ориентироваться на малые выборки. Точно вычислить коэффициент корреляции по классической формуле на малых выборках не получается. Так при вычислении по классической формуле:

$$r(x, y) = \frac{1}{16} \sum_{i=1}^{16} \frac{(E(x) - x_i) \cdot (E(y) - y_i)}{\sigma(x) \cdot \sigma(y)}, \quad (1)$$

независимые выборки могут давать ошибки в интервале от -0.75 до +0.75.

В биометрии остро стоит вопрос о быстром автоматическом обучении больших сетей искусственных нейронов на малых выборки, например, объемом в 16 примеров образа «Свой». В свою очередь персептроны уступают по эффективности обогащения данных квадратичным нейронам [1, 2]. Ожидается, что замена персептронов квадратичными нейронами Махалонобиса позволит снизить число входов у одного нейрона с 16 до 4 при одновременной замене выходных бинарных квантователей на квантователи с тремя или четырьмя выходными состояниями.

Проблема перехода от нейронов с линейным накоплением данных к нейронам с накоплением данных в квадратичном пространстве состоит от плохо обусловленной задачи обращения корреляционных матриц при обучении классических нейронов Махалонобиса. Если обучение квадратичных нейронов выполнять подбором независимых входных биометрических

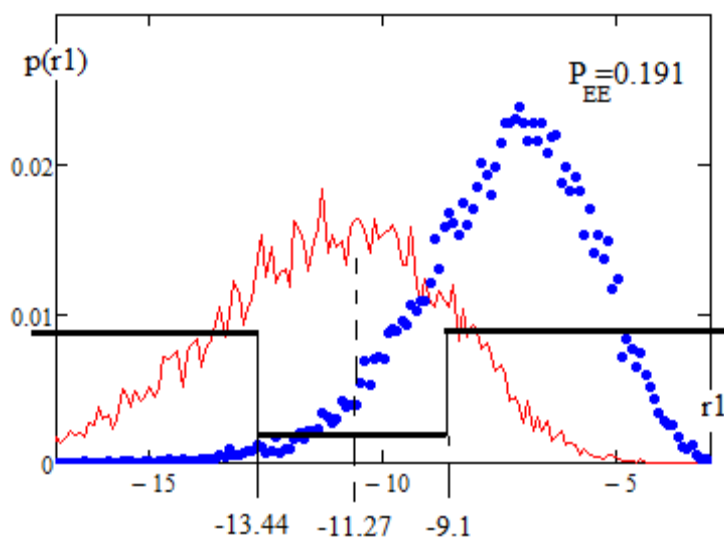
параметров, то по своей эффективности нейроны Махаланобиса и нейроны Евклида оказываются тождественны.

Таким образом, проблема быстрого автоматического обучения квадратичных нейронов сводится к регуляризации вычислений при классической оценке коэффициентов корреляции (1). Следует отметить, что в прошлом веке усилиями представителей математической статистики [3] было создано порядка 60-ти критериев, для проверки гипотезы независимости. Очевидно, что известные статистические критерии можно использовать совместно [4, 5], представив каждый из известных статистических критериев в форме эквивалентного искусственного нейрона. Тогда сеть из 60-ти искусственных нейронов на анализируемые данные малой выборки будет откликаться 60-ти битным выходным избыточным кодом. Свертывание кодовой избыточности позволяет обнаружить и устранить противоречия в частных решениях каждого из искусственных нейронов.

Следует так же подчеркнуть, что классическая формула Эджуорта-Эдлтона-Пирсона (1) является одним из самых мощных статистических критериев. В связи с этим имеет смысл исследовать близкие к этой формуле аналоги [6, 7]. Для примера рассмотрим критерий, вычисляемый по следующей формуле:

$$r1(x, y) = \frac{1}{16} \sum_{i=1}^{16} \frac{(x_i - \max(x)) \cdot (y_i - \min(y))}{|x_i - \max(x)| + |y_i - \min(y)|} \quad (2)$$

В этой формуле в числителе математические ожидания входных данных заменены на их экстремумы. В знаменателе в место стандартных отклонений используются их оценки. Результаты численного эксперимента по использованию (2) как нового статистического критерия отражены на рисунке 1.



**Рисунок 1. Высокая линейная разделимость малых независимых выборок и сильно коррелированных данных  $r=0.5$**

Из рисунка 1 видно, что у нового статистического критерия вероятность ошибок первого и второго рода составляет 0.191, что примерно на 33% хуже, чем у классического критерия (1).

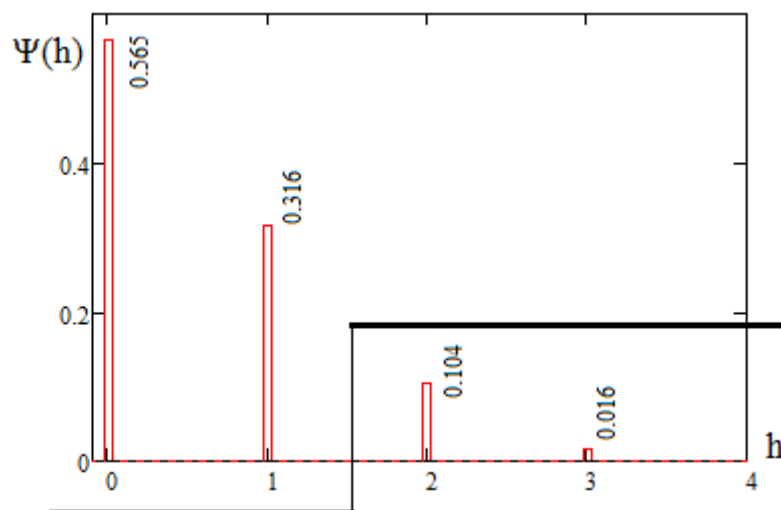
Очевидно, что по аналогии с первым критерием (2) мы можем построить второй такой же критерий:

$$r_2(x, y) = \frac{1}{16} \sum_{i=1}^{16} \frac{(x_i - \min(x)) \cdot (y_i - \max(y))}{|x_i - \min(x)| + |y_i - \max(y)|} \quad (3)$$

Распределение его выходных состояний полностью повторяет распределения, представленные на рисунке 1.

Очевидно, что каждый из перечисленных выше трех статистических критериев можно представить в форме трех искусственных нейронов, если после их сумматоров подключить бинарные выходные квантователи. В итоге мы получаем сеть из трех искусственных нейронов, на ее выходе мы имеем трех битный код. При обнаружении тремя нейронами выполнения гипотезы независимости мы будем наблюдать код «000». Если только один из нейронов дает состояние «1», то такое состояние также следует рассматривать как подтверждение гипотезы независимости. Все кодовые состояния с двумя единицами в разрядах и кодовое состояние «111» будем рассматривать как подтверждение зависимости исследуемых данных.

Формально подсчет единиц в коде является не чем иным как вычисление расстояния Хэмминга между идеальным кодом «000» и анализируемым кодом. На рисунке 2 приведены значения амплитуд вероятности четырех линий спектра расстояний Хэмминга.



**Рисунок 2. Дискретный спектр четырех линий расстояний Хэмминга на фоне порогового принятия решений**

Как видно из рисунка 2 использование трех критериев позволяет снизить вероятности ошибок первого и второго рода до величины 0.12 по сравнению с величиной 0.144 у классического критерия. Мы наблюдаем снижение вероятности ошибок в 1.2 раза. Это эквивалентно увеличению объема анализируемой выборки примерно на 50%. То есть при использовании для вычислений только классической формулы (1) эквивалентной точности удалось бы добиться при увеличении выборки с 16 опытов до 24 опытов.

**Заключение.** Для достижения доверительной вероятности 0.99 требуется около 600 бинарных нейронов, которые будут объединены в нейросеть. Существенно более эффективными являются искусственные квадратичные нейроны с многоуровневыми выходными квантователями [8].

### Библиографический список

1. Сулавко А. Е., Ложников П. С., Иванов А. И., Золотарева Т. А. Оценка информативности биометрических данных динамики рукописного почерка при обогащении искусственными нейронами в линейном пространстве и корреляционными нейронами в квадратичном пространстве //Безопасность информационных технологий : сб. науч. ст. по материалам III Всерос. науч.-техн. конф. : в 2 т. Пенза : Изд-во ПГУ, 2021. Т. 1. С. 47–54.

2. Zolotareva, T.A. Regularization of automatic training of Mahalanobis neurons for small samples of examples of the “Own” image/ T.A. Zolotareva, A.I. Ivanov, O.V. Selishchev, D.M. Skudnev // E3S Web of Conferences 224, 01024 (2020) ТРАСЕЕ-2020.

3. Кобзарь А.И. Прикладная математическая статистика. Для инженеров и научных работников. М.: ФИЗМАТЛИТ, 2006 г., 816 с.

4. Иванов А.И., Золотарева Т.А. Искусственный интеллект в защищенном исполнении: синтез статистико-нейросетевых автоматов многокритериальной проверки гипотезы независимости малых выборок биометрических данных. Препринт. //Изд-во «Пензенского государственного университета», Пенза, 2020, 105 с.

5. Иванов А. И., Серикова Ю. И., Золотарева Т. А., Полковникова С. А. Многокритериальная нейросетевая оценка коэффициентов корреляции для обработки малых выборок биометрических данных // Известия высших учебных заведений. Поволжский регион. Технические науки. 2021. № 1. С. 13–22.

6. Золотарева Т.А. Аналогии корреляционного критерия Пирсона / Информатика и вычислительная техника и управление. Серия. Естественные и технические науки. №4, 2021, с.90-97.

7. V. I. Volchikhin, A. I. Ivanov, T. A. Zolotareva and D. M. Skudnev Synthesis of four new neuro-statistical tests for testing the hypothesis of independence of small samples of biometric data // Journal of Physics: Conference Series 2094 (2021) 032013.

8. Малыгина, Е. А. Биометрико-нейросетевая аутентификация: перспективы применения сетей квадратичных нейронов с многоуровневым квантованием биометрических данных : препринт / Е. А. Малыгина. – Пенза: Изд-во ПГУ, 2020. – 114 с.

# PREDICTION OF THE REQUIRED NUMBER OF BINARY NEURONS WHEN VERIFICATION OF THE HYPOTHESIS OF DATA INDEPENDENCE

*Zolotareva Tatyana A., Ivanov Alexander I.*

Lipetsk State Pedagogical University  
named after P.P. Semenov-Tyan-Shansky,  
398020, Russia, Lipetsk, Lenin street, 42, zolotarevatatyana2016@yandex.ru

The approximate number of statistical criteria for testing the hypothesis of independence of two small samples in 16 experiments is estimated for decision making. It is shown that in order to generalize several statistical criteria, each of the criteria should be represented by an equivalent artificial neuron.

**Key words:** statistical test, small samples, data independence hypothesis testing.

УДК 004.032.26

## РОЛЬ СИНТЕТИЧЕСКИХ ДАННЫХ В РАЗВИТИИ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

*Рабчевский Андрей Николаевич*

Пермский государственный национальный исследовательский университет,  
614990, Россия, г. Пермь, ул. Букирева, 15, ran@psu.ru

В статье представлено описание синтетических данных, области их применения, методов генерации, систем генерации синтетических данных, оценки их качества. Также показана роль синтетических данных в развитии современных технологий искусственного интеллекта.

**Ключевые слова:** синтетические данные, искусственный интеллект, нейросетевые технологии.

**Введение.** В процессе проектирования современных нейросетевых моделей исследователи часто сталкиваются с проблемой недоступности достаточного количества данных для их обучения, а также с неравномерностью или разреженностью этих данных. Нередко случается, что реальных данных просто не существуют, так как область исследований еще только формируется. Также существует проблема конфиденциальности реальных персональных данных или медицинских данных пациентов, которые используются в процессе обмена между исследователями или в процессе тестирования различных нейросетевых систем. Во всех этих случаях на помощь приходят синтетические данные.

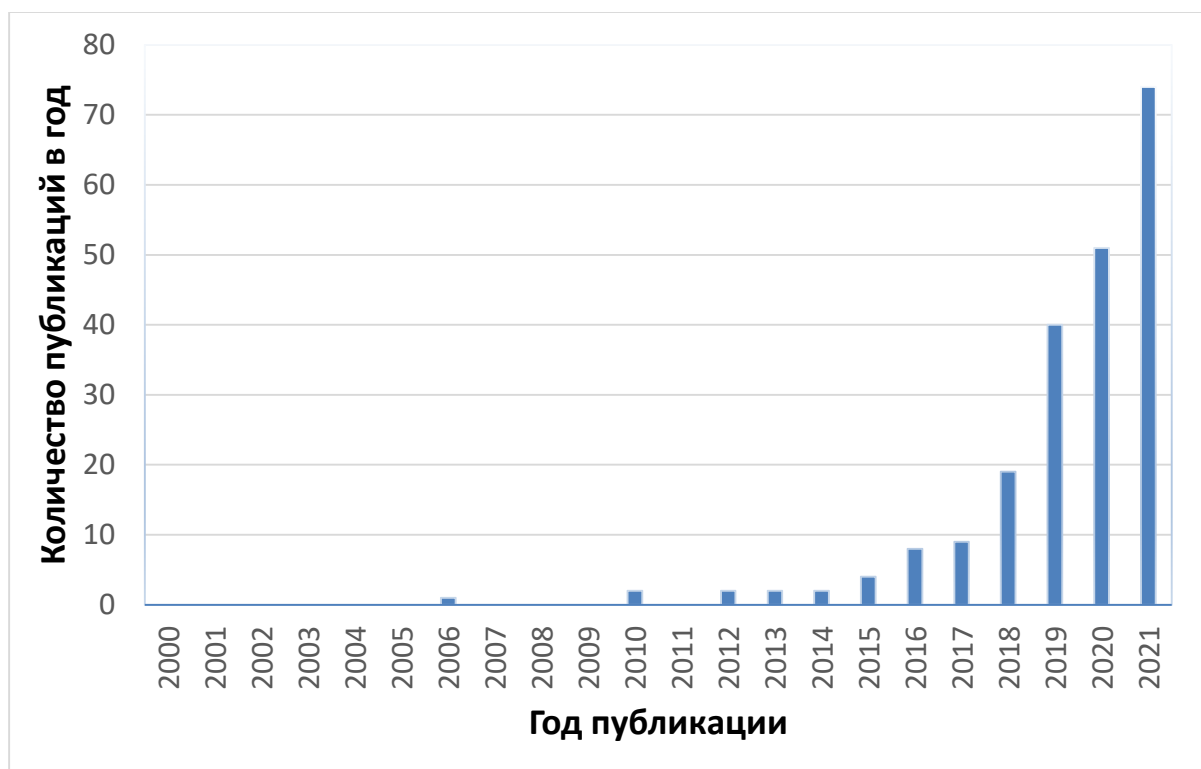


Как следует из названия, синтетические данные, это данные, которые созданы искусственно, а не в результате реальных событий. Они часто создаются с помощью алгоритмов и используются для широкого спектра действий.

Одно из первых упоминаний о применении синтетических данных встречается в связи с разработкой и тестированием системы обнаружения вторжений DARPA в 1998 и 1999 годах [1]. Тестовые данные содержали сетевой трафик и файлы журнала системных вызовов из смоделированной большой компьютерной сети. Атакующие данные были сгенерированы синтетически на основе сценариев возможных атак, а фоновые данные – с помощью программных автоматов, имитирующих использование различных услуг. Использование синтетических данных позволило разработчикам смоделировать и протестировать различные сценарии вторжения, которые ранее еще не встречались.

В тех случаях, когда новые службы еще только тестируются перед вводом в эксплуатацию, данные для обучения нейросетей могут просто отсутствовать. В этом случае для тестирования нужны синтезированные данные. Например, авторы статьи [2] уже в 2002 применяли синтетические данные при создании системы детектирования мошенничества.

С тех пор использование синтетических данных стало применяться все чаще. На рисунке 1 показан график роста количества ежегодных публикаций, содержащих в названии «синтетические данные», по данным электронной библиотеки arxiv.org<sup>1</sup> за период 2000-2021 годы.



**Рисунок 1. Ежегодное количество публикаций об использовании синтетических данных**

<sup>1</sup> Электронная библиотека Cornell University доступна по адресу <https://arxiv.org>

С января по август 2022 года в этой библиотеке уже опубликовано 85 статей и их количество продолжает расти. Также продолжают расширяться и сферы применения синтетических данных.

**Области применения синтетических данных.** Наибольшее количество публикаций связано с распознаванием образов и компьютерным зрением. Это связано с тем, что доминирующий метод сбора данных в распознавании образов основан на веб-данных и ручных аннотациях, а для таких проблем компьютерного зрения, как стерео или оценка оптического потока, этот подход неосуществим, потому что люди не могут вручную ввести поле потока с точностью до пикселя. Отсюда и всплеск числа подходов, полагающихся исключительно на синтетические данные для обучения. В статье [3] авторы предлагают использование синтетически сгенерированных данных с целью обучения глубоких сетей таким задачам, а также несколько способов генерации таких данных и оценки влияния свойств набора данных на производительность и свойства обобщения результирующих сетей.

Синтетические данные широко используются для распознавания и классификации различных объектов, таких как:

- продукты питания [4,5],
- бытовые отходы [6],
- объекты розничной торговли [7],
- промышленные детали [8],

а также в таких областях распознавания, как:

- распознавание лиц [9–15],
- распознавание сцен [16,17],
- распознавании позы, рук и жестов [18–20],
- анализ толпы и подсчет пешеходов [21–25]
- при разработке летательных аппаратов [26–29],
- в автономном вождении [30–35],
- в робототехнике [36–40]
- распознавание речи и звуков [41–46],
- распознавание текста [47–50] в различных естественных условиях.

Особое место синтетические данные занимают в медицине. По большей части это связано с конфиденциальностью медицинских данных пациентов. Большое количество работ посвящено использованию синтетических данных в медицинской документации [51] и применению реалистичных синтетических электронных медицинских карт [52–54]. Представлены также инструментальные методы генерации различных синтетических изображений:

- поражений печени [55],
- головного мозга [56,57],
- костных структур для грудных рентгенограмм [58],
- синтетических образцов компьютерной томографии грудной клетки [59] и др.

Также встречаются примеры использования синтетических данных, например, для развития глубокой моторики младенцев [60], для обнаружения и

коррекции физических упражнений [61], для решения проблемы асинхронии вентиляции легких [62], для обнаружения инфицированных коронавирусом участков легких на КТ-изображениях [63], классификации пневмонии [64] и т.д.

Синтетические данные применяются и в других областях науки и техники:

- социальные науки [65–68],
- экономика [69–72],
- лингвистика [73–79],
- биология и генетика [80–83],
- математика [84],
- астрономия [85],
- промышленность [86–88].

Методам и алгоритмам генерации синтетических данных также посвящено большое количество работ. Рост популярности и доступности синтетических данных обязан бурному развитию алгоритмов генеративных adversarial сетей (GAN), которые способны создавать множество синтетических примеров на основе небольшого количества реальных данных, а также развитию платформ CAD моделирования и систем виртуальной реальности.

Несмотря на стремительный рост количества примеров использования синтетических данных, в работах [89–92] и многих других большое внимание уделяется их качеству.

Рост популярности синтетических данных также объясняется тем, что стоимость их генерации существенно ниже, чем поиск и разметка реальных данных. Так, например, при создании датасета [93] для классификации ролей пользователей социальных сетей [94–96] трудоемкость получения синтетического датасета была примерно в 1000 раз ниже, чем поиск и разметка реальных данных.

Сегодня существует целая индустрия по производству синтетических данных, список некоторых представителей которой представлен здесь [97].

**Заключение.** Представленные данные убедительно показывают, что синтетические данные стали неотъемлемым элементом современных нейросетевых моделей и систем. Они используются во все большем количестве областей науки и техники, начиная от медицины и генетики, и заканчивая математикой и астрономией. С помощью синтетических данных улучшаются нейросетевые алгоритмы, дополняются недостающие данные, генерируются несуществующие данные, обеспечивается конфиденциальность, а также снижается стоимость построения нейросетевых систем. К сожалению, большая часть работ, представленных в данной статье, опубликованы в зарубежных изданиях, что говорит о том, что в России использование синтетических данных развито не так сильно. Однако, автор надеется, что данная работа поможет обратить внимание российских исследователей к этому эффективному инструменту, что поможет развитию искусственного интеллекта в России.

### Библиографический список

1. Lippmann R.P. et al. Evaluating intrusion detection systems: the 1998 DARPA off-line intrusion detection evaluation // Proceedings DARPA Information Survivability Conference and Exposition. DISCEX'00. IEEE Comput. Soc. P. 12–26.

2. Lundin E., Kvarnström H., Jonsson E. A Synthetic Fraud Data Generation Methodology. 2002. P. 265–277.
3. Mayer N. et al. What Makes Good Synthetic Training Data for Learning Disparity and Optical Flow Estimation? // *Int J Comput Vis*. 2018. Vol. 126, № 9. P. 942–960.
4. Rajpura P.S., Hegde R.S., Bojinov H. Object Detection Using Deep CNNs Trained on Synthetic Images // *CoRR*. 2017. Vol. abs/1706.06782.
5. Park D. et al. Deep Learning based Food Instance Segmentation using Synthetic Data // *CoRR*. 2021. Vol. abs/2107.07191.
6. Grard M. et al. Object segmentation in depth maps with one user click and a synthetically trained fully convolutional network // *CoRR*. 2018. Vol. abs/1801.01281.
7. Hinterstoisser S. et al. An Annotation Saved is an Annotation Earned: Using Fully Synthetic Training for Object Instance Detection // *CoRR*. 2019. Vol. abs/1902.09967.
8. Back S. et al. Segmenting Unseen Industrial Components in a Heavy Clutter Using RGB-D Fusion and Synthetic Data // *CoRR*. 2020. Vol. abs/2002.03501.
9. Queiroz R. et al. Generating Facial Ground Truth with Synthetic Faces // 2010 23rd SIBGRAPI Conference on Graphics, Patterns and Images. 2010. P. 25–31.
10. Kortylewski A. et al. Priming Deep Neural Networks with Synthetic Faces for Enhanced Performance // *CoRR*. 2018. Vol. abs/1811.08565.
11. Kortylewski A. et al. Analyzing and Reducing the Damage of Dataset Bias to Face Recognition With Synthetic Data // 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). 2019. P. 2261–2268.
12. Kortylewski A. et al. Can Synthetic Faces Undo the Damage of Dataset Bias to Face Recognition and Facial Landmark Detection? 2018.
13. Wood E. et al. Fake It Till You Make It: Face analysis in the wild using synthetic data alone // *CoRR*. 2021. Vol. abs/2109.15102.
14. Khan F. et al. High-Accuracy Facial Depth Models derived from 3D Synthetic Data. arXiv, 2020.
15. Qiu H. et al. SynFace: Face Recognition with Synthetic Data // *CoRR*. 2021. Vol. abs/2108.07960.
16. Handa A. et al. SceneNet: Understanding Real World Indoor Scenes With Synthetic Data. 2015.
17. Papon J., Schoeler M. Semantic Pose Using Deep Networks Trained on Synthetic RGB-D // 2015 IEEE International Conference on Computer Vision (ICCV). IEEE, 2015. P. 774–782.
18. Malik J. et al. DeepHPS: End-to-end Estimation of 3D Hand Pose and Shape by Learning from Synthetic Depth // 2018 International Conference on 3D Vision (3DV). 2018. P. 110–119.
19. Rahman M.M. et al. Effect of Kinematics and Fluency in Adversarial Synthetic Data Generation for ASL Recognition with RF Sensors // *IEEE Trans Aerosp Electron Syst*. Institute of Electrical and Electronics Engineers (IEEE), 2022. P. 1.
20. Rad M., Oberweger M., Lepetit V. Feature Mapping for Learning Fast and Accurate 3D Pose Inference from Synthetic Images // *CoRR*. 2017. Vol. abs/1712.03904.

21. Ekbatani H.K., Pujol O., Seguí S. Synthetic Data Generation for Deep Learning in Counting Pedestrians // ICPRAM. 2017.
22. Khadka A.R. et al. Learning how to analyse crowd behaviour using synthetic data // Proceedings of the 32nd International Conference on Computer Animation and Social Agents. New York, NY, USA: ACM, 2019. P. 11–14.
23. Wang Q. et al. Learning from Synthetic Data for Crowd Counting in the Wild // CoRR. 2019. Vol. abs/1903.03303.
24. Fabbri M. et al. MOTSynth: How Can Synthetic Data Help Pedestrian Detection and Tracking? // CoRR. 2021. Vol. abs/2108.09518.
25. Nabati M. et al. Using Synthetic Data to Enhance the Accuracy of Fingerprint-Based Localization: A Deep Learning Approach // IEEE Sens Lett. Institute of Electrical and Electronics Engineers (IEEE), 2020. Vol. 4, № 4. P. 1–4.
26. Madaan R., Maturana D., Scherer S. Wire detection using synthetic data and dilated convolutional networks for unmanned aerial vehicles // 2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). 2017. P. 3487–3494.
27. Marcu A. et al. SafeUAV: Learning to Estimate Depth and Safe Landing Areas for UAVs from Synthetic Data // ECCV Workshops. 2018.
28. Kiefer B., Ott D., Zell A. Leveraging Synthetic Data in Object Detection on Unmanned Aerial Vehicles // CoRR. 2021. Vol. abs/2112.12252.
29. Liu W., Liu J., Luo B. Can Synthetic Data Improve Object Detection Results for Remote Sensing Images? // CoRR. 2020. Vol. abs/2006.05015.
30. Moiseev B. et al. Evaluation of Traffic Sign Recognition Methods Trained on Synthetically Generated Data. 2013. P. 576–583.
31. Ros G. et al. The SYNTHIA Dataset: A Large Collection of Synthetic Images for Semantic Segmentation of Urban Scenes // 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2016. P. 3234–3243.
32. You Y. et al. Virtual to Real Reinforcement Learning for Autonomous Driving // CoRR. 2017. Vol. abs/1704.03952.
33. Devaranjan J., Kar A., Fidler S. Meta-Sim2: Unsupervised Learning of Scene Structure for Synthetic Data Generation // CoRR. 2020. Vol. abs/2008.09092.
34. Hahner M. et al. Semantic Understanding of Foggy Scenes with Purely Synthetic Data // CoRR. 2019. Vol. abs/1910.03997.
35. McCraith R., Neumann L., Vedaldi A. Real Time Monocular Vehicle Velocity Estimation using Synthetic Data // CoRR. 2021. Vol. abs/2109.07957.
36. Zhang Y. et al. UnrealStereo: A Synthetic Dataset for Analyzing Stereo Vision // CoRR. 2016. Vol. abs/1612.04647.
37. Triyonoputro J.C., Wan W., Harada K. Quickly Inserting Pegs into Uncertain Holes using Multi-view Images and Deep Network Trained on Synthetic Data // CoRR. 2019. Vol. abs/1902.09157.
38. Kollar T. et al. SimNet: Enabling Robust Unknown Object Manipulation from Pure Synthetic Data via Stereo // CoRR. 2021. Vol. abs/2106.16118.
39. Feng Q. et al. Look, Cast and Mold: Learning 3D Shape Manifold from Single-view Synthetic Data // CoRR. 2021. Vol. abs/2103.04789.

40. Lips T., de Gusseme V.-L., wyffels F. Learning Keypoints from Synthetic Data for Robotic Cloth Folding. arXiv, 2022.
41. Li J. et al. Training Neural Speech Recognition Systems with Synthetic Speech Augmentation // CoRR. 2018. Vol. abs/1811.00707.
42. Rosenblatt L. et al. Differentially Private Synthetic Data: Applied Evaluations and Enhancements // CoRR. 2020. Vol. abs/2011.05537.
43. Huang Y. et al. Learning generic feature representation with synthetic data for weakly-supervised sound event detection by inter-frame distance loss // CoRR. 2020. Vol. abs/2011.00695.
44. Hu T.-Y. et al. Synt++: Utilizing Imperfect Synthetic Data to Improve Speech Recognition. arXiv, 2021.
45. Fazel A. et al. SynthASR: Unlocking Synthetic Data for Speech Recognition // CoRR. 2021. Vol. abs/2106.07803.
46. Howard N. et al. A Neural Acoustic Echo Canceller Optimized Using An Automatic Speech Recognizer And Large Scale Synthetic Data. arXiv, 2021.
47. Jaderberg M. et al. Synthetic Data and Artificial Neural Networks for Natural Scene Text Recognition. 2014.
48. Gupta A., Vedaldi A., Zisserman A. Synthetic Data for Text Localisation in Natural Images // CoRR. 2016. Vol. abs/1604.06646.
49. Eggert C., Winschel A., Lienhart R. On the Benefit of Synthetic Data for Company Logo Detection // Proceedings of the 23rd ACM international conference on Multimedia. New York, NY, USA: ACM, 2015. P. 1283–1286.
50. Tang Z. et al. Stroke-Based Scene Text Erasing Using Synthetic Data // CoRR. 2021. Vol. abs/2104.11493.
51. Lombardo J., Moniz L. A Method for Generation and Distribution of Synthetic Medical Record Data for Evaluation of Disease-Monitoring Systems // Johns Hopkins APL Technical Digest (Applied Physics Laboratory). 2008. Vol. 27.
52. McLachlan S., Dube K., Gallagher T. Using the CareMap with Health Incidents Statistics for Generating the Realistic Synthetic Electronic Healthcare Record // 2016 IEEE International Conference on Healthcare Informatics (ICHI). IEEE, 2016. P. 439–448.
53. Guan J. et al. Generation of Synthetic Electronic Medical Record Text // 2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). 2018. P. 374–380.
54. Melamud O., Shivade C. Towards Automatic Generation of Shareable Synthetic Clinical Notes Using Neural Language Models // CoRR. 2019. Vol. abs/1905.07002.
55. Frid-Adar M. et al. GAN-based synthetic medical image augmentation for increased CNN performance in liver lesion classification // Neurocomputing. 2018. Vol. 321. P. 321–331.
56. Han C. et al. GAN-based synthetic brain MR image generation // 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018). 2018. P. 734–738.
57. Dikici E. et al. Constrained Generative Adversarial Network Ensembles for Sharable Synthetic Data Generation. arXiv, 2020.

58. Gozes O., Greenspan H. Bone Structures Extraction and Enhancement in Chest Radiographs via CNN Trained on Synthetic Data // CoRR. 2020. Vol. abs/2003.10839.
59. Das H.P. et al. Conditional Synthetic Data Generation for Robust Machine Learning Applications with Limited Pandemic Data // CoRR. 2021. Vol. abs/2109.06486.
60. Yang C.-Y. et al. Unsupervised Domain Adaptation Learning for Hierarchical Infant Pose Recognition with Synthetic Data. arXiv, 2022.
61. Duczak N., Kerzel M., Wermter S. Continual Learning from Synthetic Data for a Humanoid Exercise Robot // CoRR. 2021. Vol. abs/2102.10034.
62. van Diepen A. et al. A Model-Based Approach to Synthetic Data Set Generation for Patient-Ventilator Waveforms for Machine Learning and Educational Use // CoRR. 2021. Vol. abs/2103.15684.
63. Yazdekhasty P. et al. Segmentation of Lungs COVID Infected Regions by Attention Mechanism and Synthetic Data. arXiv, 2021.
64. Schaudt D. et al. Improving COVID-19 CXR Detection with Synthetic Data Augmentation. arXiv, 2021.
65. Ясницкий Л.Н., Бржевская А.С., Черепанов Ф.М. О возможностях применения методов искусственного интеллекта в сфере туризма // Сервис plus. 2010. № 4. С. 111-115.
66. Little C. et al. Generative Adversarial Networks for Synthetic Data Generation: A Comparative Study. arXiv, 2021.
67. Tiwald P., Ebert A., Soukup D.T. Representative & Fair Synthetic Data // CoRR. 2021. Vol. abs/2104.03007.
68. Hu J., Akande O., Wang Q. Multiple Imputation and Synthetic Data Generation with the R package NPBayesImputeCat. arXiv, 2020.
69. Heaton J.B., Witte J. Synthetic Financial Data: An Application to Regulatory Compliance for Broker-Dealers // Journal of Financial Transformation (Forthcoming). 2019.
70. Heaton J.B., Witte J. Machine-Learning-Generated Synthetic Data for Regulation Best Interest Compliance // SSRN Electronic Journal. 2019.
71. Heaton J.B., Witte J.H. Generating Synthetic Data to Test Financial Strategies and Investment Products for Regulatory Compliance // SSRN Electronic Journal. 2019.
72. Koenecke A., Varian H. Synthetic Data Generation for Economists. arXiv, 2020.
73. Cheng S. et al. AR: Auto-Repair the Synthetic Data for Neural Machine Translation // CoRR. 2020. Vol. abs/2004.02196.
74. Wolf F., Fink G.A. Annotation-free Learning of Deep Representations for Word Spotting using Synthetic Data and Self Labeling // CoRR. 2020. Vol. abs/2003.01989.
75. Puri R. et al. Training Question Answering Models From Synthetic Data // CoRR. 2020. Vol. abs/2002.09599.
76. Bogoychev N., Sennrich R. Domain, Translationese and Noise in Synthetic Data for Neural Machine Translation // CoRR. 2019. Vol. abs/1911.03362.

77. Zaratiana U. Contrastive String Representation Learning using Synthetic Data // CoRR. 2021. Vol. abs/2110.04217.
78. Nicosia M., Qu Z., Altun Y. Translate & Fill: Improving Zero-Shot Multilingual Semantic Parsing with Synthetic Data // CoRR. 2021. Vol. abs/2109.04319.
79. Stahlberg F., Kumar S. Synthetic Data Generation for Grammatical Error Correction with Tagged Corruption Models // CoRR. 2021. Vol. abs/2105.13318.
80. van den Bulcke T. et al. SynTReN: a generator of synthetic gene expression data for design and analysis of structure learning algorithms // BMC Bioinformatics. 2006. Vol. 7, № 1. P. 43.
81. Giuffrida M.V., Scharr H., Tsaftaris S.A. ARIGAN: Synthetic Arabidopsis Plants using Generative Adversarial Network // CoRR. 2017. Vol. abs/1709.00938.
82. Ward D., Moghadam P., Hudson N. Deep Leaf Segmentation Using Synthetic Data. 2018.
83. Fangbemi A.S. et al. ZooBuilder: 2D and 3D Pose Estimation for Quadrupeds Using Synthetic Data // CoRR. 2020. Vol. abs/2009.05389.
84. Firoiu V. et al. Training a First-Order Theorem Prover from Synthetic Data // CoRR. 2021. Vol. abs/2103.03798.
85. Suwa Y. et al. Observing Supernova Neutrino Light Curves with Super-Kamiokande. III. Extraction of Mass and Radius of Neutron Stars from Synthetic Data // Astrophys J. American Astronomical Society, 2022. Vol. 934, № 1. P. 15.
86. Duquesnoy M. et al. Machine Learning-Assisted Multi-Objective Optimization of Battery Manufacturing from Synthetic Data Generated by Physics-Based Simulations. arXiv, 2022.
87. Leonardi R. et al. Egocentric Human-Object Interaction Detection Exploiting Synthetic Data. arXiv, 2022.
88. Muench S. et al. Performance Assessment of different Machine Learning Algorithm for Life-Time Prediction of Solder Joints based on Synthetic Data. arXiv, 2022.
89. Jerome P. Reiter. Releasing multiply imputed, synthetic public use microdata: an illustration and empirical study // Royal Statistical Society. 2005. Vol. 168, № 1. P. 185–205.
90. McLachlan S. Realism in Synthetic Data Generation. Palmerston North, New Zealand: School of Engineering and Advanced Technology Massey University Palmerston North, New Zealand, 2017.
91. Nowruzi F.E. et al. How much real data do we actually need: Analyzing object detection performance using synthetic and real data // CoRR. 2019. Vol. abs/1907.07061.
92. Linjordet T., Balog K. Sanitizing Synthetic Training Data Generation for Question Answering over Knowledge Graphs // Proceedings of the 2020 ACM SIGIR on International Conference on Theory of Information Retrieval. New York, NY, USA: Association for Computing Machinery, 2020. P. 121–128.
93. Рабчевский А.Н., Заякин В.С. База данных для классификации ролей пользователей социальных сетей. Свидетельство государственной регистрации базы данных для ЭВМ №2021621533 от 15.07.2021. 2021.



94. Рабчевский А. Н. Нейросетевая система классификации пользователей социальных сетей и экспертный способ ее создания. . Москва: Нейрокомпьютеры и их применение. XVIII Всероссийская научная конференция. Тезисы докладов. , 2020. 298–299 p.

95. Рабчевский А. Н. Применение нейро-сетевой фильтрации для оптимизации алгоритмов выявления наиболее влиятельных узлов в социальных сетях // Нейрокомпьютеры и их применение. XIX Всероссийская научная конференция: тезисы докладов. Москва, 2021.

96. Рабчевский Е.А. et al. ЭКСПЕРТНЫЙ СПОСОБ ФОРМИРОВАНИЯ ОБУЧАЮЩИХ ВЫБОРОК НА ПРИМЕРЕ СОЗДАНИЯ НЕЙРОСЕТЕВОЙ СИСТЕМЫ КЛАССИФИКАЦИИ ПОЛЬЗОВАТЕЛЕЙ СОЦИАЛЬНЫХ СЕТЕЙ // НЕЙРОКОМПЬЮТЕРЫ: РАЗРАБОТКА, ПРИМЕНЕНИЕ. 2020. Vol. 22, № 5. P. 54–63.

97. Dilmegani Cem. The Ultimate Guide to Synthetic Data: Uses, Benefits & Tools [Electronic resource] // <https://research.aimultiple.com/synthetic-data/>. 2018

## **THE ROLE OF SYNTHETIC DATA IN THE DEVELOPMENT OF ARTIFICIAL INTELLIGENCE**

*Rabchevskiy Andrey N.*

Perm State University

Str. Bukireva, 15, Perm, Russia, 614990, ran@psu.ru

The article presents a description of synthetic data, the scope of their application, generation methods, systems for generating synthetic data, assessment of their quality. It also shows the role of synthetic data in the development of modern artificial intelligence technologies.

**Keywords:** synthetic data, artificial intelligence, neural network technology.

УДК 004.032.26

## **ИСПОЛЬЗОВАНИЕ DBLATEX В СОЧЕТАНИИ С ASCIIDOCTOR ДЛЯ ОФОРМЛЕНИЯ НАУЧНО-ИССЛЕДОВАТЕЛЬСКИХ РАБОТ СТУДЕНТОВ ПРОГРАММИСТСКИХ СПЕЦИАЛЬНОСТЕЙ В СООТВЕТСТВИИ С ДЕЙСТВУЮЩИМИ СТАНДАРТАМИ**

*Воробьев Лев О.*

Донецкий национальный технический университет (г. Донецк)

lev.vorobyov@rambler.ru

Использование dblatex в сочетании с asciidoctor для оформления научно-исследовательских работ студентов программистских специальностей. Проводится оценка форматов представления текстовых документов с

точки зрения наличия возможностей форматирования текста, удобства редактирования и наличия метаданных. Предлагается новый способ оформления учебных документов с помощью пакетов dblatex и asciidoctor.

**Ключевые слова:** семантическая разметка текста, принцип единого источника, текстовый процессор.

**Введение.** Оформление НИРС со вставками исходного кода представляет проблему для студентов, потому что каждый блок исходного кода нужно оформлять и подписывать единообразно, а это очень утомительно при большом количестве таких вставок. Кроме того, исходный код постоянно обновляется, что приводит к необходимости его вставки в отчет.

Для решения этой проблемы предлагается использование современных средств семантической разметки текста.

Под семантической разметкой текста понимается такой формат представления документов, где содержимое отделено от оформления.

Кроме того, в современных условиях необходимо импортозамещение зарубежного ПО, в частности текстового процессора MS Word.

Цель: на примере задачи оформления НИРС по программированию предложить наиболее эффективный метод подготовки документов.

Задачи:

- провести сравнительный анализ средств оформления текстовых документов, чтобы выделить наиболее подходящий инструмент для оформления НИРС по программированию с поддержкой вставок исходного кода, формул, автоматической нумерации рисунков, таблиц и пр.;
- упростить оформление НИРС по программированию в соответствии с требованиями ГОСТ 7.32-2017 путём отделения правил оформления от фактического текстового содержания, используя выделенные в предыдущем шаге инструменты.

**Исследование.** Существуют следующие форматы представления электронных текстовых документов: DocX, ODT, HTML, ePub, LaTeX, AsciiDoc, Markdown, FictionBook и простой текст.

Форматы DocX и ODT – это бинарные форматы представления электронных текстовых документов. Для их создания и редактирования используются пакеты MS Word или LibreOffice/OpenOffice Writer.

HTML — гипертекстовый формат для его просмотра используется браузер. Характеристики гипертекстовых файлов также относятся и к документам ePub, поскольку последние представляют собой сжатый набор гипертекстовых файлов.

AsciiDoc или Markdown – текстовые файлы, в которых определённые фрагменты текста выделяются символами форматирования. Для получения документа с визуальным форматированным текстом используются специальные утилиты [1].

LaTeX — формат документов для научной литературы, разработанный Дональдом Кнуттом в 1984 году. Файлы с расширением .tex представляют со-

бой текстовые файлы с командами форматирования. О написании этих команд рассказывается в книге [2].

FictionBook — это формат электронных книг, разработанный Дмитрием Грибовым в 2004 году. Файлы с расширением .fb2 представляют собой XML с текстом и метаданными [5]. А файлы с расширением .fb3 представляют собой сжатую папку с XML файлами, в одном из которых хранятся метаданные, а в другом — текст. Файлы .fb3 можно создать через онлайн-редактор <http://fb3edit.org>.

**Таблица 1**

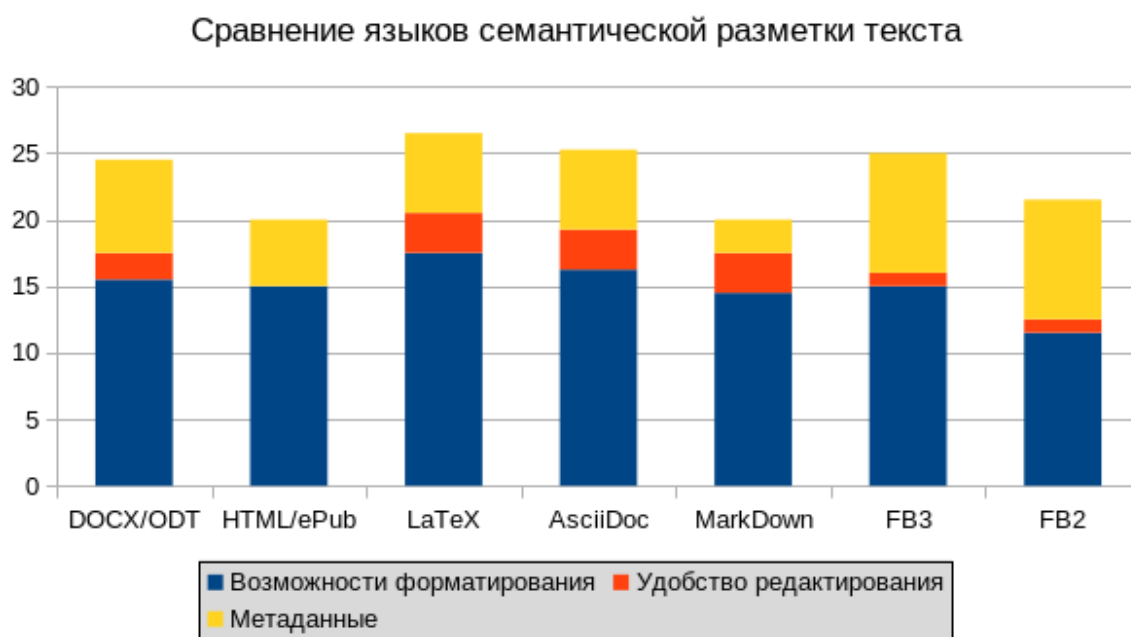
**Сравнительная таблица форматов подготовки документов**

Возможности семантической разметки текста	Формат документов							
	.docx	.html	.tex	.adoc	.md	.fb3	.fb2	.txt
<b>Возможности форматирования</b>								
1. Заголовки 1–6 уровней	1	1	1	1	1	1	1	0
2. Вставка изображений	1	1	1	1	1	1	1	0
3. Таблицы	1	1	1	1	1	1	0,5	0
4. Колонки	1	0,5	1	0,25	0	0	0	0
5. Многоуровневые списки	1	1	1	1	1	1	0	0
6. Выделение курсивом	1	1	1	1	1	1	1	0
7. Выделение жирным	1	1	1	1	1	1	1	0
8. Подчёркивание	1	1	1	0,5	0	1	0	0
9. Зачёркивание	1	1	1	0,5	1	1	1	0
10. Верхний индекс	1	1	1	1	1	1	1	0
11. Нижний индекс	1	1	1	1	1	1	1	0
12. Формулы	1	0	1	1	1	0	0	0
13. Гиперссылки	1	1	1	1	1	1	1	0
14. Стихи	0,5	0,5	0,5	1	0,5	1	1	0
15. Цитаты	0	1	1	1	1	1	1	0
16. Эпиграф	0,5	0,5	1	1	0,5	1	1	0
17. Горизонтальная линия	0,5	1	1	1	1	0	0	0
18. Разрыв страницы	1	0,5	1	1	0,5	1	0	0
Итого:	15,5	15	17,5	16,25	14,5	15	11,5	0
<b>Удобство редактирования</b>								
1. Сноски	1	0	1	1	1	1	1	0
2. Библиография	1	0	1	1	1	0	0	0
3. Встраивание исходных кодов в документ	0	0	1	1	1	0	0	0
Итого:	2	0	3	3	3	1	1	0
<b>Метаданные</b>								
1. Название документа	1	1	1	1	1	1	1	0
2. Автор	1	1	1	1	0,5	1	1	0
3. Язык	0,5	1	0,5	1	0	1	1	0
4. Дата создания (публикации)	1	0	1	1	0,5	1	1	0
5. Переводчик	0,5	0	0	0	0	1	1	0
6. Язык оригинала	0,5	0	0	0	0	1	1	0
7. Аннотация	0,5	1	1	1	0,5	1	1	0
8. Ключевые слова	1	1	1	1	0	1	1	0
9. Организация (издательство)	1	0	0,5	0	0	1	1	0
Итого:	7	5	6	6	2,5	9	9	0
Всего:	24,5	20	26,5	25,25	20	25	21,5	0
Ранг	4	6	1	2	6	3	5	

Разделим требования к средствам подготовки документов на три категории: наличие возможностей форматирования, удобства редактирования и наличия метаданных. Результаты оценочного сравнения приведены в таблице 1.

Здесь критериями оценки в ячейках таблицы являются: 1 — возможность присутствует; 0,5 — возможность есть, но либо она работает не всегда, либо её использование неочевидно и/или требует значительных усилий; 0,25 — возможность работает не всегда и её использование неочевидно и/или требует больших усилий; 0 — возможность отсутствует.

На рисунке 1 изображена накопительная диаграмма, в которой синий столбец показывает рейтинг по возможностям форматирования, синий с красным показывает, в том числе, и удобство редактирования, а все три цвета отражают суммарное количество баллов по трём критериям.



**Рисунок 1. Рейтинг форматов документов по критериям возможностей форматирования, удобства редактирования и наличия метаданных**

**Обсуждение результатов сравнения.** DocX не поддерживает вставку кода в документ с автоматическим обновлением. При изменении кода приходится заново вставлять код в документ. Т.е. нужно переписывать почти весь документ, чтобы распечатать новые изменения. Это послужило одной из причин отказа от Word и перехода на LaTeX (mikTeX + WinShell), где есть команда `input`, которая вставляет содержимое файла в документ. Но LaTeX отнимает много времени на написание команд, таких как `begin`, `end`, `textrt`, `verb` и т.п., поэтому пришлось перейти на Markdown, а затем на более удобный AsciiDoc с простыми командами.

Кроме того, в Word очень сложно редактировать большие документы, поскольку нужно долго листать документ, чтобы найти нужную страницу. В LaTeX, AsciiDoc это решается созданием отдельного файла для каждой главы и главного файла, который с помощью команды LaTeX `input` или AsciiDoc `include` включает все главы в один документ.

Когда нужно создавать несколько документов с одинаковым оформлением, в Word для этого используются шаблоны. Когда нужно редактировать несколько документов одновременно, созданных из одного шаблона все изменения в оформлении нужно повторять во всех документах и в шаблоне. В LaTeX или AsciiDoc настройки оформления можно поместить в один файл. И все изменения оформления будут распространяться на все документы, его использующие.

AsciiDoc файлы можно редактировать в блокноте и производить HTML или DocBook вызовом asciidoc из командной строки. Либо использовать плагины AsciiDoc для IntelliJ IDEA и при редактировании документа сразу получать результат на экране.

Сначала за неимением работающего с кириллицей конвертера приходилось переводить AsciiDoc в LaTeX с помощью скрипта на Perl. Но впоследствии обнаружилась программа dblatex, которая может перевести DocBook в LaTeX, и буквы не будут искажаться в выходном документе. Правда, dblatex использует bibtex, а нужен biblathex для генерации библиографических описаний в списке литературы на русском языке.

### **Метод подготовки документов в AsciiDoc в соответствии с ГОСТ 7.32**

Для подготовки документов с помощью AsciiDoc используем следующие инструменты: AsciiDoctor; Pandoc; Dblatex; TexLive; GNU make; AsciiDoctor-bibtex; Setzer; AsciiDoctor-fb2; AsciiDoctor-diagrams; PlantUML, TikZed.

Для настройки этих пактов необходимо поместить файлы из репозитория [3] в каталог .dblathex в домашней директории пользователя.

После установки и настройки вышеперечисленных пакетов создаём простой текстовый файл semantic.adoc и Makefile следующего содержания.

### **Пример Makefile для преобразования AsciiDoc в DocX и PDF**

```
TARGET = semantic.pdf
TEX = xetex
LATEX ?= xelatex
LATEXOPT = -shell-escape
ADOCOPTS = -a lang=ru -d article
DBSTYLE = gost-article
TMPDIR = texfiles
LATEXOPT += --output-directory=./${TMPDIR}
EVAL = $(shell sed -ne '/TEXINPUTS=./' p' ${TMPDIR}/env_tex)

all: ${TARGET}

${TMPDIR}/%.aux: %.tex
${EVAL} ${LATEX} ${LATEXOPT} ${OUTDIR} $<
```

```
%.pdf: %.tex ${TMPDIR}/%.aux
${EVAL} ${LATEX} ${LATEXOPT} ${OUTDIR} $<
mv ${TMPDIR}/$@ .
```

```
%.tex: %.adoc Makefile
asciidoc -b docbook5 -o - $< ${ADOCOPTS} | \
dblatex -ttex -T ${DBSTYLE} -d --tmpdir=./${TMPDIR} -o $@ -
```

```
%.docx: %.adoc Makefile ${DBSTYLE}.docx
asciidoc -b docbook5 -o - $< ${ADOCOPTS} | \
pandoc -f docbook -t docx -o $@ --reference-doc=${DBSTYLE}.docx
```

Для того, чтобы преобразовать в PDF, используем утилиту GNU make.

Об оформлении документов в формате AsciiDoc подробно написано в документации [6].

Для создания автоматически пронумерованного в порядке появления списка литературы сначала создаётся текстовый файл библиографической базы данных с помощью программы Setzer или Winshell, а затем подключается к файлу AsciiDoc с помощью расширения AsciiDoctor-bibtex. Для создания библиографической ссылки в текст вставляется команда cite.

Теперь фактическое текстовое содержимое размещается в файлах .adoc, а общие стили оформления в папке .dblatex, что реализует принцип семантической разметки текста.

Выше описан способ генерации PDF из текстового файла AsciiDoc. Для того, чтобы создать DocX нужно добавить файл DBSTYLE.docx, как описано в руководстве [7] и отредактировать в нём стили. Затем командой make <filename>.docx получить файл DocX со стилями из файла DBSTYLE.docx и текстовым содержимым из файла .adoc. Это реализует принцип единого источника, что может быть полезно при кросс-платформенной разработке.

Кроме того, следует отметить, что есть возможность преобразования в FictionBook для просмотра на мобильных устройствах, но при этом отсутствует возможность настраивать оформление.

### Существующие проблемы

На данный момент существуют следующие программные недочёты в предложенной схеме, которые можно устранить:

1. Модуль asciidoctor-biblatex добавляет лишние literal теги в выводе docbook;
2. Dblatex не переводит список источников в latex;
3. AsciiDoctor-fb2 добавляет лишние пустые строки в xml;
4. AsciiDoctor-fb2 не переводит блок abstract в аннотацию;
5. AsciiDoctor-fb2 не вставляет обязательный тег version в document-info блока описания;
6. AsciiDoctor-fb2 игнорирует параметр keywords в текстовом документе;

7. AsciiDoctor не использует теги poetry для оформления стихов в docbook;

8. Dblatex использует несовместимый с biblatex пакет natbib, который требует не работающий в последнее время bibtex 8-bit;

9. Setzer не открывает файл библиографической базы данных из проводника в xfce;

10. AsciiDoctor-diagrams не поддерживает кириллицу в диаграммах tikz.

Кроме вышеперечисленных проблем существуют ещё затруднения, связанные с самой схемой редактирования документов в формате AsciiDoc:

1. При создании таблиц необходимо вручную вводить ширину столбцов и смотреть на результирующий документ;

2. Слова в ячейках таблицы могут выходить за рамки ячеек;

3. Очень сложно отключить переносы слов;

4. Для редактирования графических изображений нужно использовать внешний редактор (например, TikZed) или вводить команды отрисовки графиков вручную;

5. Для того, чтобы выделить жирным или курсивом всю таблицу нужно выделять каждую ячейку отдельно.

Но среди минусов есть и плюсы использования AsciiDoc для форматирования научной документации:

1. Не нужно перенумеровывать ссылки на литературу при компоновке документа из нескольких статей, ссылки нумеруются автоматически;

2. Не нужно следить за расположением подписей таблиц и рисунков, все подписи располагаются единообразно;

3. Не нужно проверять нумерацию заголовков первого-шестого уровней, заголовки нумеруются автоматически;

4. Начертание шрифта меняется только путём вставки символов форматирования в текст документа, что позволит избежать нарушений правил оформления;

5. Документы в формате AsciiDoc лучше сжимаются и сохраняются в системе контроля версий, что экономит дисковое пространство;

6. Документ AsciiDoc можно разделить на несколько файлов и редактировать их отдельно, что облегчает навигацию по документу. При конвертации все фрагменты объединяются в один документ.

7. Созданный документ можно конвертировать в несколько разных форматов (принцип единого источника).

## Выводы

В результате сравнительного анализа было обнаружено, что по всем параметрам лидирует LaTeX, но LaTeX требует больших усилий для подготовки документов, поэтому более разумным компромиссом для подготовки документов со встроенными фрагментами программного кода является AsciiDoc. Если же программного кода в тексте документа нету, сам документ небольших размеров, с небольшим количеством иллюстраций и таблиц, то подойдёт MS Word или его аналоги.

Предложенный способ оформления НИРС в формате AsciiDoc позволяет студентом программистских специальностей добавить свой отчёт в систему контроля версий и отслеживать изменения вместе с кодом программы, легко помещать код программы в отчёт и не волноваться по поводу того, что код в отчёте может устареть, упрощает следование стандарту ГОСТ 7.32-2017 по части оформления текста поскольку избавляет от необходимости повторять одни и те же действия по созданию оформления в каждом новом документе. Однако у данного подхода существуют недостатки, которые приходится исправлять регулярными выражениями.

### Библиографический список

1. Портнов И. Подготовка технической документации с использованием asciidoc и DocBook [Электронный ресурс] — Режим доступа: <http://ipor-tnov.blogspot.com/2011/03/asciidoc-docbook.html>
2. Львовский С.М. Набор и вёрстка в системе LATEX, 3-е изд. — 2003 — 448 с.
3. DocBook-GOST-7.32 [Электронный ресурс] — Режим доступа: <https://github.com/lvorobyov/DocBook-GOST-7.32>
4. Pgf manual [Электронный ресурс] — Режим доступа: <https://mirror.datacenter.by/pub/mirrors/CTAN/graphics/pgf/base/doc/pgfmanual.pdf>
5. Грибов Д.П. FictionBook 3.0 beta – краткое описание [Электронный ресурс] — Режим доступа: [http://fictionbook.org/index.php/FictionBook\\_3.0\\_beta\\_-\\_краткое\\_описание](http://fictionbook.org/index.php/FictionBook_3.0_beta_-_краткое_описание)
6. Lars Vogel, Jennifer Nerlich *Using AsciiDoc and AsciiDoctor to write documentation – Tutorial* [Online] — URL: <https://www.vogella.com/tutorials/AsciiDoc/article.html>
7. Robin Moffatt *Converting from AsciiDoc to Google Docs and MS Word* [Online] — URL: <https://rmoff.net/2020/04/16/converting-from-asciidoc-to-google-docs-and-ms-word/>

## THE USE OF DBLATEX IN COMBINATION WITH ASCIIDOCTOR FOR THE DESIGN OF RESEARCH PAPERS OF STUDENTS OF PROGRAMMING SPECIALTIES

*Vorobyov Lev O.*

Donetsk National Technical University (Donetsk)

The formats of presentation of text documents are evaluated in terms of the availability of text formatting capabilities, ease of editing and the availability of metadata. A new way of registration of educational documents with the help of dblatex and ascii doctor packages is proposed.

**Keywords:** semantic markup of the text, the principle of a single source, a word processor.



## ТЕРМОДИНАМИЧЕСКИЕ СЛЕДСТВИЯ УРАВНЕНИЯ ВАН ДЕР ВААЛЬСА В ОБЛАСТИ КРИТИЧЕСКОГО СОСТОЯНИЯ

*Калиниченко Борис Артёмович*

Донецкий национальный технический университет,  
Донецкая Народная Республика, г. Донецк, ул. Артёма, 58, kreefore@mail.ru

На основе уравнения Ван дер Ваальса выполнен расчёт температурных зависимостей некоторых термодинамических свойств на линии сосуществующих фаз. Полученные теоретические результаты сопоставляются с экспериментальными данными.

**Ключевые слова:** уравнение состояния, термодинамические функции, фазовый переход, критическая точка, плотности сосуществующих фаз.

**Введение.** В 1873 году И.Д. Ван дер Ваальс, выступая в Лейденском университете с докладом о результатах своей научной работы, представил новое приближённое уравнение состояния, позволяющее не только качественно, но и, в некоторых случаях, количественно описывать равновесные термодинамические свойства простых жидкостей и газов в широком диапазоне изменения параметров состояния, включая критическую область [1].

Заметим, что во второй половине 19-го века в России исследования критического состояния жидкости проводились, в частности, А.Г. Столетовым и Д.И. Менделеевым, который 1861 году ввёл понятие температуры абсолютного кипения или, в современной терминологии, критической температуры жидкости [2]. Однако, побудительными мотивами для Ван дер Ваальса при создании теории конденсации были, по-видимому, результаты экспериментальной работы, выполненной Т. Эндрюсом в 1869 году.

**Теория и эксперимент.** Известно, что взаимодействие между частицами идеального газа отсутствует, а в реальных газах между ними действуют силы отталкивания и притяжения на малых и больших расстояниях соответственно. Эти факты послужили основанием для введения в уравнение состояния идеального газа специальных поправок, благодаря которым удалось воспроизвести экспериментально наблюдаемые явления. В результате было получено уравнение состояния реального газа

$$P = \frac{RT}{V-b} - \frac{a}{V^2}, \quad (1)$$

( $R$  – универсальная газовая постоянная) связывающее давление  $P$ , объём  $V$  и температуру  $T$  одного моля газа с параметрами  $a$  и  $b$ , характеризующими взаимодействие между частицами вещества. Первый из них —  $a$  отвечает силам

притяжения, приводящим к уменьшению давления, а второй —  $b$ , учитывает размер частиц, что можно интерпретировать как влияние сил отталкивания.

Один из вариантов нахождения значений параметров  $a$  и  $b$  состоит в решении системы уравнений

$$\left\{ \left( \frac{\partial P}{\partial V} \right)_c = 0, \left( \frac{\partial^2 P}{\partial V^2} \right)_c = 0 \right\}, \quad (2)$$

определяющей вместе с уравнением (1) координаты критической точки  $(P_c, V_c, T_c)$  на термодинамической поверхности  $P = P(V, T)$ . Здесь и далее индекс “с” указывает на принадлежность величин к критической точке.

Связь критических параметров вещества с его индивидуальными параметрами  $a$  и  $b$  устанавливается равенствами

$$P_c = \frac{a}{27b^2}, \quad V_c = 3b, \quad RT_c = \frac{8a}{27b}, \quad (3)$$

получаемыми при решении системы (2).

Из (3) и уравнения (1) следует, что критическая сжимаемость флюида Ван дер Ваальса не зависит от параметров  $a, b$  и равна

$$Z_c = \frac{P_c V_c}{RT_c} = \frac{3}{8} = 0,375.$$

Для реальных газов эта величина  $Z_c < 3/8$ , например, для четырёх инертных газов она принимает значения от 0,290 до 0,307.

С помощью (3), привлекая экспериментальные значения  $P_c$  и  $V_c$  для конкретного вещества, определяются значения  $a$  и  $b$  этого вещества, используемые, в частности, с целью представления уравнения (1) в безразмерной форме в приведённых переменных  $\Pi = P/P_c$ ,  $\varphi = V/V_c$ ,  $\tau = T/T_c$

$$\Pi(\varphi, \tau) = \frac{8\tau}{3\varphi - 1} - \frac{3}{\varphi^2}. \quad (4)$$

Дифференцируя уравнение (4) по  $\tau$ , найдём значение производной в критической точке

$$\left( \frac{\partial \Pi}{\partial \tau} \right)_c = \frac{8}{3\varphi - 1} \Big|_{\varphi=1} = 4,$$

несколько отличающееся от значения 6,0 для реальных газов [3].

Уравнение (4) не содержит величин  $a$  и  $b$ , характеризующих конкретное вещество, поэтому выражает закон соответственных состояний, который хорошо выполняется для группы инертных газов и поэтому является удобным математическим инструментом для расчётов их свойств.

К достоинствам уравнения Ван дер Вальса следует отнести возможность предсказания фазового перехода жидкость-газ с критической точкой, как указано выше, и установления границы области метастабильных состояний – спинодали, определяемой первым уравнением системы (2). Граница области двухфазных состояний может быть найдена с помощью правила равных площадей, предложенного Максвеллом и связанного с условиями термодинамического равновесия фаз, состоящего в равенстве давлений и химических потенциалов сосуществующих фаз:

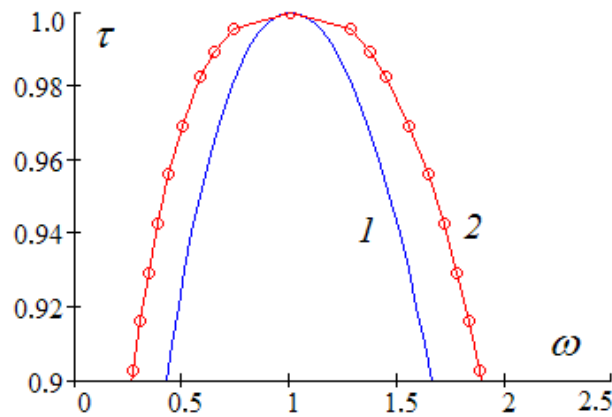
$$\begin{cases} \Pi(\varphi_1, \tau) = \Pi(\varphi_2, \tau), \\ \mu(\varphi_1, \tau) = \mu(\varphi_2, \tau). \end{cases} \quad (5)$$

где  $\mu(\varphi, \tau) = \mu_{id}(\varphi, \tau) - k_B T_c \left[ \tau \ln \left( 1 - \frac{1}{3\varphi} \right) - \frac{\tau}{3\varphi - 1} + \frac{9}{4\varphi} \right]$  – химический потенциал,

$k_B$  – постоянная Л.Больцмана,  $\mu_{id}(\varphi, \tau) = k_B T_c \tau \ln \left( \frac{n_c \lambda_c^3}{\varphi \tau^{3/2}} \right)$ ,  $\lambda_c = \frac{h}{\sqrt{2\pi m k_B T_c}}$ ,

$h$  – постоянная М. Планка.

Решение системы (5) при заданной температуре  $\tau$  позволяет найти плотности сосуществующих фаз  $\omega_1 = 1/\varphi_1$ ,  $\omega_2 = 1/\varphi_2$ , с помощью которых можно построить бинодаль (линию сосуществующих фаз) и выполнить расчёты свойств газовой и жидкой фазы на бинодали. На рисунке 1 изображены экспериментальная и теоретическая бинодали.



**Рисунок 1. Кривые сосуществующих фаз: 1 – уравнение (1); 2 – экспериментальные данные для аргона [4].**

Интегрируя по объёму  $V$  соотношение  $\left( \frac{\partial U}{\partial V} \right)_T = T \left( \frac{\partial P}{\partial T} \right)_V - P$  [5], правая часть которого для уравнения (1) равна  $T \left( \frac{\partial P}{\partial T} \right)_V - P = \frac{a}{V^2}$ , получим

$$U(T, V) = \int \frac{a}{V^2} dV = -\frac{a}{V} + U_0.$$

Постоянную интегрирования  $U_0$  найдём с учётом того, что при  $V \rightarrow \infty$  реальный газ становится идеальным, поэтому  $U_0 = U_{id} = C_V^{id}T$ ,  $C_V^{id} = 3R/2$  – изохорная теплоёмкость идеального газа. Тогда внутренняя энергия ван-дерваальсовского флюида будет равна

$$U(T, V) = -\frac{a}{V} + \frac{3}{2}RT. \quad (6)$$

Из (6) следует, что изохорная теплоёмкость газа Ван дер Ваальса не зависит от температуры

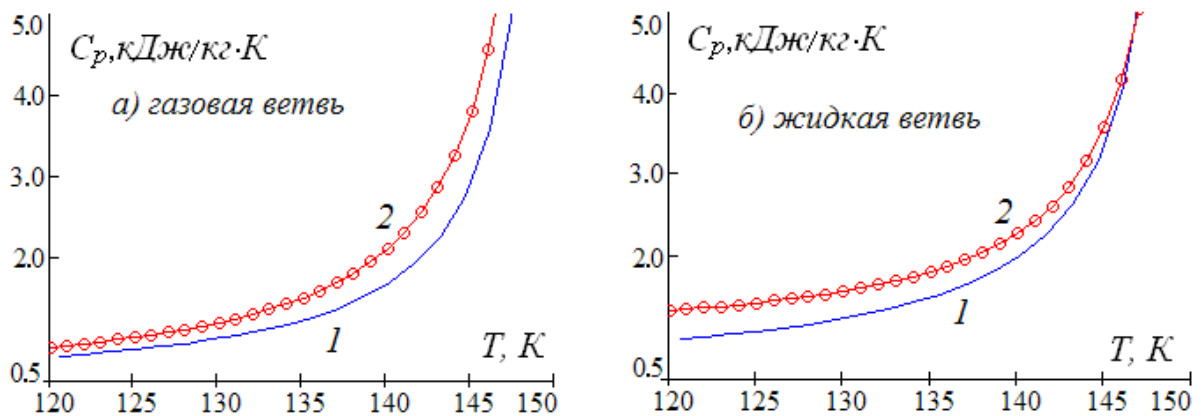
$$C_V = \left( \frac{\partial U}{\partial T} \right)_V = \frac{3R}{2} = C_V^{id}.$$

Однако, данные измерений показывают, что температурная зависимость существует, а на бинодали теплоёмкости  $C_V$  и жидкости, и газа возрастают с приближением температуры к критической  $T_c$ .

Изобарная теплоёмкость вычисляется по формуле

$$C_P(T) = C_V - \frac{(\partial P / \partial T)_V^2}{(\partial P / \partial V)_T} = C_V + R \left( 1 - \frac{9\omega}{4\tau} \left( 1 - \frac{\omega}{3} \right)^2 \right)^{-1}, \quad (7)$$

с помощью которой, учитывая плотности  $\omega_1, \omega_2$  жидкости и газа, построены зависимости  $C_P(T)$ , изображённые на рисунке 2.

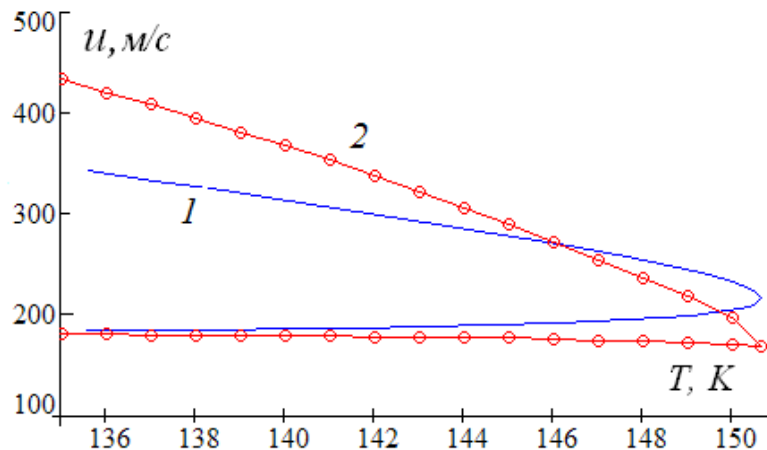


**Рисунок 2. Изобарная теплоёмкость: 1 – расчёт по уравнению (1); 2 – экспериментальные данные для аргона [4].**

На рисунке 3 представлены экспериментальные данные скорости звука  $u(T)$  в жидком и газообразном аргоне на линии сосуществующих фаз [4] и соответствующие значения  $u(T)$ , рассчитанные по следующей формуле с учётом (1) и решений системы (5)

$$u(T) = v \left( \frac{TM}{C_V} \left( \frac{\partial P}{\partial T} \right)_V^2 - M \left( \frac{\partial P}{\partial V} \right)_T \right)^{1/2}, \quad (8)$$

где  $v$  – удельный объём,  $M$  – молярная масса вещества,  $C_V$  – молярная теплоёмкость.



**Рисунок 3. Зависимость скорости звука на бинадали: 1 – расчёт по уравнению (1); 2 – экспериментальные данные для аргона [4].**

Из формулы (8) и рисунка 3 видно, что  $u(T_c) \neq 0$ , так как теплоёмкость флюида Ван дер Ваальса  $C_V \neq \infty$ , что характерно для многих уравнений состояния. Кроме того, наблюдается качественное отличие в поведении теоретической и экспериментальной кривых, отвечающих газообразному состоянию (нижние ветви кривых  $u(T)$  на рисунке 3).

Более успешным в количественном отношении уравнение Ван дер Ваальса оказалось при расчётах изобарной теплоёмкости, скорости звука, коэффициента Джоуля-Томсона и других термодинамических характеристик в накритической области, где  $P > P_c$  при температурах от 100 до 400 К.

**Заключение.** Уравнение Ван дер Ваальса имеет важное значение при изучении некоторых курсов физики, связанных с конденсированным состоянием вещества, т.к. оно может выступать в качестве удобного объекта для демонстрации возможностей аналитических методов при исследовании термодинамики различных модельных систем. Несмотря на свой солидный возраст (без малого 150 лет), уравнение не утратило своей актуальности, поскольку до сих пор является источником “вдохновения” для специалистов-теплофизиков, конструирующих на основе канонической формы различные модификации, позволяющие улучшить теоретическое описание экспериментальной картины [6-9].

Результаты научной деятельности Ван дер Ваальса были рассмотрены Нобелевским комитетом, в результате чего в 1910 году учёный был удостоен Нобелевской премии по физике за работы по созданию уравнений агрегатных состояний газов и жидкостей

*Автор считает приятным долгом выразить свою благодарность И.К. Локтионову за интересное обсуждение, ценные советы и замечания.*

## Библиографический список

1. Van der Waals J.D., Ph. D. Thesis, University of Leiden, 1873 / Waals J.D. van der. The continuity of the liquid and gaseous states. – In: Phys. Mem. L.: Taylor and Francis, 1890, vol. 1, part 3, p. 333-496.
2. Менделеев Д.И. Сочинения // под общ. редакцией акад В.Г. Хлепина, куратор тома В.Я. Курбатов, изд-во АН СССР, Ленинград, Москва. Т. V, 1947 310 С.
3. Каганер М.Г. Максимумы термодинамических свойств и переход газа к жидкости в надкритической области / Журнал физической химии. 1958. Т. XXXI1. №2. С.332-340.
4. Stewart R.B., Jacobsen R.T. Thermodynamical Properties of Argon from the TriplePoint to 1200 K with Pressures to 1000 MPa / J. Phys. Chem. Ref. Data. 1989. V.18. №2. P. 639.
5. Сычев В.В. Дифференциальные уравнения термодинамики / М. “Высшая школа”, 1991, 224 С.
6. Wei Zong, Changming Xiao, Yongkai Zhu. Modified Van der Waals equation and law of corresponding states / Physica A, 471 (2017) 295-300.
7. Воробьев В.С., Апфельбаум Е.М. Обобщенные законы подобия на основе некоторых следствий уравнения Ван-дер-Ваальса / Теплофизика высоких температур. 2016, Т.54, №2, С. 186-196.
8. Каплун А.Б., Мешалкин А.Б. Простое уравнение состояния жидкости и газа типа уравнения Ван-дер-Ваальса / Журнал физической химии. 2001, Т.75, №12, С.2135-2141.
9. Каплун А.Б., Мешалкин А.Б. Приближённые и высокоточные уравнения состояния однокомпонентных нормальных веществ / Журнал физической химии. 2006, Т.80, №11, С.2097-2102.

## THERMODYNAMIC EFFECTS OF VAN DER WAALS EQUATION IN THE AREA OF CRITICAL CONDITION

*Kalinichenko Boris A.*

Donetsk National Technical University,  
Donetsk People's Republic, Donetsk, st. Artyoma, 58, kreefore@mail.ru

Based on the Van der Waals equation, the temperature dependencies of some thermodynamic properties on the line of coexisting phases were calculated. The obtained theoretical results are compared with experimental data.

**Keywords:** equation of state, thermodynamic functions, phase transition, critical point, densities of coexisting phases.

## МЕТОДЫ АДАПТАЦИИ НЕЧЕТКИХ СИСТЕМ УПРАВЛЕНИЯ

*Хижняков Юрий Николаевич, Южаков Александр Анатольевич,  
Никулин Вячеслав Сергеевич, Сторожев Сергей Александрович*

Пермский национальный исследовательский политехнический университет,  
614000, Россия, г. Пермь, Комсомольский пр., 26, kalif23@yandex.ru

В статье представлен обзор и описание различных методов адаптации для нейро-нечетких систем. Рассмотрено дальнейшее развитие нечетких систем управления с целью придания им адаптивных свойств с учетом требований системы реального времени.

**Ключевые слова:** нечеткие нейронные сети, генетический алгоритм, база знаний, функция принадлежности, фаззификатор, дефаззификатор.

**Введение.** Из анализа мягких вычислений как сочетание нейронного и нечеткого управлений следует:

– субъективное определение вида функции принадлежности и ее параметров в продукционных правилах.

– определение оптимальной структуры нечетких нейронных сетей (ННС) типа ANFIS в задачах обучения с учителем (аппроксимация обучающего сигнала с требуемой ошибкой и с минимальным количеством продукционных правил в базе знаний (БЗ)).

– применение генетического алгоритма (ГА) в задачах многокритериального управления при наличии дискретных ограничений на параметры объекта управления.

Возможны следующие варианты развития интеллектуальных систем управления (ИСУ) в контуре управления с ПИД-регулятором.

### Интеллектуальные ИСУ первого поколения

- Выбор функции принадлежности зависит от уровня квалификации эксперта и его объективности.
- Проблема формирования БЗ НР, так как отсутствует алгоритм формирования БЗ.
- Требуется знаний по выбору вариантов реализации фаззификации, нечеткого вывода и дефаззификации.

### Интеллектуальные ИСУ второго поколения

Дополнительное введение **интеллектуальной глобальной обратной связи** (ИГОС) позволило извлекать объективные знания непосредственно из самого динамического поведения регулирующего органа (РО) и исполни-

тельного устройства САУ. Генетический алгоритм (ГА) извлекает информацию об оптимальном сигнале управления, а нечеткая нейронная сеть (ННС) аппроксимирует данный сигнал с помощью заданной структуры нейронной сети. Основным блоком в структуре является система моделирования оптимального сигнала управления (СМОСУ) с помощью ГА и выбранного критерия качества. На основе СМОСУ можно построить БЗ НР, представленной данной ННС, настраиваемой методом обратного распространения ошибки (ОРО) либо методом Левенберга-Марквардта.

### **Интеллектуальные ИСУ третьего поколения**

За счет наличия отрицательной обратной связи и глобальной интеллектуальной обратной связи реализован принцип не разрушения нижнего уровня управления в соответствии с приоритетом уровней качества управления. Тем самым было определено узкое место в структуре искусственных систем управления (ИСУ), например, персептрон. Для повышения уровня робастности был введен **оптимизатор баз знаний (ОБЗ)**, который формирует БЗ НР с целью увеличения количественных мер и частных критериев качества управления.

Рассмотрим дальнейшее развитие нечетких систем управления с целью придания им адаптивных свойств с учетом требований системы реального времени. К ним относятся следующие методы адаптации.

#### **1. Алгоритм адаптации Уидроу-Хоффа**

Уидроу и Хофф модифицировали персептронный алгоритм Ф. Розенблатта, дополнительно введя сигмоидную функцию активации вместо пороговой функции активации [1].

Суть адаптации заключается в итерационной процедуре сведения до нуля минимума квадрата ошибки.

Достоинством алгоритма Уидроу-Хоффа является отсутствия ограничения на вид функций принадлежности (терма) и ее расположение в нормированном интервале, а недостаток – сравнительно низкое быстродействие.

#### **2. Модифицированный метод наименьших квадратов [2]**

Для повышения быстродействия предлагается алгоритм адаптации, основанный на модификации метода наименьших квадратов [1].

Данный метод применим только для линейных терм, основания которых равны и равномерно расположены в нормированном интервале. Число терм терм-множества фаззификатора всегда должно быть равно числу подинтервалов разбиения нормированного интервала.

#### **3. Алгоритм последовательного обучения [2]**

Данный метод относится к методам последовательной идентификации линейных систем в реальном времени, основанный на принципе обучения с моделью импульсной характеристики.



Особенность данного метода адаптации заключается в том, что требуется одна итерация для коррекции степеней принадлежности фаззификатора и отсутствие ограничений на число терм в терм-множестве, которые можно располагать на интервале  $0...+1$  с равным основанием последних.

#### 4. Метод математического ожидания или метод «пройденного» пути

Метод предполагает применение треугольных терм, число которых равно числу входных переменных.

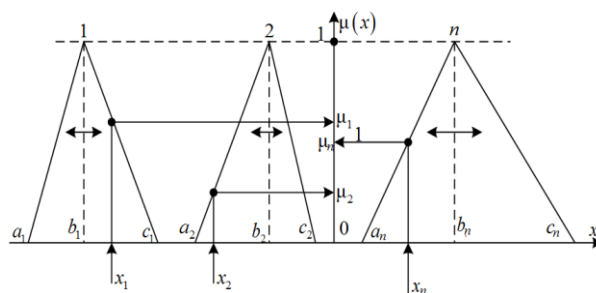
На рис. 1 показано расположение треугольных терм 1, 2, ...,  $n$  входов и синглтоны по всем переменным  $x_1, x_2, \dots, x_n$ , с помощью которых вычисляются степени принадлежности (выход фаззификатора).

Математическое описание треугольных терм фаззификатора выполнено согласно выражению [2].

$$\mu_i(x_i, a_i, b_i, c_i) = \begin{cases} 0, & x_i \leq a_i; \\ \frac{x_i - a_i}{b_i - a_i}, & a_i < x_i \leq b_i; \\ \frac{c_i - x_i}{c_i - b_i}, & b_i < x_i < c_i; \\ 0, & x_i \geq c_i; \end{cases}$$

Координата  $b_i$  вычисляется как среднее арифметическое значение входа блока адаптации за последние  $m$  тактов (значение  $m$  является настройкой блока адаптации а также расстояния между  $a$  и  $b$ ,  $b$  и  $c$ ).

Данный метод адаптации снимает субъективизм при расположении терм на нормированном интервале.



**Рисунок 1.** К пояснению вычисления степеней принадлежности блока адаптации, где  $x_i$  – текущие значения параметров ГТД;  $a_i, b_i, c_i$  – параметры треугольных функций принадлежности;  $\mu_i$  – степени принадлежности терм

Данный метод адаптации снимает субъективизм при расположении терм на нормированном интервале.

#### Метод активного эксперимента

Метод активного эксперимента предполагает определение базы знаний для преобразования в блоке нечеткой информации, который содержит адап-

тивный фаззификатор, блок правил, дефаззификатор, блок констант. Алгоритм работы предполагает для каждого входа в блоке фаззификации использовать две ассиметричные сигмоидные функции принадлежности (терм-множества). Значения из блока правил поступают на дефаззификатор. Дефаззификация происходит по методу среднего взвешенного [2].

### Метод смещения

Адаптация нечетких регуляторов заключается в смещении терм-множеств фаззификатора вдоль горизонтальной оси. Величина смещения определяется текущими значениями контуров управления [3-5].

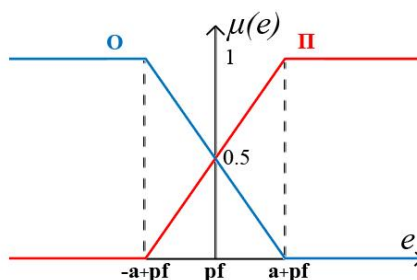


Рисунок 2. Терм-множество

Математическое описание терм-множества

$$\mu_{\pi}(e, a, pf) = \begin{cases} 0, & e \leq -a + pf; \\ \frac{e + a - pf}{2 \cdot a}, & -a + pf < e < a + pf; \\ 1, & e \geq a + pf; \end{cases}$$

$$\mu_0(e, a) = \begin{cases} 1, & e \leq -a + pf; \\ \frac{a + pf - e}{2 \cdot a}, & -a + pf < e < a + pf; \\ 0, & e \geq a + pf. \end{cases}$$

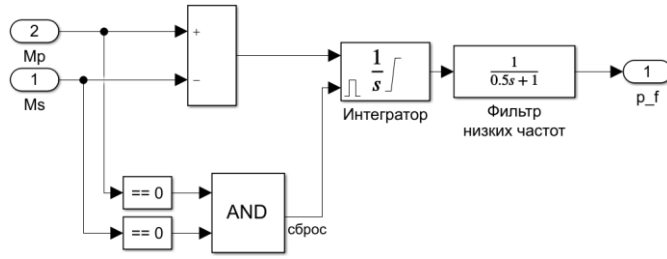
Расположение терм «О» и «П» представлено на рис. 2. Значение параметра « $pf$ » сдвигает терм-множество вдоль горизонтальной оси.

На рис. 3 приведен блок адаптации метода смещения. Во время работы контура приемистости (см. «Mr») терм-множества сдвигаются вправо (знак «+»), а во время работы контура сброса (см. «Ms») – влево (знак «-»). Интегратор имеет ограничения минимального и максимального значений  $\pm l$  (крайнее левое и крайнее правое положения терм-множеств). Сигнал сброса формируется в момент отключения контуров приемистости и сброса. Фильтр обеспечивает плавное перемещение терм-множеств.

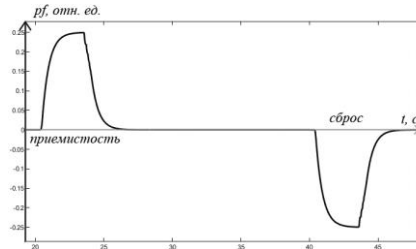
Пример изменения параметра « $pf$ » во время приемистости и сброса приведен на рис. 4.

### Метод разности степеней принадлежности

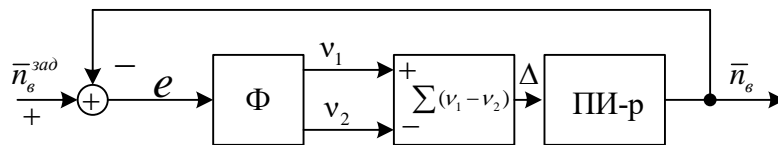
На рис. 5 приведена  $i$ -структурная схем блока адаптации.



**Рисунок 3. Структурная схема блока адаптации, где «Mr», «Ms» – индикаторы работы контуров приемистости и сброса; «p\_f» – значение сдвига терм-множеств**



**Рисунок 4. Значение параметра «p\_f» на приемистости и сбросе**



**Рисунок 5. Структурная схема блока адаптации**

Пусть в фаззификаторе заложены, например, сигмоидные функции

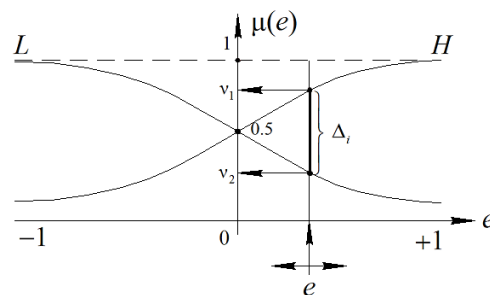
$$H = [1 + \exp(-b(e - C))]^{-1} \text{ при } b < 0;$$

$$L = [1 + \exp(-b(e - C))]^{-1} \text{ при } b > 0$$

Далее с помощью синглтона определяется разность степеней принадлежности как показано на рис. 6.

$$\Delta_i = (v_1 - v_2);$$

$$out = (k_1 \cdot \Delta_{li} + \int k_2 \cdot \Delta_{li} dt).$$



**Рисунок 6. Сигмоидные функции принадлежности**

Выход ПИ-регулятора блока адаптации подается в сумматор контура [5-6]. Особенностью рассмотренного метода адаптации является ограничение на количество терм в терм-множестве (не более двух), которые должны располагаться на интервале  $-1 \dots +1$  с равными основаниями.

**Заключение.** Представлен обзор и описание различных методов адаптации для нейро-нечетких систем. Рассмотрено дальнейшее развитие нечетких систем управления с целью придания им адаптивных свойств с учетом требований системы реального времени. Предложен вариант определения структуры с адаптивными свойствами нечетких нейронных сетей типа ANFIS в задачах обучения с учителем (аппроксимация обучающего сигнала с требуемой ошибкой и с минимальным количеством продукционных правил в базе знаний).

### Библиографический список

1. Мусатов, М. В. Анализ моделей метода наименьших квадратов и методов получения оценок / М. В. Мусатов, А. А. Львов // Вестник Саратовского государственного технического университета. – 2009. – Т. 4. – № 2(43). – С. 137-140.
2. Хижняков Ю.Н. Нечеткое, нейронное и гибридное управление: учеб. пособие / Ю.Н. Хижняков. – Пермь: Изд-во Перм. нац. исслед. политехн. ун-та, 2013. 303 с.
3. Storozhev, S. Gas Turbine Engine Combustion Chamber Pollutant Meter Using Neural Technology / S. Storozhev, Yu. Khizhnyakov, V. Nikulin // Proceedings of the 2021 IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering (ElConRus), 26-29 Jan., 2021. P. 934-936.
4. Чичерова Е.В. Использование алгоритмов нечёткой логики для управления частотой вращения силовой турбины газотурбинного двигателя // Труды МАИ. 2015. №81. URL: <http://trudymai.ru/published.php?ID=57812>.
5. Дорф, Р. Бишоп, Р. Современные системы управления / Пер. с англ. Б.И. Копылова – М.: Лаборатория базовых знаний, 2002.– 832 с.
6. Ротач, В.Я. Теория автоматического управления: учеб. для вузов. 3-е изд., перераб. и доп. – М. Изд-во МЭИ, 2004. – 400 с.

## METHODS FOR ADAPTATION OF FUZZY CONTROL SYSTEMS

*Khizhnyakov Yury N., Yuzhakov Alexander A.,  
Nikulin Vyacheslav S., Storozhev Sergey A.*

Perm National Research Polytechnic University,  
614000, Russia, Perm, Komsomolsky pr. 26, kalif23@yandex.ru

The article presents an overview and description of various adaptation methods for neuro-fuzzy systems. The further development of fuzzy control systems is considered in order to give them adaptive properties, taking into account the requirements of a real-time system.

**Keywords:** fuzzy neural networks, genetic algorithm, knowledge base, membership function, fuzzifier, defuzzifier.

*Научное издание*

**Искусственный интеллект в решении  
актуальных социальных и экономических проблем XXI века**

Сборник статей по материалам Восьмой всероссийской  
научно-практической конференции с международным участием

(г. Пермь, 25–26 октября 2022 г.)

Издается в авторской редакции  
Компьютерная верстка: *К.Н. Богдановой*

---

Подписано в печать 20.10.2022. Формат 60×90/16  
Усл. печ. л. 34,41. Тираж 28 экз. Заказ № 1402/2022.

---

Издательский центр  
Пермского государственного  
национального исследовательского университета  
614990, г. Пермь, ул. Букирева, 15

Отпечатано с готового оригинал-макета  
в типографии издательства Пермского национального  
исследовательского политехнического университета  
614990, г. Пермь, Комсомольский пр., 29, к. 113.  
Тел.: + 7 (342) 219-80-33.