



САНКТ-ПЕТЕРБУРГСКИЙ  
ГОСУДАРСТВЕННЫЙ  
ЭЛЕКТРОТЕХНИЧЕСКИЙ  
УНИВЕРСИТЕТ «ЛЭТИ»  
им. В.И. Ульянова (Ленина)



ПАВЛОВСКИЙ ЦЕНТР  
«ИНТЕГРАТИВНАЯ  
ФИЗИОЛОГИЯ - МЕДИЦИНЕ,  
ВЫСОКОТЕХНОЛОГИЧНОМУ  
ЗДРАВООХРАНЕНИЮ  
И ТЕХНОЛОГИЯМ  
СТРЕССОУСТОЙЧИВОСТИ»



ИНСТИТУТ ФИЗИОЛОГИИ  
им. И.П. Павлова  
РОССИЙСКОЙ  
АКАДЕМИИ НАУК



**IEEE**

РОССИЙСКАЯ  
СЕВЕРО-ЗАПАДНАЯ  
СЕКЦИЯ IEEE

**NEURONT.ETU.RU**

# V МЕЖДУНАРОДНАЯ КОНФЕРЕНЦИЯ ПО НЕЙРОННЫМ СЕТЯМ И НЕЙРОТЕХНОЛОГИЯМ

*20 июня 2024*

**NeuroNT'24**

## СБОРНИК ДОКЛАДОВ



Санкт-Петербург, 2024

Санкт-Петербургский государственный электротехнический университет  
«ЛЭТИ» им. В.И. Ульянова (Ленина) (СПбГЭТУ «ЛЭТИ»)

Павловский центр «Интегративная физиология – медицине, высокотехнологичному  
здоровоохранению и технологиям стрессоустойчивости»

Институт физиологии им. И.П. Павлова Российской академии наук (ИФ РАН)

Российская Северо-Западная секция Международного института  
инженеров электротехники и электроники IEEE

---

**V МЕЖДУНАРОДНАЯ КОНФЕРЕНЦИЯ  
ПО НЕЙРОННЫМ СЕТЯМ  
И НЕЙРОТЕХНОЛОГИЯМ  
(NeuroNT'2024)**

**20 июня 2024 года**

***Сборник докладов***

**Санкт-Петербург  
2024**

УДК 004.032.26

ISBN 978-5-7629-3369-8

V Международная конференция по нейронным сетям и нейротехнологиям (NeuroNT'2024): сб. докладов. СПб.: СПбГЭТУ «ЛЭТИ». 2024. – 234 с.

### Международный программный комитет

Почетный председатель	Проф. Ханнес Топфер (Германия)
Сопредседатели	Проф. М.С. Куприянов (СПбГЭТУ «ЛЭТИ», Россия) Проф. Ю.Е. Шелепин (ИФ РАН, Россия)
Члены международного программного комитета	Dr. S.P. Shiva Prakash, Associate Professor, Department of Information Science and Engineering, JSS Science and Technology University, Mysuru, Karnataka, India Dr. Balaraman Ravindran, Prof., Computer Science and Engineering & Robert Bosch Centre for Data Science and AI, Indian Institute of Technology Madras, Chennai, Tamil Nadu, India Dr. Thanh Nghi Nguyen, Prof., Head of the Department of Physics and Mathematics, Hanoi University of Technology, Vietnam Dr. Nigel Foreman, Prof., Middlesex university, Department of Psychology, London, UK Проф. Ю.А. Шичкина (СПбГЭТУ «ЛЭТИ», Россия) Доц. М.М. Заславский (СПбГЭТУ «ЛЭТИ», Россия) Проф. В.О. Полякова (СПбГУ, Россия) Доц. О.И. Буренева (СПбГЭТУ «ЛЭТИ», Россия) Доц. О.Л. Петросян (СПбГУ, Россия) Проф. С.А. Вакуленко (ИПМаш РАН, Россия) Доц. М.П. Чмелевский (ФГБУ «НМИЦ им. В.А. Алмазова», Россия) Доц. Е.А. Пустозеров (СПбГЭТУ «ЛЭТИ», Россия) Проф. О.В. Войцеховская (ФГБУН Ботанический институт им. В.Л. Комарова РАН, Россия) Р.О. Малашин, (ИФ РАН, Россия) Е.Ю. Шелепин (ИФ РАН, Россия)

### Организационный комитет

Председатель	Доц. А.А. Семенов (СПбГЭТУ «ЛЭТИ», Россия)
Члены организационного комитета	Доц. Я.А. Бекенева (СПбГЭТУ «ЛЭТИ», Россия) Проф. А.С. Красичков (СПбГЭТУ «ЛЭТИ», Россия) К.Ф. Котченко (ИФ РАН, Россия) А.А. Бойко (ИФ РАН, Россия) О.Н. Журавлева (СПбГЭТУ «ЛЭТИ», Россия)

ISBN 978-5-7629-3369-8

© СПбГЭТУ «ЛЭТИ», 2024

# Реализация биоподобных поведенческих функций мобильного робота на сегментной спайковой модели нейрона

В. В. Иванова<sup>1,2</sup>, А. М. Корсаков<sup>2</sup>, А. А. Демчева<sup>2,3</sup>, А. В. Бахшиев<sup>3</sup>

<sup>1</sup>Санкт-Петербургский государственный электротехнический университет «ЛЭТИ» им. В.И. Ульянова (Ленина)

<sup>2</sup>Центральный научно-исследовательский и опытно-конструкторский институт робототехники и технической кибернетики

<sup>3</sup>Санкт-Петербургский политехнический университет Петра Великого

overtime@gmail.com, a.korsakov@rtc.ru, a.demcheva@rtc.ru, palexab@gmail.com

**Аннотация.** В данной работе рассматривается реализация биоподобных поведенческих функций робота, а именно, взаимодействие моделей болевых ощущений и прогностических функций мозжечка на сегментной спайковой модели нейрона. В качестве мобильного робота была выбрана робототехническая платформа **Omegabot**.

**Ключевые слова:** сегментная спайковая модель нейрона; модель болевых ощущений; модель прогностических функций мозжечка; поведенческие функции

## I. ОБЗОР ЛИТЕРАТУРЫ

### A. Навигация мобильных роботов

О важности обеспечения обнаружения и предотвращения столкновений автономного мобильного робота с препятствием упоминается во множестве статей по робототехнике [1, 2]. У данной задачи, посвященной движению робота в пространстве, существует множество решений [3–7], и в большинстве из них предполагается адаптация робота к окружающей среде. Способность адаптироваться к переменным свойствам мира, в который помещен робот, является перспективным направлением разработки.

Создание автономных роботов осуществляют, в том числе, с применением искусственных нейронных сетей (ИНС) благодаря таким свойствам, как адаптивность, возможность параллельно выполнять множество однотипных операций. Биологическая достоверность поведения моделей, основанных на ИНС, может быть рассмотрен как альтернативный подход при создании навигационных систем для мобильных роботов.

В статье [8] рассмотрено движение робота вдоль стены, при этом робот не различает объекты, от внешней среды он получает только показания с датчиков расстояния. Поступающая с дальномеров информация преобразуется в импульсы, которые подаются на вход ИНС.

Результаты получены в рамках выполнения государственного задания Минобрнауки России №075-00697-24-00 от 27.12.2023 «Исследование и разработка биоподобной системы управления поведением мобильных роботов на базе энергоэффективных программно-аппаратных нейроморфных средств» (FNRG-2022-0016 1021060307690-3-1.2.1;2.2.2).

В статье [2], посвященной аналогичной задаче, предлагается решение с помощью спайковых нейронных сетей (SNN – Spiking Neuron Networks) и нейронов Ижикевича [9].

### B. Сегментная спайковая модель нейрона

Сегментная спайковая модель нейрона (Compartmental Spiking Neuron Model – CSNM) [10] воспроизводит строение биологического нейрона и его реакцию на внешние воздействия. CSNM способна распознавать паттерны импульсов, которые поступают ей на вход. При этом паттерны импульсов различаются только результатом временного кодирования. Преимущество сети из таких нейронов заключается в том, что она обладает способностью к обучению на малом количестве примеров. Структурная схема CSNM приведена на рис. 1.

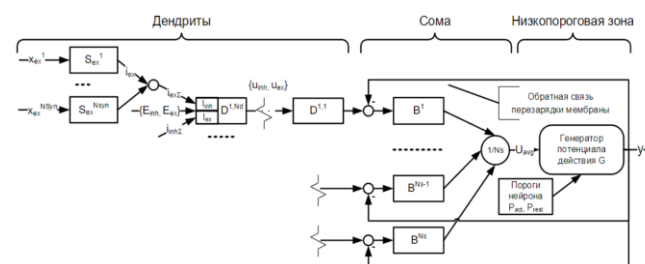


Рис. 1. Структура сегментной спайковой модели нейрона

CSNM характеризует набор нескольких гиперпараметров: размер нейрона, совокупность длин дендритов и число возбуждающих и тормозных синапсов для каждого сегмента мембраны. Они определяют конфигурацию модели обеспечивают формирование нужной реакции нейрона на входное воздействие. Так как нейроны обмениваются только информацией о моменте времени возникновения спайка, то расчет систем дифференциальных уравнений для каждого нейрона осуществляется обособленно от других.

В число смоделированных с помощью CSNM поведенческих функций входят условно-рефлекторная деятельность [11], возникновение болевых ощущений у живых организмов, прогностические функции мозжечка.

Модель болевых ощущений (детектор нештатных ситуаций) [12] предполагает два возможных события: штатную и нештатную ситуации. В первом случае на вход модели поступает паттерн, который похож на тот, который использовался в качестве обучающего примера, и реакция модели отсутствует. В противном случае система генерирует выходной предупреждающий сигнал.

Модель прогностических функций мозжечка (предиктор) [13] – это сеть элементарных ячеек, выполняющих операции сложения и вычитания по аналогии с биологическими нейронами (рис. 2).

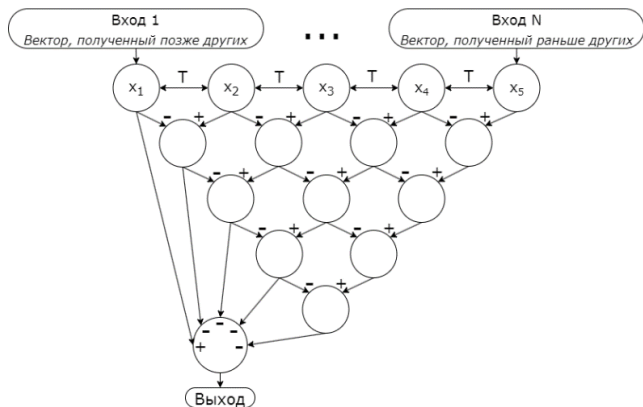


Рис. 2. Матмодель прогностических функций мозжечка

При этом выход модели – это предсказание следующего значения входного сигнала.

## II. ИНСТРУМЕНТАРИЙ

Эксперименты проводились как на модельной симуляции мобильного робота в среде Gazebo, так и на реальной робототехнической платформе «Omegabot».

### A. ROS2 и Gazebo

ROS2 (Robot Operating System, операционная система робота) представляет собой набор библиотек и инструментов разработчика соответствующих приложений. С ее помощью настраивались не только визуальная составляющая модели робота, рассматриваемого в данной статье, но и его физические характеристики.

Gazebo – это среда моделирования, в которую интегрировался робот, чтобы определить, как при настроенных параметрах он будет взаимодействовать с окружающим миром при текущих инерции, гравитации, освещенности и других параметрах окружающего мира.

### B. Мобильный робот «Omegabot»

Представляет собой робототехническую платформу с программируемыми модулями. Состоит из колесной платформы, контроллера, микрокомпьютера RPi3, базовой основы, аккумулятора блока и блока питания. Помимо этого, для проведения экспериментов использовался подключенный к базовой основе ультразвуковой дальномер HC-SR04.

Принцип работы дальномера HC-SR04 заключается в следующем. Логический уровень выхода Trig меняется с 0 на 1, когда на выход подается единичный импульс определенной длительности (10 мкс), преобразующийся

в восемь импульсов с частотой 40 КГц. Эти импульсы через излучатель передаются во внешнюю среду, после чего логический уровень Trig снова становится равен 0. После отражения от препятствия импульсы фиксируются приемником и обеспечивают изменение логического уровня выхода Echo с 0 на 1. Время задержки сигнала на эхолотаторе преобразовывается в расстояние до объекта.

Точность измерения дальномера определяется многими факторами. Чем больше расстояние до объекта, тем сильнее затухает сигнал и возрастает вероятность его отражения от соседних препятствий. Температура и влажность воздуха также оказывают влияние на показания датчика, поскольку скорость распространения ультразвука зависит от параметров окружающей среды.

Для того чтобы обеспечить запуск скрипта управления, к робототехнической платформе подключались экран, мышь и дополнительный блок питания.

## III. РЕАЛИЗАЦИЯ

### A. Система управления мобильным роботом на основе биоподобного блока «Рефлекс-Мозжечок-Боль»

Действие формировалось на основе системы принятия решений. Если при определенном состоянии неоднократно срабатывала модель болевых ощущений, то элемент системы управления «рефлекторная дуга» блокировал действие. Элемент системы управления «мозжечок + боль» позволял предсказывать блокировку действия до того, как срабатывал болевой нейрон. Если действие не выполнялось, то в блок системы принятия решений поступал соответствующий сигнал, и случайным образом формировалось другое действие из числа доступных роботу.



Рис. 3. Система «Рефлекс-Мозжечок-Боль»

В данной работе были реализованы только два элемента системы управления (выделены линиями «штрих» и «штрих-пунктир» на рис. 3) и блок фильтрации данных методом наименьших квадратов.

Скрипт для управления роботом был написан на языке программирования Python.

### B. Реализация в среде моделирования

Чтобы избежать столкновения мобильного робота с препятствием, из всех значений данных, полученных с

дальномера, вычиталась некоторая величина, одинаковая для всех экспериментов – 0.5 метра. Она соответствовала так называемому расстоянию до виртуальной стены, у которой и останавливались модель и робот. Этот параметр можно было настраивать в процессе движения робота – расстояния до препятствия в момент нажатия какой-либо клавиши становилось расстоянием до невидимой стены.

В случае реализации элемента системы управления «боль» в среде Gazebo остановка робота определялась наличием срабатывания обученного нейрона. Если это происходило, то модель робота останавливалась, и в терминал выводилось соответствующее сообщение.

При реализации элемента системы управления «боль + мозжечок» болевой нейрон был реализован также как и в предыдущем случае, но модель робота останавливалась раньше, чем он успевал сработать. Наличие реализации поведенческой функции мозжечка позволило предсказывать расстояние до невидимой стены на основе имеющихся данных и передавать их на второй нейрон, который срабатывал немного раньше, чем болевой (рис. 4).

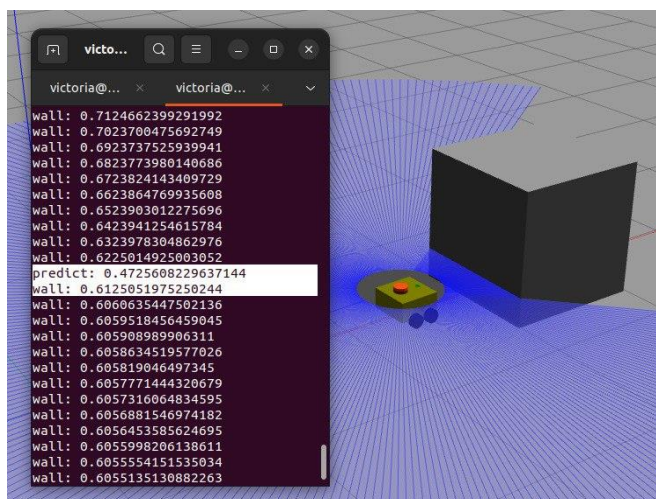


Рис. 4. Симуляция остановки робота по срабатыванию нейрона

То, насколько раньше срабатывал предсказывающий нейрон по сравнению с болевым, определялось расстоянием, которое проезжал робот между двумя соседними измерениями расстояния до препятствия – чем оно больше, тем дальше от препятствия останавливался робот.

### С. Реализация на реальном роботе

В связи с тем, что данные, полученные с дальномера, не всегда соответствовали действительности (в них присутствовали выбросы, которые значительно влияли на работу блока «мозжечок»), то также была реализована фильтрация поступающих данных. По методу наименьших квадратов строилась линия тренда, по которой расстояния пересчитывались, и уже на их основе осуществлялось предсказание (рис. 5).

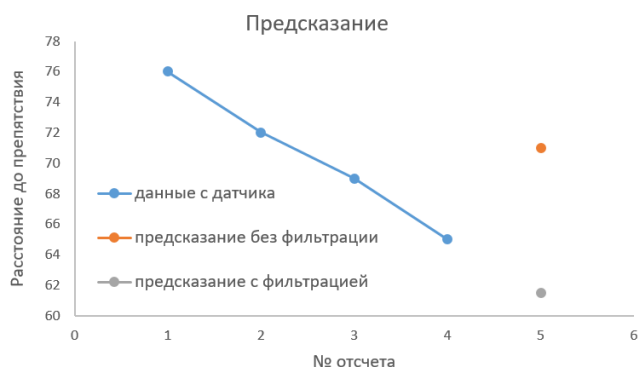


Рис. 5. Построение линии тренда

Реализация элементов системы управления мобильным роботом осуществлялась по аналогии с тем, что было смоделировано в среде Gazebo. Синей линией на рис. 6 обозначена виртуальная стена.

В случае одного нейрона при остановке у препятствия, помимо вывода соответствующего сообщения в терминал, на роботе загорался красный светодиод (рис. 6, слева).

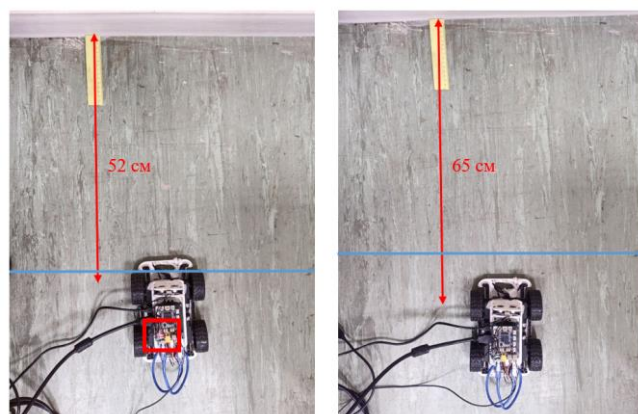


Рис. 6. Остановка робота у препятствия

При наличии предсказывающего нейрона остановка происходила раньше и примерно соответствовала модельным данным (рис. 6, справа). При этом красный светодиод не загорался, а значит, болевой нейрон не успевал сработать.

## IV. ЗАКЛЮЧЕНИЕ

В работе были реализованы два блока системы управления мобильным роботом на основе биоподобного блока «Рефлекс-Мозжечок-Боль».

Перспектива дальнейшей работы заключается в том, чтобы реализовать всю схему «Рефлекс-Мозжечок-Боль», тем самым обучив мобильного робота избегать препятствий, находящихся рядом с ним.

### СПИСОК ЛИТЕРАТУРЫ

- [1] Soumare, S., Ohya, A., and Yuta, S. (2002). Real-time obstacle avoidance by an autonomous mobile robot using an active vision sensor and a vertically emitted laser slit // In Proceedings of the IEEE International Conference on Robotics and Automation, pages 3068–3073.
- [2] Mantovani R.G., Ayrosa P. Use of spiking neural networks for robot navigation // Tech. Rep, 2013.

- [3] Horiuchi T.K. (2009). A spike-latency model for sonar-based navigation in obstacle fields. *Trans. Cir. sys. Part I*, 56(11):2393–2401.
- [4] Zhu Y. et al. Rule-based reinforcement learning for efficient robot navigation with space reduction // *IEEE/ASME Transactions on Mechatronics*. 2021. Т. 27. №. 2. С. 846-857.
- [5] A.F. Foka and P.E. Trahanias, "Predictive autonomous robot navigation," // *IEEE/RSJ International Conference on Intelligent Robots and Systems*, Lausanne, Switzerland, 2002, pp. 490–495 vol.1
- [6] Frommberger L. Learning to behave in space: A qualitative spatial representation for robot navigation with reinforcement learning // *International Journal on Artificial Intelligence Tools*. 2008. Т. 17. №. 03. С. 465-482.
- [7] Remazeilles A., Chaumette F. Image-based robot navigation from an image memory // *Robotics and Autonomous Systems*. 2007. Т. 55. №. 4. С. 345-356.
- [8] E. Nichols, L. J. McDaid, N. Siddique. Biologically Inspired SNN for Robot Control // *IEEE Transactions on Cybernetics*, 2012. V. 43(1). P. 115–128.
- [9] Izhikevich E.M. Simple model of spiking neurons // *IEEE Transactions on neural networks*, 2003. V. 14(6). P. 1569–1572
- [10] Bakhshiev A., Demcheva A., Stankevich L. (2021) CSNM: The Compartmental Spiking Neuron Model for Developing Neuromorphic Information Processing Systems // *International Conference on Neuroinformatics*, pp. 327–333.
- [11] Корсаков А.М., Бахшиев А.В., Астапова Л.А., Станкевич Л.А. (2021). Реализация поведенческих функций на спайковых нейронных сетях. // *Информатика и автоматизация*, 20(3), 591-622.
- [12] Демчева А.А., Корсаков А.М., Бахшиев А.В. Моделирование болевых ощущений для решения задач бинарной классификации // *Экстремальная робототехника: Центральный научно-исследовательский и опытно-конструкторский институт робототехники и технической кибернетики*. 2022. Т. 1. №. 1. С. 355-361.
- [13] Предупреждение возникновения критических ситуаций в сложных технических системах с использованием нейроморфного подхода / А.А. Демчева [и др.] // *Робототехника и техническая кибернетика*. Т. 11. № 4. Санкт-Петербург: ЦНИИ РТК. 2023. С. 281-291. Текст : непосредственный

# Оценка состояния человека на основе данных о его работе на компьютерной клавиатуре

Даньтин Ма

Факультет компьютерных технологий и информатики  
Санкт-Петербургский государственный электротехнический университет  
«ЛЭТИ» им. В.И. Ульянова (Ленина)

dantingma65@gmail.com

**Аннотация.** В данном исследовании представлена разработанная нами модель для оценки эмоциональных состояний человека, использующая данные, полученные в результате взаимодействия с компьютерной клавиатурой. В модель интегрирован механизм обратной связи, что значительно повышает точность предсказаний эмоциональных состояний. Благодаря систематическому анализу динамики использования пользователем клавиатуры и наблюдению за отдельными факторами окружающей среды, наш подход не только предсказывает состояние пользователя, но и адаптируется к изменениям в поведении пользователя. Эффективность модели была подтверждена путем проведенных экспериментов. Способность модели повышать адаптивность к человеку и персонализация пользовательских интерфейсов улучшает взаимодействие человека с интеллектуальной системой. Разработанная система показывает направления улучшений в области человеко-машинных интерфейсах, предоставляя основу для будущей интеграции в более сложные системы, направленные на всеобъемлющую оценку психофизиологических состояний.

**Ключевые слова:** механизмы обратной связи, машинное обучение, взаимодействие человека и машины, оценка эмоционального состояния

## I. ВВЕДЕНИЕ

В быстро развивающуюся цифровую эпоху взаимодействие людей с различными технологическими устройствами в повседневной жизни и на работе постоянно увеличивается. Это все более тесное взаимодействие между человеком и машиной не только меняет способы работы, но и оказывает глубокое влияние на психологическое и физиологическое здоровье человека. Поэтому крайне важно точно оценивать состояние людей при использовании этих технологий, что не только помогает повысить эффективность работы, но и важно для обеспечения здоровья и благополучия пользователей.

С постоянным прогрессом технологий, направленных на создание пользовательского интерфейса повышается возможность анализировать данные о взаимодействиях

пользователей с устройствами для изучения их психологических и физиологических состояний. В частности, клавиатуры и мыши [1], мониторинг поведенческих моделей при использовании этих устройств могут эффективно отражать уровень стресса, эмоциональные изменения и уровень усталости пользователя [2, 3]. Эти подходы к анализу поведения предоставляют возможности для реального и ненавязчивого мониторинга состояний, являясь экономичными и эффективными решениями [4].

В настоящее время управление стрессом и сохранение эмоционального здоровья на рабочем месте становятся важными задачами для предприятий и организаций. Реальное время мониторинга психофизиологического состояния сотрудников позволяет своевременно корректировать рабочую нагрузку, предотвращая здоровьесберегающие проблемы, такие как профессиональное выгорание и различные травмы. Кроме того, мониторинг состояния человека также способствует оптимизации дизайна пользовательского интерфейса, делая его более удобным для пользователя, что, в свою очередь, улучшает производительность человека и интеллектуальной системы.

Данное исследование направлено на изучение того, как данные о взаимодействии человека с различными устройствами, такими как клавиатура, мышь и другие, могут быть использованы для оценки психологических и физиологических состояний пользователей, а также как эта информация может помочь в дизайне более эргономичных пользовательских интерфейсов. Результаты исследования могут быть вкладом в дизайн взаимодействия человека с компьютером, управление здоровьем на рабочем месте и оптимизацию рабочего процесса.

## II. СВЯЗАННЫЕ РАБОТЫ

В области взаимодействия человека с компьютером исследователи активно изучают, как можно более эффективно фиксировать состояния пользователей с помощью данных о поведении, полученных с клавиатуры и других устройств ввода. Эти устройства позволяют рассчитывать такие метрики, как скорость печати, интервалы между нажатиями клавиш, продолжительность и сила нажатий и т. п. Существующие исследования показали, что эти метрики

Работа выполнена при поддержке Минобрнауки России в рамках соглашения № 075-15-2021-291 от 15.04.2022г о предоставлении гранта в форме субсидий из федерального бюджета на осуществление государственной поддержки создания и развития научного центра мирового уровня «Павловский центр «Интегративная физиология – медицине, высокотехнологичному здравоохранению и технологиям стрессоустойчивости»



способны отражать изменения в эмоциональном состоянии, уровне усталости и стрессе пользователей.

Исследование Monrose и Rubin 1997 года стало пионерским в изучении возможности использования анализа динамики клавиатурного ввода для аутентификации пользователей при вводе паролей [5]. Их метод, заключающийся в записи таких деталей, как интервалы времени и продолжительность нажатий при вводе паролей, успешно идентифицировал отдельных пользователей. Это открытие подчеркнуло потенциал анализа динамики клавиатурного ввода в области безопасности, хотя он в значительной степени зависел от постоянства поведения пользователей, не учитывая возможные изменения, вызванные эмоциональными или окружающими условиями.

За этим последовали исследования Gunetti и Picardi в 2005 году, которые расширили применение анализа динамики клавиатурного ввода, предложив метод идентификации пользователей и их эмоциональных состояний через анализ свободного текста [6]. Они попытались создать модель распознавания эмоций, статистически анализируя скорость печати и привычки нажатия клавиш пользователей. Несмотря на то, что этот метод мог предоставлять обратную связь о эмоциональном состоянии пользователя в реальном времени, его предположение о стабильности поведения отдельного лица ограничивало его применение в меняющихся условиях.

В 2007 году исследование Guven и коллег указало на то, что поведенческие модели при вводе коротких текстов, таких как пароли, легко предсказуемы и подражаемы, поэтому они рекомендовали проводить анализ поведения клавиатуры на более длинных текстах для повышения безопасности аутентификации [7]. Хотя этот подход увеличивал безопасность систем паролей, он также мог увеличить нагрузку на пользователя.

Кроме того, с развитием технологий анализа данных, последние исследования, такие как работы Hernandez и других (2014) и Li и других (2018), использовали сложные алгоритмы машинного обучения для анализа более тонких параметров динамики клавиатуры, таких как сила нажатия и продолжительность, для определения стресса и эмоциональных изменений у пользователей [8, 9]. Эти методы достигли прогресса в повышении точности распознавания эмоций, однако они также сталкиваются с проблемами обработки больших объемов данных и обучения алгоритмов.

Исследование Giancardo и других (2016) открыло новую область в мониторинге здоровья с помощью анализа поведения путем отслеживания характеристик ввода данных на клавиатуре, обнаружив, что незначительные изменения в моделях использования клавиатуры могут использоваться для раннего обнаружения симптомов неврологических заболеваний, таких как болезнь Паркинсона [10]. Этот метод ненавязчивого мониторинга здоровья предлагает новый инструмент для ранней диагностики, хотя его широкое применение все еще требует дальнейшей проверки.

Эти исследования демонстрируют значительный потенциал и сложности использования клавиатур и других устройств ввода для оценки состояния

пользователей в практических приложениях. Будущие исследования должны более глубоко изучить, как интегрировать различные показатели поведения и факторы окружающей среды для повышения точности и применимости этих методов, а также обратить внимание на учет пользовательского опыта в таких системах и обеспечение безопасности и конфиденциальности данных.

### III. ПРИМЕР СОЗДАНИЯ СИСТЕМЫ ДЛЯ ОПРЕДЕЛЕНИЯ ЭМОЦИОНАЛЬНЫХ СОСТОЯНИЙ ЧЕЛОВЕКА

#### A. Разработка программного обеспечения для сбора и анализа данных

Разработанное для этого исследования программное обеспечение предназначено для мониторинга эмоционального состояния человека каждые 15 минут. Программа состоит из двух основных компонентов:

- **Полный сбор данных и маркировка эмоций:** фиксирует все взаимодействия с клавиатурой, периодически предлагая пользователям указывать свое эмоциональное состояние (active, average, inactive, tired). Программное обеспечение регистрирует дату, время и нажатые клавиши, записывая «по keystroke» в периоды неактивности. Оно выбирает от 5 до 10 позиций клавиш для анализа.
- **Персонализированный сбор данных и маркировка:** отслеживает семь конкретных клавиш (включая spacebar, backspace, delete и стрелки) и запрашивает обновления эмоционального состояния каждые 15 минут.

Анализ значимости признаков играет ключевую роль в оптимизации производительности модели машинного обучения, фильтруя избыточные признаки для фокусировки на предсказательных элементах и предотвращения переобучения [11]. Это исследование использует две стратегии выбора признаков после предварительной обработки данных для повышения эффективности модели. Первая сочетает пороговую вариативность [12] и рекурсивное устранение признаков (RFE) [13] для исключения признаков с низкой вариативностью и определения ключевых характеристик с помощью классификатора случайных лесов. Вторая стратегия использует комбинацию алгоритмов (включая chi-square tests, F-tests, and mutual information) для выделения признаков, критически важных для производительности модели. Выбранные наборы признаков стандартизируются и оцениваются с помощью кросс-валидации на 10 сегментах, обеспечивая точность и надежность модели.

#### B. Предварительная обработка собранных данных с клавиатуры

В ходе исследования были собраны данные об активности на клавиатуре у шестнадцати участников в течение двух недель, насчитывающие около 250 000 записей. Каждая запись содержит дату, время, конкретную нажатую клавишу и эмоциональное состояние пользователя (например, активное, неактивное). Для анализа данные были агрегированы в

15-минутные интервалы, что упрощает статистическую оценку и идентификацию шаблонов.

ТАБЛИЦА I. ПРИМЕР НАБОРА КЛЮЧЕВЫХ ДАННЫХ

Day	Time	Button	Human state
2023/3/18	23:37:29,851	'o'	
2023/3/18	23:37:30,037	Key.space	
2023/3/18	23:37:30,273	','	
2023/3/18	23:37:49,477	Key.ctrl_1	Active

- **Day and Time:** Каждая запись документирует дату и точное время (в миллисекундах) активности на клавиатуре.
- **Button:** Конкретные клавиши, нажатые во время активности.
- **Human State:** Отражает состояние индивида, записанное с интервалом в 15 минут с использованием метода ближайшего соседа.

Для анализа данных был выбран подход с временными окнами. Этот подход позволяет сосредоточиться на фиксации параметров эмоционального состояния в интервалах: за десять минут до и пять минут после запрашивания состояния пользователя. Этот метод позволяет рассчитывать метрики, такие как частота нажатий клавиш, интервалы и средние значения, необходимые для обучения модели.

Первичный анализ данных заключался в оценке частот нажатий клавиш в различных состояниях, выявляя вариации во взаимодействии с клавиатурой в зависимости от эмоциональных состояний. Полученные результаты можно разбить на два класса:

- **High Activity State:** Характеризуется частым использованием определенных клавиш и более быстрым ритмом печати, что свидетельствует об интенсивных действиях, например, письме.
- **Low Activity State:** Показывает более длинные интервалы и большую вариативность, что предполагает минимальное взаимодействие с клавиатурой, возможно, во время перерывов.

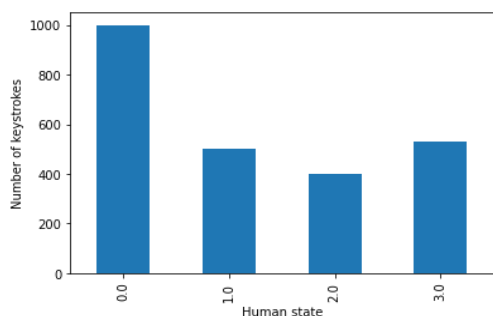


Рис. 1. Общее количество нажатий клавиш по состояниям

ТАБЛИЦА II. ЧАСТОТА КЛАВИШ В АКТИВНОМ И НЕАКТИВНОМ СОСТОЯНИЯХ

Human state	Most frequently pressed key	Count	Least frequently pressed keys	Count
0.0	'a'	77.400000	'8'	0.066667
2.0	Key.space	40.187500	'!'	0.062500

Рис. 1 и табл. 2 демонстрируют примеры использования клавиш в активном и неактивном состояниях, для различных классов состояния человека.

ТАБЛИЦА III. СТАТИСТИКА ИНТЕРВАЛОВ ВРЕМЕНИ

state	interval_mean	interval_std
Active	49.204269	1855.359851
Average	103.909600	7079.682762
Tired	71.000213	5864.714652
Inactive	188.253002	11017.443970

Анализ времени нажатий клавиш и интервалов, приведенных в табл. 3, позволил выявить различные варианты взаимодействия с клавиатурой в четырех эмоциональных состояниях:

- **Active State:** Демонстрирует самый короткий средний интервал между нажатиями клавиш, что указывает на быструю и последовательную печать, часто наблюдаемую во время сосредоточенной деятельности, такой как целенаправленный набор текста или программирование.
- **Moderate State:** Представляет более широкий диапазон интервалов нажатий клавиш, отражая смешение печати текста и других действий (например, чтение или размышления), что приводит к сбалансированному темпу.
- **Tired State:** Хотя средний интервал немного короче, чем в среднем состоянии, увеличенное стандартное отклонение предполагает нерегулярные варианты набора текста, вероятно, из-за снижения концентрации или частых пауз.
- **Inactive State:** Характеризуется самыми длинными средними интервалами и наибольшей вариативностью, что указывает на минимальное взаимодействие с клавиатурой, возможно, во время перерывов или задач, не связанных с печатью.

C. Улучшение процесса принятия решений через механизмы обратной связи

Механизмы обратной связи имеют ключевое значение в разработке интеллектуальных помощников человека, способствуя динамичному взаимодействию между пользователями и системой. Дизайн таких систем подразумевает использование продвинутых моделей с использованием технологий ИИ для анализа данных, включая обработку биометрических и видеоданных для определения эмоциональных состояний, тем самым повышая точность принятия решений. Обратная связь от пользователя играет двойную роль:

**Подтверждение и обучение:** Положительная обратная связь подтверждает и укрепляет решения системы, во время как отрицательная обратная связь, предоставленная в виде текста или голоса, анализируется с помощью ИИ (например, с использованием техник BERT и NER) для идентификации вариантов принятия решений и уточнения модели.

**Итеративное улучшение:** Система развивается, используя три ключевых процесса:

- Анализ объяснений пользователя: уточняет модель принятия решений, принимая и понимая обратную связь пользователя.

- Генерация ответов, похожих на человеческие: адаптирует обратную связь, чтобы соответствовать когнитивным паттернам человека.

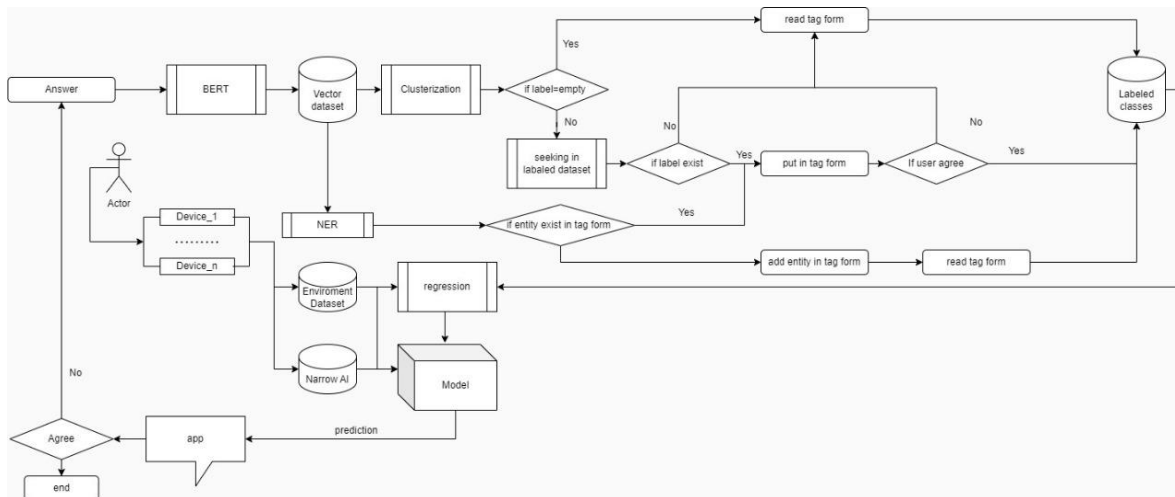


Рис. 2. Архитектура модуля обратной связи

- Оценка состояний и поведения пользователя: использует обратную связь пользователя для оценки эмоциональных состояний и динамики поведения, обеспечивая целостное понимание взаимодействий с пользователем.

Такой подход способствует циклу непрерывного обучения и адаптации, стремясь к оптимальным решениям, которые повышают как точность решений, так и удовлетворенность пользователя.

#### D. Организация обратной связи

Система улучшает понимание и точность прогнозирования эмоциональных состояний, включая модуль обратной связи. Этот модуль анализирует активность на клавиатуре и факторы окружающей среды, используя многослойный перцептрон для оптимизации точности прогноза. Регулярное взаимодействие пользователя укрепляет связь между прогнозами и реальными эмоциональными состояниями, служа важным входом для улучшения модели.

Влияние окружающей среды на эмоциональные состояния оценивается через внешние API, такие как OpenWeatherMap [14], которые предоставляют подробные данные об окружающей среде (температура, влажность, давление), обновляемые каждые 15 минут. Это позволяет определять изменения в окружающей среде, влияющие на эмоциональные состояния и учитывать их в прогнозах. Пример данных об окружающей среде приведен в табл. 4.

ТАБЛИЦА IV. Данные окружающей среды: 27 АВГУСТА 2023

Time	Temperature	Humidity	Pressure	Temperature Change	Humidity Change	Pressure Change
14:27	21.14	56	1007	17.28	76	10
14:38	21.14	56	1007	17.28	76	10

Пользователи улучшают точность прогнозов, предоставляя обратную связь о своих состояниях, которую система обрабатывает с помощью техник текстового анализа и оценки состояний. Сложная логическая модель интегрирует эти данные с данными о среде, применяя правила, связанные с конкретными условиями, для описания эмоциональных состояний. Эта модель постоянно уточняется через механизм динамической корректировки веса, который перенастраивает влияние машинного обучения и логических моделей на основе обратной связи пользователя. Это обеспечивает отзывчивость системы как к вводу пользователя, так и к изменениям в окружающей среде, поддерживая высокую точность и актуальность прогнозов.

Динамическая корректировка чувствительна к обратной связи, увеличивая вес моделей, которые пользователи считают более точными. Регулярный анализ обратной связи позволяет непрерывно оптимизировать модель, отражая адаптивность системы к восприятиям пользователей и контексту окружающей среды.

## IV. РЕЗУЛЬТАТЫ

### A. Экспериментальный дизайн и методология

В этом исследовании оценивалось влияние двух механизмов обратной связи на точность прогнозирования эмоций: простой системы обратной связи и усовершенствованной системы с логической моделью и динамическим взвешиванием. Простая система использовала базовые ответы согласие/несогласие, в то время как усовершенствованная система предоставляла подробную текстовую обратную связь и корректировала прогнозы на основе ввода пользователя и факторов окружающей среды.

### В. Сравнение точности прогнозов

Первоначальный анализ обратной связи показал:

- Предпочтение нейронной сети – 40 %.
- Логической модели – 30 %.
- 30 % не имели предпочтения, находя обе модели одинаково точными или неточными.

Предпочтения пользователей менялись в зависимости от внешних условий, таких как праздники, что подчеркивает важность адаптивного взвешивания моделей для улучшения точности. Подробный сравнительный анализ показал, что внедрение логической модели и механизма динамической корректировки веса увеличило среднюю точность прогнозов системы с 70 % до 85 %.

Рис. 3 демонстрирует сравнение точности между двумя системами, показывая значительное улучшение с усовершенствованным механизмом обратной связи.

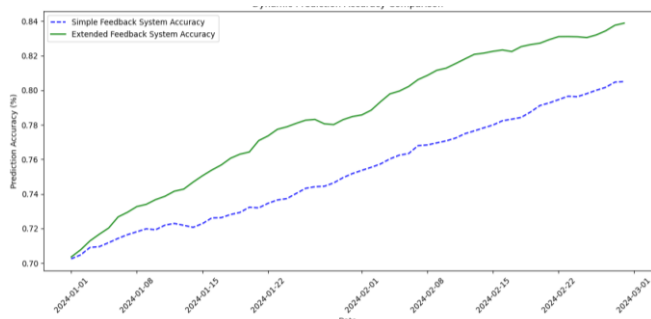


Рис. 3. Точность простых и расширенных систем обратной связи

### С. Тенденции динамической корректировки весов моделей

Анализ показал четкие тенденции в корректировке весов моделей между буднями и выходными/праздниками:

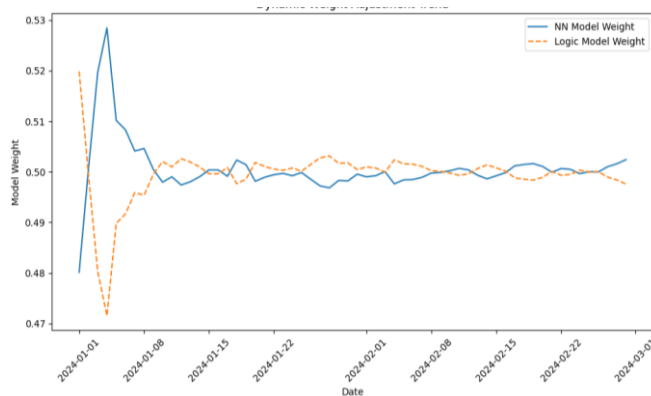


Рис. 4. Корректировка весов модели: будни против выходных/праздничных дней

В будние дни наблюдалось увеличение веса модели нейронной сети на 13 %, что лучше отражало поведение пользователя через активность на клавиатуре. В выходные/праздничные дни вес логической модели увеличивался на 18 %, показывая ее эффективность в анализе свободной пользовательской обратной связи в периоды стабильной активности пользователя.

### V. ЗАКЛЮЧЕНИЕ

Разработанная модель для оценки состояния человека на основе данных его работы на компьютерной клавиатуре показала, что внедрение структурированных механизмов обратной связи значительно повышает точность прогнозирования эмоциональных состояний. Проведенное тестирование модели подтвердило эффективность данного подхода, что позволяет системе адаптироваться к изменениям в поведении пользователя и условиях окружающей среды. Созданное программное обеспечение может быть использовано для повышения адаптивности и персонализации пользовательских интерфейсов, что улучшает взаимодействие между человеком и компьютером. В будущем разработанная модель и программное обеспечение станут частью более комплексной системы по оценке психофизиологического состояния человека, что расширит их применение и углубит взаимопонимание между пользователями и технологиями.

### СПИСОК ЛИТЕРАТУРЫ

- [1] D. Gunetti, C. Picardi, and G. Ruffo, "Keystroke analysis of different languages: a case study," in Lecture Notes in Computer Science, Springer Berlin/Heidelberg, pages 133–144, 2005.
- [2] C. Epp, M. Lipages old, and R. L. Mandryk, "Identifying emotional states using keystroke dynamics," in Proc. of the SIGCHI Conference on Human Factors in Computing Systems, pages 715–724, 2011.
- [3] A. Pentel, "Emotions and user interactions with keyboard and mouse," in Proc. 8th Int. Conf. Inf., Intell., Syst. Apages I. (IISA), pages 420–425, Aug. 2017.
- [4] J. Smith, A. Johnson, and R. Brown, "Effects of stress and fatigue on keyboard input activities: An experimental study," Computers in Human Behavior, 78: 191–199, 2020.
- [5] F. Monrose and A. Rubin, "Authentication via Keystroke Dynamics," in Proceedings of the 4th ACM conference on Computer and communications security - CCS '97, 1997.
- [6] D. Gunetti, C. Picardi, and G. Ruffo, "Keystroke analysis of different languages: a case study," in Lecture Notes in Computer Science, Springer Berlin / Heidelberg, pages 133–144, 2005.
- [7] O. Guven, S. Akyokus, M. Uysal, and A. Guven, "Enhanced Password Authentication through Keystroke Typing Characteristics," in Proc. Int. Conf. on Intelligent Systems, pages 317–322, 2007.
- [8] J. Hernandez, R. Paredes, J. Wainer, and J. D. Villarroel, "Stress Detection Using Keystroke Dynamics: Towards Non-Invasive Monitoring of Computer Users," IEEE Transactions on Affective Computing, 5(3): 274–283, 2014.
- [9] Q. Li, Y. Wang, F. Liu, and B. Xu, "The influence of psychological and physiological states on typing patterns: A systematic review," International Journal of Human-Computer Interaction, 34(7): 611–625, 2018.
- [10] L. Giancardo, Á. Sánchez-Ferro, T. Arroyo-Gallego, I. Butterworth, C. S. Mendoza, P. Montero, et al., "Computer keyboard interaction as an indicator of early Parkinson's disease," Scientific Reports, 6(1):34468, 2016.
- [11] P. Zimmermann, S. Guttormsen, B. Danuser, and P. Gomez, "Affective computing - a rationale for measuring mood with mouse and keyboard," Inter. J. of Occ. Saf. and Ergo., 9(4): 539–551, 2003.
- [12] J. Li, K. Cheng, S. Wang, et al., "Feature selection: A data perspective," ACM Computing Surveys (CSUR), 50(6):1–45, 2017.
- [13] Z. Hou, Q. Hu, and W. L. Nowinski, "On minimum variance thresholding," Pattern Recognition Letters, 27(14):1732–1743, 2006.
- [14] P. M. Granitto, C. Furlanello, F. Biasioli, et al., "Recursive feature elimination with random forest for PTR-MS analysis of agroindustrial products," Chemometrics and Intelligent Laboratory Systems, 83(2):83–90, 2006.
- [15] "API documentation," OpenWeatherMap, available at <https://openweathermap.org/api>.

# Транскодирование JPEG-изображений при помощи нейросетевого предсказания коэффициентов ДКП

И. С. Подцепко, Е. А. Беляев

Национальный исследовательский университет ИТМО

i.podtsepko@niuitmo.ru

**Аннотация.** Настоящая работа посвящена вопросу дополнительного сжатия без потерь изображений, представленных в формате JPEG. Предложен новый способ кодирования коэффициентов дискретного косинусного преобразования (ДКП) при помощи нейронных сетей, который осуществляется следующим образом: на первом этапе выполняется энтропийное декодирование коэффициентов ДКП согласно алгоритму JPEG. Полученные коэффициенты разбиваются псевдослучайным образом на две группы. Коэффициенты первой группы обнуляются, после чего выполняется обратное ДКП. Полученное изображение подаётся на вход нейронной сети, которая обучена улучшать качество декодированных изображений. В области ДКП это означает, что на местах обнулённых коэффициентов первой группы появляются похожие на них коэффициенты, то есть, нейронная сеть предсказывает значения коэффициентов первой группы по коэффициентам из второй группы. Затем все коэффициенты кодируются согласно алгоритму JPEG, однако вместо значений коэффициентов первой группы кодируется ошибка предсказания, относительно предсказанных коэффициентов. Результаты экспериментов показывают, что такой способ позволяет дополнительно повысить степень сжатия файлов формата JPEG.

**Ключевые слова:** транскодирование без потерь; JPEG; дискретное косинусное преобразование; остаточное сжатие

## I. ВВЕДЕНИЕ

JPEG [1] – это стандарт сжатия изображений, который был разработан в 1992 году Joint Photographic Experts Group. До настоящего времени он остается наиболее распространенным форматом сжатия изображений из-за своей низкой вычислительной сложности. Однако качество сжатия JPEG значительно уступает более современным алгоритмам, таким как WebP, HEIF, AVIF и т. д.

В современном мире цифровых устройств фотографии, сохраненные в формате JPEG, стали неотъемлемой частью социальной жизни. С каждым днем все больше людей используют социальные сети, такие как ВКонтакте и Telegram, для обмена сообщениями и изображениями. Эти веб-сайты хранят огромное количество JPEG-изображений на своих серверах. Более того, массовое использование JPEG-изображений приводит к увеличению спроса на локальные и облачные хранилища.

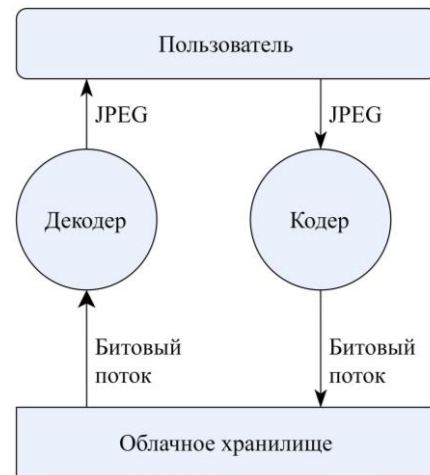


Рис. 1. Архитектура внедрения транскодера JPEG

Следовательно, исследование способов снижения затрат на хранение JPEG-изображений в облачном хранилище или на персональных компьютерах становится все более важной и актуальной задачей. При этом чаще всего речь идет именно о более эффективном хранении без потерь, так как если пользователь загружает в облачное хранилище изображение – он ожидает получить его в будущем в неизменном виде.

Для решения проблемы хранения больших объемов изображений JPEG транскодер может использоваться, как показано на рис. 1. При загрузке в облачное хранилище JPEG файл дополнительно сжимается без потерь транскодером, а при запросе от пользователя транскодируется в исходное представление.

Настоящая статья рассматривает алгоритм внутреннего предсказания ДКП-коэффициентов с помощью нейронной сети. В процессе работы предлагаемого алгоритма JPEG изображение декодируется до коэффициентов ДКП, после чего часть из них обнуляется. Такое преобразование вносит в изображения определенные искажения, которые могут исправляться специально обученной нейронной сетью. Далее на места обнуленных коэффициентов записываются ошибки их предсказания, и изображение кодируется энтропийным кодером из алгоритма JPEG. На выходе транскодера получается корректный файл JPEG, но меньшего объема. Данный подход позволяет лучше устранять избыточность, возникающую между блоками изображения, так как реализует предсказание не

только DC, но и AC-коэффициентов. Метод является совместимым с другими существующими транскодерами JPEG и в среднем позволяет добиться сжатия на 3,26 %.

Настоящая статья имеет следующую структуру. В разделе II рассматриваются существующие алгоритмы сжатия изображений: JPEG, JPEG2000 [3], HEVC [4], JPEG-LS [5] и другие. Также в этом разделе приводится краткая сводка относительно существующих на сегодняшний день подходов к транскодированию JPEG-изображений с потерями и без потерь. Раздел III содержит краткий обзор и более подробное описание предлагаемого авторами алгоритма транскодирования. В разделе IV приводятся детальная информация о том, как была обучена модель и как она тестировалась, а также результаты экспериментов. В разделе V содержатся выводы из проделанной работы, а в разделе VI – перспективы дальнейших исследований в данной области.

## II. ОБЗОР ЛИТЕРАТУРЫ

### A. Алгоритм JPEG

При сжатии алгоритмом JPEG [2] изображение проходит несколько этапов преобразования. Сначала оно переводится из цветового пространства RGB в YCbCr. Компонент Y (luminance) представляет собой яркостную информацию пикселя и определяет его интенсивность. Он кодирует черно-белую информацию изображения. Компоненты Cb (chroma blue) и Cr (chroma red) вычисляются как разности между цветом пикселя и его яркостью, то есть являются цветовыми компонентами. Преимущество такого представления состоит в том, что в силу особенностей зрительной системы человека значимость яркости превышает значимость цвета. Другими словами, наш глаз лучше фиксирует изменения именно в яркостном компоненте. Это говорит о психовизуальной избыточности. По этой причине на втором шаге алгоритма JPEG выполняется прореживание (subsampling) цветовых компонент, то есть уменьшение их размерностей.

Далее каждый из компонентов разбивается на блоки  $8 \times 8$  пикселей, которые кодируются отдельно и независимо друг от друга. К ним применяется двумерное дискретное косинусное преобразование (ДКП), в котором используются косинусные функции для разложения сигнала на сумму синусоид с разными частотами:

$$f(x, y, u, v) = a(u)a(v) \cos\left(\frac{(2x+1)u\pi}{2N}\right) \cos\left(\frac{(2y+1)v\pi}{2N}\right),$$

где

$$a(u) = \begin{cases} \frac{1}{\sqrt{N}}, & u = 0 \\ \frac{2}{\sqrt{N}}, & u \neq 0. \end{cases}$$

Коэффициенты, полученные после ДКП, разделяют на DC и AC. DC-коэффициент или коэффициент постоянной составляющей, находящийся в верхнем левом углу блока, представляет собой компонент, отвечающий за среднее значение яркости или интенсивности цвета изображения. AC-коэффициенты – это высокочастотные компоненты блока.

На следующем этапе работы алгоритма JPEG коэффициенты ДКП поэлементно делятся с округлением вниз на матрицы квантования, которые, как правило, разные для разных каналов изображения. Данный этап называется квантованием. Абсолютные значения элементов матриц квантования зависят от параметра качества (quality factor, QF), передаваемого кодеру. Часть коэффициентов становятся нулями, а оставшиеся кодируются энтропийным кодером, сначала с помощью кодирования длин серий (run-level encoding, RLE), а затем кодом Хаффмана. Абсолютные значения коэффициентов записываются в выходной битовый поток с помощью прямого кода.

Необходимо отметить, что вместо непосредственного кодирования DC-коэффициентов кодируются разность между квантованным текущим и предыдущим DC-коэффициентами. Данный подход называется кодированием с предсказанием. Его преимущество состоит в том, что он позволяет устранять избыточность между соседними блоками, которая возникает из-за того, что в реальных изображениях можно наблюдать самоподобие.

### B. Эффективные методы сжатия изображений

На сегодняшний день в области сжатия изображений без потерь есть большие успехи. Так, например, следующие форматы позволяют достичь большей степени сжатия без потерь при заданном качестве картинки: JPEG2000, основанный на вейвлет-преобразовании и арифметическом кодировании; BPG [6], основанный на HEVC; JPEG-LS, CALIC [7], FLIF [8] применяющие контекстные модели для использования высокоуровневой корреляции; а также ряд других алгоритмов, таких как HEVC, которые исследуют более сложные методы внутреннего предсказания.

К сожалению, вышеперечисленные форматы, во-первых, не используются также широко как JPEG из-за своей высокой сложности, во-вторых, являются неэффективными для транскодирования JPEG-изображений без потерь. Связано это в частности с округлением значений Y, Cb, Cr и коэффициентов ДКП при кодировании и ограничении значений R, G, B при декодировании. Однако, необходимо обратить внимание на то, что все эти подходы основаны на пространственной корреляции, наблюдаемой между соседними участками изображений.

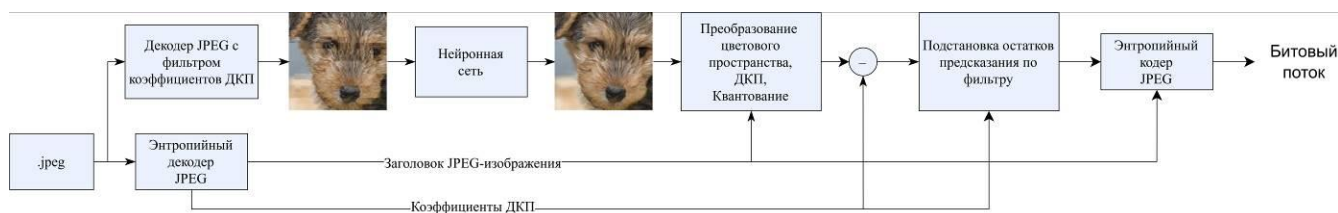


Рис. 1. Общая схема работы транскодера.



Рис. 2. Общая схема работы транскодера

### С. Транскодирование JPEG-изображений с потерями

Возможность дополнительного сжатия именно JPEG-изображений выглядит привлекательной, причем это может быть как транскодирование с потерями, так и без. Первую задачу успешно выполняют такие программы как JPEG TinyPNG [9], Mozjpeg [10] и Guetzli [11]. Они достигают высокой степени сжатия за счет объединения цветов, изменения порядка сканирования и использования особенностей человеческого зрения. Однако во многих прикладных сценариях сжатие с потерями недопустимо.

### Д. Транскодирование JPEG-изображений без потерь

Утилиты, реализующие транскодирование без потерь, также существуют. Одним из самых ранних и распространенных решений является Jpegtran. Данная программа выполняет ряд дополнительных преобразований изображения, таких как поворот блоков, таблиц квантования и т. д. Однако задача является открытой и исследования в этой области продолжаются. Так, одним из наиболее эффективных транскодеров стал LLJPEG [12], разработанный в 2023 г. и являющийся гибридным алгоритмом транскодирования JPEG без потерь, объединяющий сразу несколько эффективных подходов к кодированию изображений. Во-первых, данный метод отказывается от фиксированного размера блока, который итеративно вычисляется для каждого участка изображения. Начиная со значения  $64 \times 64$ , на каждой итерации вычисляется функция стоимости кодирования блока и пока это является эффективным – блок уменьшается. Минимальный допустимый размер блока –  $4 \times 4$ . Во-вторых, LLJPEG использует внутреннее предсказание коэффициентов ДКП аналогичное внутреннему предсказанию HEVC, где для вычисления значений в блоке используется наиболее эффективное направление из 35-и, указывающих на блоки слева и сверху. В-третьих, код Хаффмана заменяется более эффективным – контекстно-адаптивным бинарным арифметическим кодером (context-adaptive binary arithmetic coding, CABAC).

Благодаря объединению вышеперечисленных техник эффективность сжатия LLJPEG достигает от 18 % до

50 %, что на сегодняшний день является наилучшими показателями.

## III. ПРЕДЛАГАЕМЫЙ АЛГОРИТМ

### А. Мотивация

Эффективность сжатия зависит от нескольких факторов. С одной стороны, можно оптимизировать энтропийный кодер. Например, строить оптимальные таблицы Хаффмана для конкретного изображения двухпроходным алгоритмом, а не использовать универсальные, либо вовсе отказываться от кода Хаффмана в пользу арифметического кода. С другой стороны, внутреннее предсказание позволяет уменьшать количество ненулевых коэффициентов, а также их абсолютные значения. Это также положительно сказывается на степени сжатия, так как серии из нулей кодируются эффективно благодаря применению RLE, а меньшим по модулю коэффициентам обычно соответствуют более короткие кодовые слова.

Как было упомянуто выше, задача внутреннего предсказания уже эффективно решается. Однако для предсказания всегда используются конечные алгоритмы, которые обычно ищут зависимости между кодируемым блоком одним или несколькими соседними блоками из того же компонента.

В основе настоящей работы лежит идея о том, что для предсказания коэффициентов ДКП можно использовать нейронную сеть, так как она способна искать более сложные взаимосвязи в изображении, например, между:

- блоком и его соседями в любом из направлений;
- компонентом Y и компонентами Cr, Cb;
- удаленными участками изображения.

### В. Обзор алгоритма

В настоящей статье предлагается новый метод внутреннего предсказания коэффициентов ДКП и транскодер, основанный на данном методе внутреннего предсказания. На рис. 2 представлена общая схема транскодера.

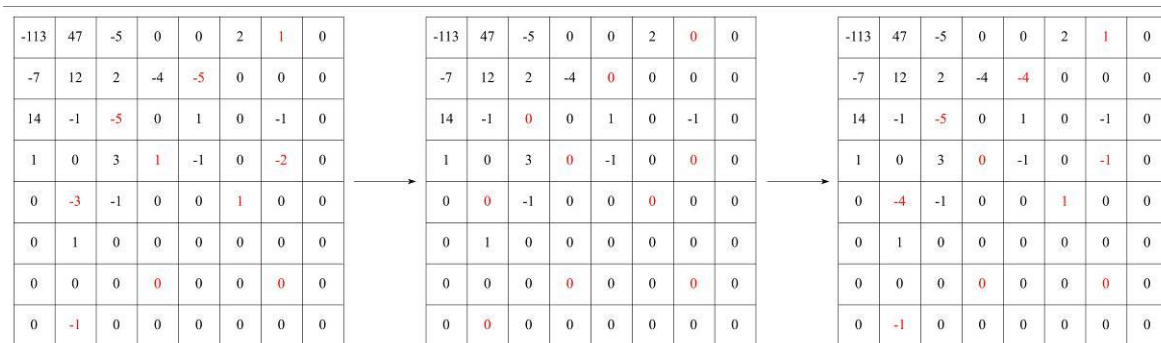


Рис. 3. Пример обнуления и предсказания блока изображения: слева – исходный блок, посередине – блок после обнуления, справа – блок после обработки изображения нейронной сетью

При кодировании из исходного JPEG-изображения извлекаются коэффициенты ДКП после применения энтропийного декодера JPEG. Далее полученные коэффициенты разбиваются псевдослучайным образом на две группы и коэффициенты первой группы обнуляются. Пример такого преобразования представлен на рис. 4.

В результате обнуления в изображение вносится искажение. В блоке становится меньше ненулевых коэффициентов, и он может быть более компактно представлен при кодировании энтропийным кодером. Основная гипотеза настоящей работы состоит в том, что нейронная сеть может компенсировать такие искажения и восстанавливать исходное изображение.

После обнуления блока к оставшимся коэффициентам применяется обратное квантование и обратное ДКП, и изображение переводится в цветовое пространство RGB. Далее оно подается на вход нейронной сети, которая обучена улучшать качество декодированных изображений. В области ДКП это означает, что на местах обнуленных коэффициентов первой группы появляются похожие на них коэффициенты (см. рис. 4), то есть, нейронная сеть предсказывает значения коэффициентов первой группы по коэффициентам из второй группы.

Затем все коэффициенты кодируются согласно алгоритму JPEG, однако вместо значений коэффициентов первой группы кодируется ошибка предсказания, относительно предсказанных коэффициентов.

Если нейронная сеть в точности предсказывает коэффициент, то вместо него в блоке остается 0, в противном случае — он заменяется меньшим по модулю коэффициентом. Таким образом, достигается дополнительное сжатие JPEG-изображения.

Схема трансдекодера представлена на рис. 3 и аналогична схеме транскодера. При декодировании создается изображение, у которого удалены те же самые коэффициенты, что и при кодировании. Оно подается на вход нейронной сети и на месте выколотых коэффициентов появляются значения, к которым остается только добавить поправку, закодированную у мест удаленных коэффициентах.

C. Фильтрация коэффициентов

Фильтрация коэффициентов реализована детерминировано, но наборы коэффициентов для каждого блока являются псевдослучайными. Такая реализация следует из следующих соображений. Во-первых, фильтр (список удаляемых коэффициентов) требуется знать при декодировании, а значит, он должен быть всегда определен. Во-вторых, если удалять одни и те же коэффициенты в соседних блоках, то некоторая часть информации может быть утеряна. Если фильтры соседних блоков существенно отличаются друг от друга — восстановление коэффициентов в блоке может происходить благодаря заимствованию информации о них из соседних блоков.

Необходимо также отметить тот факт, что фильтруются только коэффициенты яркостной компоненты, так как в стандарте JPEG и так предусмотрено существенное сокращение вклада цветовых компонент в итоговый объем файла благодаря прореживанию.

D. Внутреннее предсказание коэффициентов

Для внутреннего предсказания была реализована и обучена нейронная сеть с архитектурой QE-CNN (quality enhancement convolution neural network), представленной на рис. 5. Каждый ее слой состоит из свертки, сохраняющей размерность тензора, и функции-активации PReLU с количеством параметров равном количеству фильтров в свертке.

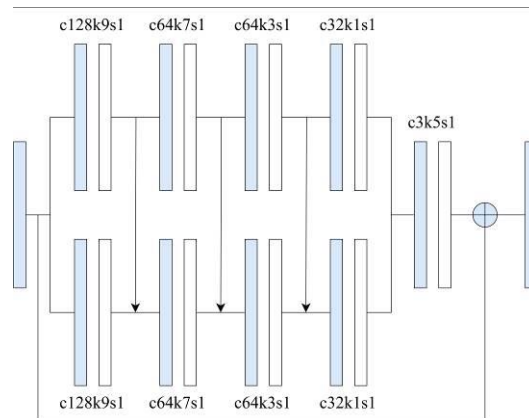


Рис. 4. Архитектура нейронной сети



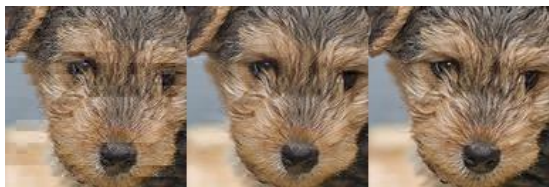


Рис. 5. Пример работы QE-CNN: слева – изображение после фильтрации ДКП коэффициентов, посередине – изображение на выходе нейронной сети, справа – исходное изображение

#### IV. РЕЗУЛЬТАТЫ

Для обучения нейронной сети было взято 50'000 изображений из набора данных ImageNET, из которых 5'000 было выделено для тестирования модели, 10'000 – для валидации и оставшиеся 35'000 – для обучения. Изображения были обрезаны до размера 320×320, затем разбиты на блоки 8×8 пикселей, в каждом из которых были удалены 16 псевдослучайных коэффициентов АС-коэффициентов, как это было описано в предыдущем разделе.

Обучение производилось на основании функции ошибки MSE в пространстве YCbCr после нормализации с использованием оптимизатора Adam с параметром learning rate равным  $1e-4$  на протяжении 100 эпох. После каждой эпохи рассчитывалась метрика MSE на валидационном наборе данных. Наименьшее значение функции ошибки было достигнуто по завершении последней эпохи и составило  $1,405e-3$ . Пример работы нейронной сети представлен на рис. 6.

Оценка качества работы транскодера выполнялась на упомянутом выше тестовом наборе данных. Результаты эксперимента показывают, что разработанный метод транскодирования позволяет сократить размер JPEG изображения в среднем на 3.26 % при медиане 3.33 %. Наилучшее достигнутое сжатие составляет 6.05 %.

#### V. ЗАКЛЮЧЕНИЕ

Для оптимизации облачного хранения изображений, закодированных в формате JPEG, был предложен алгоритм транскодирования без потерь, основанный на внутреннем предсказании коэффициентов ДКП на основе нейронной сети. Эксперименты показывают, что такой подход позволяет сократить до 6% памяти на изображениях из подмножества набора данных ImageNET.

Несмотря на то, что альтернативные системы транскодирования JPEG-изображений, такие как LLJPEG, позволяют добиться более высокого сжатия, данный подход имеет место быть и нуждается в дальнейшем исследовании, так как в результате его работы получается корректный файл JPEG. Это означает,

что данный метод может быть использован одновременно с другими методами транскодирования. К тому же, предлагаемый алгоритм затрагивает исключительно этап внутреннего предсказания, а значит, он открыт к дальнейшему улучшению, например, за счет отказа от фиксированного размера блока и использования арифметического кодера.

#### VI. ДАЛЬНЕЙШИЕ ПЕРСПЕКТИВЫ ИССЛЕДОВАНИЯ

В рамках дальнейшей работы над нейросетевым транскодером планируется, во-первых, продолжать обучение нейронной сети. Во-вторых, оценить степень сжатия для других мощностей фильтра коэффициентов ДКП – то есть при удалении меньшего и большего количества АС-коэффициентов. В-третьих, разработать гибридную систему транскодирования изображений аналогично LLJPEG, добавив оптимизации алгоритма JPEG в другие этапы его работы, такие как энтропийное кодирование и деление изображения на блоки.

#### СПИСОК ЛИТЕРАТУРЫ

- [1] G. Wallace, "The jpeg still picture compression standard," IEEE Transactions on Consumer Electronics, vol. 38, no. 1, pp. xviii–xxxiv, 1992.
- [2] G.K. Wallace, "Overview of the JPEG (ISO/CCITT) still image compression standard," Proc. SPIE, vol. 1244, pp. 220–233, Jun. 1990.
- [3] C. Christopoulos, A. Skodras, and T. Ebrahimi, "The jpeg2000 still image coding system: an overview," IEEE Transactions on Consumer Electronics, vol. 46, no. 4, pp. 1103–1127, 2000.
- [4] G.J. Sullivan, J.-R. Ohm, W.-J. Han, and T. Wiegand, "Overview of the high efficiency video coding (HEVC) standard," IEEE Trans. Circuits Syst. Video Technol., vol. 22, no. 12, pp. 1649–1668, Dec. 2012.
- [5] M.J. Weinberger, G. Seroussi, and G. Sapiro, "The LOCO-I lossless image compression algorithm: Principles and standardization into JPEG-LS," IEEE Trans. Image Process., vol. 9, no. 8, pp. 1309–1324, Aug. 2000.
- [6] F. Bellard. (2015). BPG Image Format Website. [Электронный ресурс]. URL: <http://bellard.org/bpg/>
- [7] X. Wu and N. Memon, "CALIC—A context based adaptive lossless image codec," in Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP), vol. 4, May 1996, pp. 1890–1893.
- [8] J. Sneyers. (2021). FLIF Image Format Website. [Электронный ресурс]. URL: <http://flif.info/>
- [9] A. Kyleduo. (2017). Tinypng Project. [Электронный ресурс]. URL: <https://tinypng.com/developers>
- [10] Mozilla. (2019). Mozilla Mozjpeg. [Электронный ресурс]. URL: <https://github.com/mozilla/mozjpeg>
- [11] J. Alakuijala, R. Obryk, O. Stoliarchuk, Z. Szabadka, L. Vandevenne, and J. Wassenberg, "Guetzli: Perceptually guided JPEG encoder," 2017, arXiv:1703.04421.
- [12] C. Sun, X. Fan and D. Zhao, "Lossless Recompression of JPEG Images Using Transform Domain Intra Prediction," in IEEE Transactions on Image Processing, vol. 32, pp. 88-99, 2023.

# Применение инструментов машинного обучения в системах автоматизированного проектирования технологических процессов (САПР ТП)

А. А. Лаптев<sup>1</sup>, С. Д. Третьяков

Национальный исследовательский университет ИТМО

<sup>1</sup>laptev.aleksandr@rambler.ru

**Аннотация.** В данной работе освещены различные методы проектирования технологических процессов, проведен их сравнительный анализ и выделены общие этапы. Особое внимание было уделено анализу решений в подсистеме информационного поиска, где реализуется поиск деталей-аналогов в существующих системах автоматизированного проектирования технологических процессов (САПР ТП). На основе проведенного анализа предложен новый подход к поиску деталей-аналогов и унификации деталей с применением инструментов, и алгоритмов машинного обучения. Также предложено использование нового формата базы данных в САПР ТП, что позволяет существенно упростить процесс поиска и унификации деталей. А также повысить эффективность поиска за счет сокращения времени и качественных результатов. Полученные результаты и предложения могут стать основой для улучшения процедуры проектирования технологических процессов в промышленности и повышения эффективности использования систем автоматизированного проектирования технологических процессов.

**Ключевые слова:** технологические процессы; технологическая подготовка производства; машинное обучение; поиск деталей-аналогов; маршруты обработки деталей; векторная база данных; автоматизация

## I. ОБЗОР ЛИТЕРАТУРЫ

На современном этапе порядка 70–80 % изделий машиностроения производится малыми партиями на предприятиях единичного и мелкосерийного типа производства [1]. Современные машиностроительные предприятия характеризуются сложной структурой движения потоков (технологических этапов), управленческого состава, многообразием оборудования и технологических процессов, большой номенклатурой выпускаемых изделий.

С повышением количества выпускаемой номенклатуры изделий, а также частым изменением производимых изделий, совместно с увеличением запросов на уменьшение производственного цикла, остро ставится вопрос о повышении эффективности этапа технологической подготовки производства. Очевидно, что повышение эффективности этапа технологической подготовки производства является основополагающим инструментом, благодаря которому можно повысить эффективность всего предприятия в

целом. Одним из способов достичь данной цели является внедрение и использование современных подходов [2, 3].

Технологическая подготовка производства (ТПП) является неотъемлемой частью производственного процесса, от которой в значительной степени зависит его эффективность. Согласно ГОСТ Р 50995.3.1-96 одной из задач ТПП является своевременное обеспечение производства качественными технологическими процессами, материалами, комплектующими изделиями, средствами технологического оснащения на основе использования, при их создании или приобретении, информационных массивов описаний конструкторско-технологических решений. Очевидно, что на сегодняшний день, для выполнения данной задачи используются автоматизированные системы технологической подготовки производства (АСТПП), однако необходимо отметить, что технологические процессы согласно ГОСТ 23501.108 разрабатываются именно в системах автоматизированного проектирования технологических процессов (САПР ТП).

Для выявления недочетов в разработке технологических процессов в системе автоматизированного проектирования технологических процессов (САПР ТП) рассмотрим основные методы, применяемые при проектировании технологических процессов. В настоящее время существует несколько основных подходов к проектированию технологических процессов. [4, 5, 6]:

- индивидуальное проектирование;
- проектирование на основе технологического процесса-аналога (деталь-аналог);
- типовой;
- групповой.

При необходимости проведения индивидуального проектирования технологических процессов в случае отсутствия аналогичных, происходит создание маршрута изготовления изделия и определение последовательности операций исходя из уникальных конструктивно-технологических характеристик самого изделия и возможностей применяемой производственной системы.

Процесс автоматизированного проектирования на основе аналогичных технологических процессов является многоплановым, включает различные ключевые этапы. При выборе подходящего аналогичного технологического процесса важно учитывать его высокий уровень сертификации качества, что позволяет снизить расходы на технологическое проектирование и производство.

Автоматизированного проектирования на базе технологических процессов-аналогов включают в себя следующие этапы:

- Выбор продукта для автоматизированного проектирования технологических процессов, включающий анализ конструкции изделия и его требований к производству и эксплуатации, а также отбор необходимых изделий и деталей для дальнейшего проектирования.
- Подготовка исходных данных для автоматизированного проектирования, включающая подготовку технической документации для кодирования и заполнение соответствующих полей с исходными данными.
- Генерация конструктивно-технологического кода изделия в зависимости от уровня автоматизации проектирования технологических процессов, осуществляется либо на этапе подготовки исходных данных, либо в начале выполнения задачи на компьютере.
- Поиск аналогичных технологических процессов, включающий сопоставление конструктивно-технологических кодов изделий с подобными продуктами, сохраненными в поисковых системах.
- Переработка выбранного аналогичного технологического процесса, включая разработку технологического маршрута, операций, расчет режимов обработки и норм времени, а также создание технологической документации.

В свою очередь типовой и групповой технологический процесс согласно работам Митрофанова С.П. и Соколовским А.П. базируется на предварительно унифицированных группах по конструктивным или конструктивно-технологическим признакам соответственно. После «ручного» процесса создания унифицированных групп, технологю необходимо создать или выбрать деталь прототип для типового технологического процесса и комплексную деталь для группового технологического.

Однако необходимо обратить внимание на то, что каждый их перечисленных выше методов проектирования технологических процессов включает в себя этап поиска детали-аналога, комплексной детали или детали представителя.

Согласно [7–15] в существующих системах САПР ТП поиск детали осуществляется за счет «ручного» или частично автоматизированного ввода определенных атрибутов в поисковую систему. В качестве атрибутов могут выступать габариты детали, материал, размеры

заготовки, наименования и т. д., каждый поставщик САПР ТП определяет свои атрибуты для поисков деталей в базе данных (БД).

Соответственно сегодняшнюю общую методику проектирования технологического процесса в САПР ТП можно представить следующим образом, как показано на рис. 1.

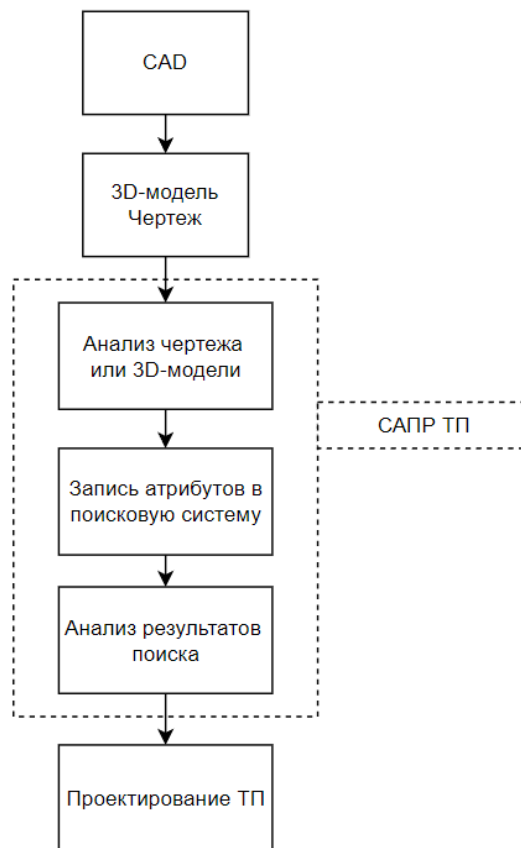


Рис. 1. Общая методика проектирования технологического процесса в САПР ТП

## II. ОСНОВНАЯ ЧАСТЬ

Согласно приведенному в первой главе анализу можно выделить следующие недостатки существующих решений в САПР ТП:

- «ручной» процесс ввода атрибутов;
- трудоемкий процесс анализа заказной детали и не менее трудоемкий процесс сравнения заказной детали с деталями в БД;
- исключительное использование реляционных и не реляционных баз данных для хранения и обработки информации.

Данные недостатки возможно компенсировать за счет применения в существующих САПР ТП инструментов и алгоритмов машинного обучения, а именно нейронных сетей. В некоторых исследованиях предлагается использовать методы машинного обучения для классификации деталей, рассмотренных ранее. Это свидетельствует о новаторском подходе к применению инструментов и алгоритмов машинного обучения в

задаче классификации деталей для их последующего поиска. Тем не менее, на сегодняшний день отсутствуют конкретные решения данной проблемы. Так в следующих работах [16, 17] авторы рассматривают важность использования машинного обучения для улучшения производительности и эффективности обработки запросов в современных системах управления базами данных (СУБД).

Соответственно при использовании нейронных сетей при поиске деталей в САПР ТП, общая методика проектирования технологических процессов будет выглядеть следующим образом:

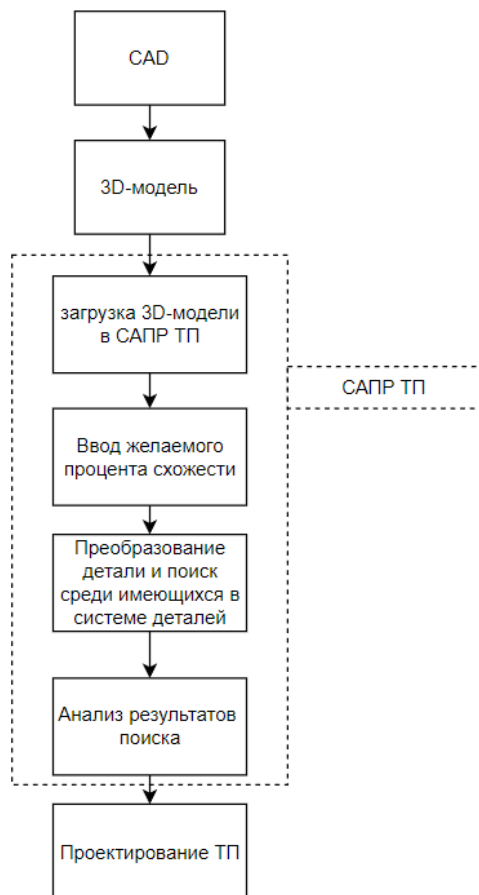


Рис. 2. Общая методика проектирования технологического процесса в САПР ТП с инструментами машинного обучения

Вследствие разработанной методики была обучена нейронная сеть на базе TensorFlow, с применением 20 моделей деталей в формате .stp и проведен эксперимент со сравнением деталей.

В обучающей выборке была модель детали, изображенной на рис. 3.

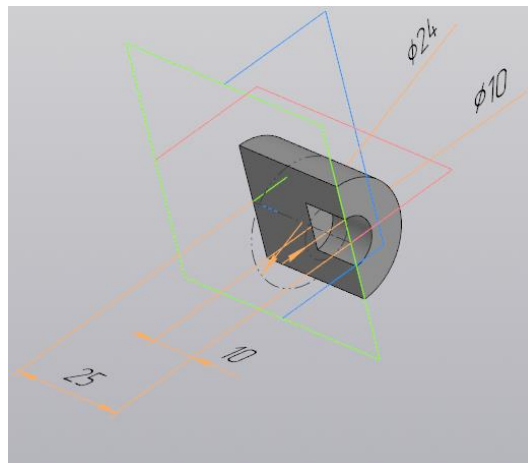


Рис. 3. Модель «А»

Тестовая модель изображена на рис. 4.

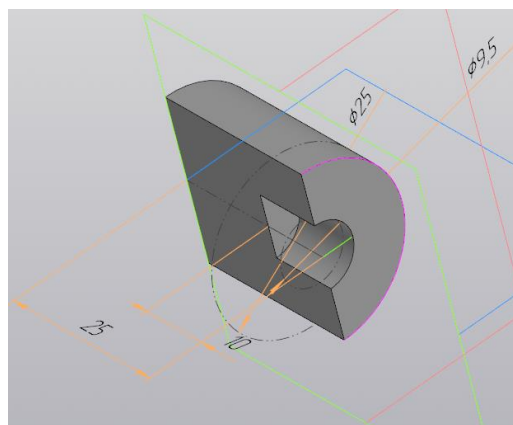


Рис. 4. Модель «В»

По результатам поиска были получены следующие результаты предсказаний с использованием обученной модели, представленные на рис. 5

```

1/1 0s.71ms/step
[[1.3697991e-07 9.9989105e+01 1.0869582e-06 1.0895673e-02 6.2170166e-08]]
  
```

Рис. 5.

Были выведены самые похожие модели и схожесть модели «А» с моделью «В» получилась 99 %. Так что полученные результаты можно оценить, как правдивые, так как данные модели очень похожи.

### III. ЗАКЛЮЧЕНИЕ

В заключение статьи можно отметить, что проведенный анализ выявил недостатки существующих решений в системах автоматизированного проектирования технологических процессов (САПР ТП), такие как «ручной» процесс ввода атрибутов, трудоемкий анализ и сравнение деталей, а также ограниченное использование баз данных. Для компенсации этих недостатков предлагается использовать инструменты и алгоритмы машинного обучения, включая нейронные сети. Применение подобных технологий может значительно улучшить процесс проектирования и оптимизации технологических процессов, сократить время на

разработку проектов и повысить их качество. Использование машинного обучения позволит автоматизировать рутинные задачи, увеличить точность анализа данных и сэкономить усилия инженеров. Таким образом, интеграция инновационных подходов в САПР ТП способствует повышению эффективности проектирования и снижению вероятности ошибок в процессе разработки технологических процессов.

#### IV. ДАЛЬНЕЙШИЕ ПЕРСПЕКТИВЫ ИССЛЕДОВАНИЯ

Дальнейшие перспективы данного исследования связаны с разработкой и усовершенствованием методов машинного обучения для поиска и классификации деталей в САПР ТП. Кроме того, возможно расширение функциональности систем за счет внедрения новых подходов и технологий, что позволит значительно улучшить процесс проектирования технологических процессов в промышленности и повысить общую эффективность использования автоматизированных систем.

#### СПИСОК ЛИТЕРАТУРЫ

- [1] Хрусталева И.Н., Любомудров С.А., Романов П.И. Автоматизация технологической подготовки единичного и мелкосерийного производства // Научно-технические ведомости СПбПУ. Естественные и инженерные науки. 2018. Т. 24, № 1. С. 113–121. DOI: 10.18721/JEST.240111.
- [2] Селиванов Сергей Григорьевич, Никитин Виталий Викторович, Дружинина Марина Александровна, Шпилова Влада Георгиевна Автоматизированная система технологической подготовки бережливого производства в машиностроении // Вестник УГАТУ = Vestnik UGATU. 2009. №1.
- [3] Колыбенко Е.Н. Принципы подхода к определению технологической элементной базы средств автоматизации для подготовки производства меха-нообработки по всему ее циклу продолжение // Справочник. Инженерный журнал. 2021. № 12 (297). С. 20–27. DOI: 10.14489/hb.2021.12.pp.020-027
- [4] Фролова И.Н., Кутилова О.И. Анализ современных систем автоматизированного проектирования технологических процессов (САПР ТП) // Труды НГТУ им. Р.Е. Алексеева. 2010. №1.
- [5] Методики автоматизированного проектирования технологических процессов // Аспект URL: <https://na-journal.ru/12-2023-tehnologii-proizvodstva/7920-metodiki-avtomatizirovannogo-proektirovaniya-tehnologicheskikh-processov> (дата обращения: 20.02.2024).
- [6] Антоненкова Т.В. Способ формального проектирования структуры технологических процессов // МНИЖ. 2023. №3 (129).
- [7] Жетесова Г.С., Юрченко В.В., Никонова Т.Ю., Кожанов М.Г., Иванов С.С., Кибико А.С. Исследование и анализ САПР ТП машиностроительного производства // Science conference CAD/EDA/Simulation in electronics: Международная научно-практическая конференция "САПР и моделирование в электронике" 2017-2021 г. 2018. №5. С. 10-19.
- [8] Бехметьев Вячеслав Иванович Исследование прикладных особенностей системы автоматизированного проектирования технологических процессов «Вертикаль» в создании эффективных технологий изготовления деталей самолетов // Наука без границ. 2020. №3 (43).
- [9] Зотова Д.О. Оценка целесообразности внедрения проекта САПР ТП Вертикаль на предприятии РКЦ «Прогресс» при помощи структуры TELOS [Электронный ресурс]: сборник тезисов докладов на конференции. // Научный диалог: экономика и управление – Чебоксары: Центр научного сотрудничества «Интерактив плюс». Режим доступа: <https://doi.org/10.21661/r-486256> (дата обращения: 01.02.2024). – DOI 10.21661/r-486256.
- [10] САПР ТП ВЕРТИКАЛЬ: технологию проектировать просто! // САПР и графика URL: <https://sapr.ru/article/16074> (дата обращения: 18.02.2024).
- [11] T-FLEX Технология — от САПР ТП к комплексной системе технологической подготовки производства // ТОП СИСТЕМЫ URL: <https://www.tfex.ru/about/publications/detail/index.php?ID=4037> (дата обращения: 02.02.2023).
- [12] Петухов Александр. Цифровая трансформация проектирования технологических процессов при подготовке инженеров-проектировщиков на базе комплекса T-FLEX PLM // Опыт использования технологий. 2022. №8. С. 52-59.
- [13] Система автоматизации технологического проектирования и подготовки производства «ТехноПро» версия 7. Книга 3. Руководство по автоматическому проектированию.
- [14] Компьютерная и программная инженерия. Международная научно-техническая конференция студентов, аспирантов и молодых ученых. 15-16 декабря 2015 года. Донецк, ДонНТУ. С. 90-95.
- [15] Ахметшин А.С. Возможности программной среды СПРУТ-ТП // 8 Всероссийская научно-практическая конференция молодых ученых с международным участием "Россия молодая". 2016. №16. С. 18-22.
- [16] Коновалов Г.Г. Применение машинного обучения для оптимизации запросов в системах управления базами данных // Международный журнал гуманитарных и естественных наук. 2023. №10-2 (85).
- [17] Дулин Сергей К., Рябцев Антон Б. Анализ подходов к оптимизации запросов в аналитических СУБД // Информационные технологии. 2023. № 3 (44). С. 73-80.

# Интеллектуальный анализ и прогнозирование динамики промышленного развития регионов России

Е. С. Митяков<sup>1</sup>, А. И. Ладынин<sup>2</sup>

МИРЭА – Российский технологический университет

<sup>1</sup>iyao@mail.ru, <sup>2</sup>andrey.ladynin@hotmail.com

С. Н. Митяков<sup>3</sup>, Т. М. Крюкова<sup>4</sup>,  
Е. А. Назарова<sup>5</sup>

Нижегородский государственный технический университет им. П.Е. Алексеева

<sup>3</sup>snmit@mail.ru, <sup>4</sup>kryukova-tamara@rambler.ru,  
<sup>5</sup>k-romanova@list.ru

**Аннотация.** В работе изложена одна из возможных моделей кластеризации субъектов Федерации на основе данных об их уровне промышленного развития. Решение задачи кластеризации осуществлено с использованием математического подхода иерархической кластеризации. Для определения пространства признаков была разработана система показателей, основанная на данных официальных статистических источников, включающая четыре индикатора, характеризующих промышленное производство в России. Для оценки и прогнозирования в ходе исследования был проведен кластерный анализ регионов согласно предложенной системе индикаторов и построены модели прогнозов индекса промышленного производства для регионов Приволжского и Центрального федеральных округов. Кластеризация проводилась по методу Варда, позволяющего сформировать иерархически вложенные кластеры. Для прогнозирования использовались регрессионные модели и методы анализа временных рядов, показавшие схожую динамику: прослеживается рост промышленного производства в соответствии с получаемыми прогнозными значениями.

**Ключевые слова:** машинное обучение, кластерный анализ, иерархическая кластеризация, прогнозирование промышленного развития, регрессионный анализ, экспоненциальное сглаживание

## I. ВВЕДЕНИЕ

Современная стратегия развития субъектов Федерации характеризуется разнообразием в уровне их развития: присутствуют значительные разрывы в финансовом обеспечении, социально-экономическом потенциале, промышленном развитии, плотности населения, природных ресурсах и других аспектах [1]. Только комплексный анализ всех факторов во взаимосвязи позволит сформировать объективное представление об экономической целесообразности и технических возможностях применения новой техники или оборудования в конкретных условиях. Необходимость создания новых подходов к промышленному развитию в региональных экосистемах обусловлена формированием эффективных методов управления промышленностью в субъектах Федерации.

Для обеспечения промышленного развития в региональных экосистемах следует разработать методологические основы, используя междисциплинарный подход с применением современных технологий управления, информационных технологий и элементов экономико-математического моделирования. В свете различий в инфраструктуре и уровне развития отраслей экономики между субъектами Федерации, важным кажется проведение информационно-статистического анализа процессов промышленного развития в регионах России. Актуальный научный дискурс включает инструментальный анализ статистических данных с применением математического моделирования [2]. Для информационной поддержки принимаемых решений эффективно применяется технология кластерного анализа, которая успешно применяется в областях с высокой степенью ответственности, таких как региональная политика или стратегическое планирование [3]. Цель настоящего исследования заключается в осуществлении экономико-математического анализа развития промышленности субъектов Российской Федерации с применением методов иерархической кластеризации и построения прогнозных моделей на основе машинного обучения и инструментария анализа временных рядов.

## II. ПОСТАНОВКА ЗАДАЧИ

Представим формализованную постановку задачи кластеризации. В контексте данной задачи конечным результатом кластерного анализа выступают группы субъектов Федерации, которые схожи по уровню промышленного развития. При этом необходимо предоставить экономическую интерпретацию для каждого кластера с учетом особенностей промышленности в соответствующем регионе. Суть задачи заключается в разделении изначальной выборки регионов на  $k$  непересекающихся кластеров  $Y = (y_1, \dots, y_k)$ . Каждый кластер должен объединять регионы, наиболее близкие по расстоянию  $r$ . Одновременно с этим регионы из разных кластеров должны значительно отличаться. Каждому анализируемому объекту  $x_n$  присваивается метка кластера  $y_k$ . Задача выполняется путем применения некоторого алгоритма кластеризации:

$$F : X \rightarrow Y.$$

Исследование выполнено за счет гранта Российского научного фонда (проект № 23-78-10009).

Таким образом, кластерный анализ нацелен на разделение данных на группы. При этом элементы одной группы должны быть максимально схожими между собой и, в то же время, обеспечить максимальное различие между элементами разных групп [4]. Для поиска решения задачи кластеризации субъектов РФ по уровню промышленного развития предлагается использовать следующий укрупненный алгоритм:

1. формирование множества регионов России, подлежащих кластерному анализу;
2. разработка системы показателей промышленного развития субъекта Федерации;
3. приведение показателей к сопоставимым единицам и пределам измерения (нормирование);
4. выбор алгоритма кластерного анализа;
5. произведение расчётов;
6. интерпретация полученных результатов.

Множеством объектов исследования в данной работе выступает перечень субъектов Федерации. Далее рассмотрим пункты 2–6 алгоритма, представленного выше.

### III. СИСТЕМА ПОКАЗАТЕЛЕЙ ДЛЯ КЛАСТЕРИЗАЦИИ РЕГИОНОВ ПО УРОВНЮ ПРОМЫШЛЕННОГО РАЗВИТИЯ

Для создания системы показателей, используемой при кластеризации субъектов Федерации по уровню промышленного развития, были выбраны статистические данные, опубликованные на официальном сайте Федеральной службы государственной статистики [5]. После этого эти данные были преобразованы в индикаторы промышленного развития (табл. 1).

ТАБЛИЦА I Показатели промышленного развития регионов

Обозначение показателя	Наименование показателя
K <sub>1</sub>	Рентабельность продаж в обрабатывающих производствах, %;
K <sub>2</sub>	Отношение объема промышленного производства в обрабатывающих отраслях к ВРП
K <sub>3</sub>	Отношение инвестиций в основной капитал в обрабатывающих производствах к ВРП
K <sub>4</sub>	Отношение среднегодового числа занятых в обрабатывающих производствах к численности населения в регионе

Первый показатель (K<sub>1</sub>) является важным финансовым индикатором, который позволяет оценить эффективность производственной деятельности в регионе. Чем выше рентабельность продаж, тем лучше для промышленности региона, поскольку это свидетельствует о том, что он умеет эффективно управлять своими расходами и обеспечивает высокий уровень прибыли от каждого реализованного товара или услуги.

Отношение объема промышленного производства в обрабатывающих отраслях к валовому региональному продукту (K<sub>2</sub>) помогает оценить вклад обрабатывающих отраслей в общую экономическую активность и развитие региона. Показатель может быть полезен для оценки степени промышленной развитости региона, его

экономической структуры и влияния промышленности на общий уровень экономического развития. Отношение инвестиций в основной капитал в обрабатывающих производствах к валовому региональному продукту (K<sub>3</sub>) показывает степень вложений в промышленность. Большое значение индикатора свидетельствует о стратегическом внимании к промышленному сектору. Отношение среднегодового числа занятых в обрабатывающих производствах к численности населения в регионе (K<sub>4</sub>) служит индикатором важности обрабатывающих производств в субъекте Федерации и показывает влияние промышленности на рынок труда.

В различных научных публикациях представлено многообразие алгоритмов кластеризации. Обычно их можно разделить на плоские, которые создают одно разбиение исходного множества объектов на кластеры, и иерархические, позволяющие формировать систему вложенных разбиений [6]. Важно отметить, что на данный момент нет универсального критерия для оценки качества кластеризации. При использовании различных алгоритмов кластерного анализа для одного набора объектов результаты могут значительно различаться, иногда принципиально отличаясь друг от друга.

В нашей задаче исходное определение количества кластеров представляет собой сложную задачу. В таких ситуациях часто применяют иерархические кластерные процедуры, которые удобно визуализировать с использованием дендрограмм, где отображается пошаговое объединение или разделение объектов в кластерах, а также метрика их близости [7].

Для построения дендрограммы необходимо определить матрицу сходства (или различия) между парами кластеров, используя выбранный критерий кластерной связи. Этот критерий определяет метод расчета расстояния между кластерами, применяемого при их объединении в ходе построения иерархии. В литературе и практике часто используются методы одиночной связи (ближайшего соседа), полной связи (дальнего соседа), средней связи и метод Варда [8]. Таким образом, в данной работе для кластерного анализа выбран иерархический метод.

### IV. ИЕРАРХИЧЕСКАЯ КЛАСТЕРИЗАЦИЯ

В рамках проведения вычислительного эксперимента для субъектов Федерации был взят 2022 год. Исходные данные представлены в табл. 2 и табл. 3. На основании представленных данных далее построим модель иерархической кластеризации.

ТАБЛИЦА II Показатели промышленного развития регионов ЦФО в 2022 г.

Регион	K <sub>1</sub>	K <sub>2</sub>	K <sub>3</sub>	K <sub>4</sub>
Белгородская область	10,5	0,7827	0,0172	0,0796
Брянская область	13,8	0,7508	0,0245	0,0718
Владимирская область	17,7	1,1264	0,0399	0,1169
Воронежская область	14	0,7994	0,0506	0,0655
Ивановская область	10,4	0,7829	0,0206	0,1056
Калужская область	5,8	1,3815	0,0895	0,1108
Костромская область	11,6	0,8408	0,0347	0,1007
Курская область	10,6	0,4821	0,0381	0,0687
Липецкая область	8,4	1,3222	0,0624	0,0880
Московская область	14,5	0,6463	0,0197	0,0685

Регион	$K_1$	$K_2$	$K_3$	$K_4$
Орловская область	14,5	0,5965	0,0199	0,0697
Рязанская область	14,3	0,8393	0,0303	0,0903
Смоленская область	11,7	0,8656	0,0243	0,0885
Тамбовская область	13,5	0,5711	0,0226	0,0680
Тверская область	8,8	0,6567	0,0192	0,0856
Тульская область	18,3	1,4350	0,0846	0,1087
Ярославская область	12,8	0,7748	0,0493	0,1059
Москва	14,9	0,1745	0,0034	0,0577

ТАБЛИЦА III ПОКАЗАТЕЛИ ПРОМЫШЛЕННОГО РАЗВИТИЯ РЕГИОНОВ ПФО В 2022 Г.

Регион	1	2	3	4
Республика Башкортостан	9,5	0,6790	0,0257	0,0616
Республика Марий Эл	9,8	0,8391	0,0265	0,0846
Республика Мордовия	13,2	1,0891	0,0447	0,0799
Республика Татарстан	19,8	0,8813	0,0402	0,0826
Республика Удмуртия	10,7	0,5166	0,0189	0,0998
Республика Чувашия	14,5	0,7727	0,0186	0,0828
Пермский край	49	0,8394	0,0646	0,0899
Кировская область	12,2	0,8285	0,0338	0,1010
Нижегородская область	19,1	0,8699	0,0453	0,1039
Оренбургская область	20,1	0,3545	0,0173	0,0558
Пензенская область	12,2	0,5565	0,0221	0,0809
Самарская область	17,1	0,5889	0,0337	0,0971
Саратовская область	9,3	0,6220	0,0229	0,0576
Ульяновская область	6,9	0,7552	0,0341	0,0975

Поскольку данные имеют различную размерность, на следующем этапе алгоритма они были нормализованы с использованием Z-преобразования. Для оценки различий в уровне промышленного развития между регионами Российской Федерации необходимо визуализировать данные с использованием дендрограммы. Так можно определить, в каких регионах этот уровень выше, а в каких – ниже.

В исследовании, в качестве метрики расстояния между регионами использовалось расстояние Евклида. Для построения дендрограмм применялся метод Варда, который выделяется тем, что оценивает расстояния между кластерами через дисперсионный анализ. Этот метод минимизирует сумму квадратов для любых двух кластеров, объединенных на каждом этапе. Более подробное описание методологии приведено в работе Варда 1963 года под названием «Иерархическая группировка для оптимизации целевой функции» [9]. В целом, этот метод оказывается эффективным, хотя он имеет тенденцию формировать кластеры меньшего размера.

Для более ясной визуализации результатов и анализа регионов в пределах Приволжского и Центрального федеральных округов на рис. 1–2 представлены данные иерархической кластеризации. В дендрограмме длина вертикальных линий соответствует евклидову расстоянию между регионами, оцененному по всем четырем критериям. Чем раньше (ниже на дереве) происходит слияние регионов, тем более схожи группы наблюдений между собой. Наблюдения, объединяющиеся ближе к вершине дерева, могут значительно отличаться по набору факторов анализа. Высота точки слияния, измеряемая по вертикальной оси, отражает степень различий между регионами внутри соответствующего федерального округа.

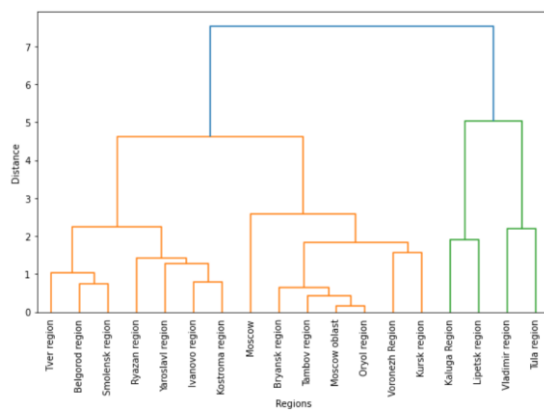


Рис. 1. Кластерный анализ регионов ЦФО

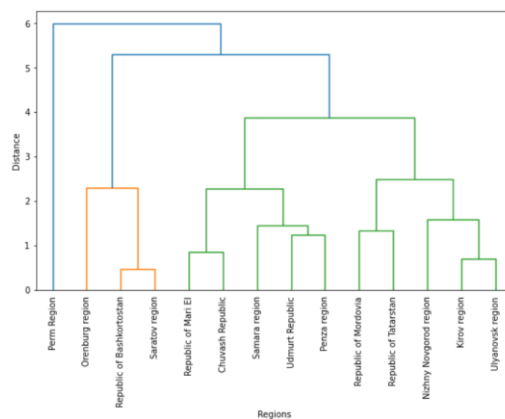


Рис. 2. Кластерный анализ регионов ПФО

Применение иерархической кластеризации для регионов с учетом уровня промышленного развития обеспечило формирование последовательных вложенных кластеров, которые эффективно представлены с использованием дендрограммы. При использовании альтернативных критериев связи, таких как метод центроидов, ближнего или дальнего соседа, а также различных методов расчета расстояния, таких как манхэттенское расстояние и др., наблюдалась незначительная изменчивость в составе кластеров. Отсутствие значительных изменений при изменении параметров кластеризации может говорить о стабильности кластерной структуры данных. На основе результатов кластерного анализа можно разрабатывать и оптимизировать политику развития промышленности в каждом кластере, учитывая особенности конкретных групп регионов.

#### V. ПРОГНОЗ ДИНАМИКИ ПРОМЫШЛЕННОГО РАЗВИТИЯ

Кластерный анализ промышленного развития регионов России позволяет сделать вывод о приоритетных направлениях инвестиций, интенсификации выпуска продукции и совершенствования механизмов государственного управления. В то же время, для анализа развития промышленного производства для задачи среднесрочного планирования, целесообразно рассмотреть прогнозы, характеризующие динамику индекса промышленного производства на горизонте трех лет. В рамках рассматриваемого периода были проанализированы значения индекса промышленного



производства и построены прогнозы с использованием моделей парной регрессии третьего порядка и экспоненциального сглаживания. Беглый анализ показывает ожидаемый рост индекса (рис. 3, 4).

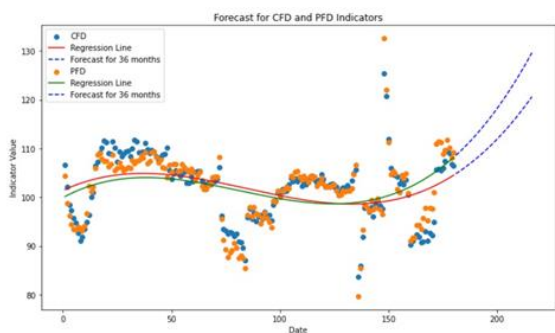
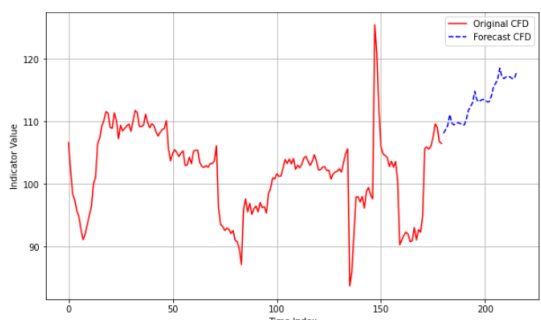
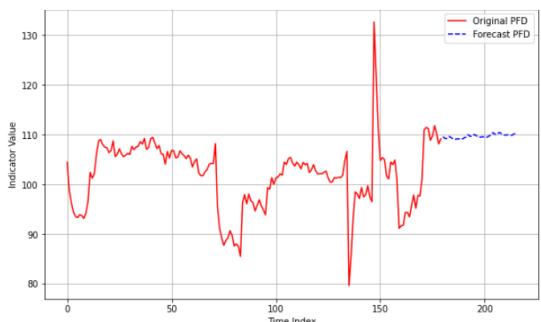


Рис. 3. Регрессионная модель прогноза индекса промышленного производства для ЦФО и ПФО



a)



b)

Рис. 4. Прогноз по методу экспоненциального сглаживания для ЦФО и ПФО

Исходя из полученных результатов, можно сделать вывод о наличии однозначного восходящего тренда в отношении производства центрального федерального округа, что подтверждается как большим углом наклона касательной в модели регрессии, проведенной в момент времени начала прогноза, так и ростом кривой прогноза по методу экспоненциального сглаживания.

Одновременно, необходимо учитывать специфику промышленной ориентации регионов. Так, сравнительно менее выраженный тренд роста прогноза индекса промышленного производства для Приволжского федерального округа объясним структурой производств и производимой номенклатуры продукции.

## VI. ВЫВОДЫ

В данной работе представлена модель кластеризации регионов страны на основе уровня промышленного развития, использующая иерархические процедуры кластеризации а также прогноз промышленного развития на ЦФО и ПФО. Для формирования системы показателей кластеризации были использованы данные, полученные из официальных статистических источников, отражающих состояние промышленности в регионах. Проведенный вычислительный эксперимент наглядно демонстрирует группы регионов с похожими паттернами промышленной активности, а также помог выявить регионы с существенными различиями в промышленности. Сравнительный анализ прогнозов, построенных с использованием регрессионных и статистических моделей, показывает, что в ближайшие три года ожидается рост промышленного производства в Центральном и Приволжском федеральных округах.

## СПИСОК ЛИТЕРАТУРЫ

- [1] Нормова Ю.В. Влияние промышленного производства на технологическое развитие регионов России // Экономика, предпринимательство и право. 2022. Том 12. № 5. С. 1475-1492. doi: 10.18334/epp.12.5.114632.
- [2] Грамович Я.В., Мусатов Д.Ю., Петрусевич Д.А. Применение беггинга в прогнозировании временных рядов. *Russian Technological Journal*. 2024;12(1):101-110. <https://doi.org/10.32362/2500-316X-2024-12-1-101-110>
- [3] Дударин П.В., Ярушкина Н.Г. Подходы к нечеткой и иерархической кластеризации и классификации ключевых показателей эффективности системы стратегического планирования российской федерации // Нечеткие системы, мягкие вычисления и интеллектуальные технологии (НСМВИТ-2017): труды VII всероссийской научной-практической конференции, Санкт-Петербург, 03–07 июля 2017 года. Том 2. Санкт-Петербург: Политехника-сервис, 2017. С. 65-74. EDN ZNMWBL.
- [4] Верхорунов А.И., Жуков В.Г. О решении задачи кластерного анализа инцидентов информационной безопасности // Решетневские чтения, 2011, №15, с. 654-655.
- [5] Федеральная служба государственной статистики. [Электронный ресурс]. Режим доступа: URL: <https://rosstat.gov.ru/> (in Russian)
- [6] Казанцев С.В., Митяков Е.С. Оценка значимости факторов динамики жизнеспособности субъектов Российской Федерации // Экономическая безопасность. 2022. Том 5. № 1. С. 155-174. doi: 10.18334/ecsec.5.1.114267.
- [7] Жамбю М. Иерархический кластер-анализ и соответствия. М.: Финансы и статистика. 1988. 345 с.
- [8] Дж. Вэн Классификация и кластер. М.: Мир, 1980. 390 с.
- [9] Ward J.H., Jr. (1963), "Hierarchical Grouping to Optimize an Objective Function" // Journal of the American Statistical Association, 58, pp. 236–244.

# Методы машинного обучения для прогнозирования результативности онлайн-курсов

М. А. Бурнаев, Е. А. Болдырева

Национальный исследовательский университет ИТМО

burnaev.maxim.a@gmail.com

**Аннотация.** Данный доклад описывает разработку модели прогнозирования результативности онлайн-курсов на основе активности в виртуальной образовательной среде. Исследование использует данные семи курсов, представленные в датасете Открытого университета Великобритании. Анализ результатов показал, что наиболее эффективными методами прогнозирования успешности прохождения курса являются случайный лес и градиентный бустинг. Лучшее значение ассигасы достигло в 76,2 %, а лучшее значение f1-метрики в 60,2 %. Однако все рассмотренные методы достигли хороших результатов, демонстрируя способность предсказания на основе доступных признаков. Таким образом, данное исследование предлагает подход к оценке успеваемости студентов на онлайн-курсах на основе методов машинного обучения. Авторам удалось улучшить предсказательную способность успешности прохождения онлайн-курсов за счет комбинирования методов прогнозирования и настройки их параметров.

**Ключевые слова:** машинное обучение; онлайн-образование; модель прогнозирования

## I. ВВЕДЕНИЕ

В современном образовании онлайн-курсы становятся все более популярными, предоставляя возможность получить образование из любой точки мира. Однако, важно обеспечить их эффективное освоение студентами, так как на данный момент люди очень часто, проходя онлайн-курсы, не доходят до финала [1].

Целью данной работы является повышение результативности онлайн-курсов. Для этого авторы данной статьи занимались разработкой прогностической модели, которая основывается на данных взаимодействия студентов с виртуальной образовательной средой. Эксперименты проводились с использованием различных методов машинного обучения: случайный лес, дерево решений, градиентный бустинг, k-ближайших соседей, логистическую регрессию, метод опорных векторов, линейный дискриминантный анализ и наивный байесовский классификатор. В ходе исследования были определены наиболее эффективные методы для прогнозирования успешности прохождения онлайн-курсов.

Данное исследование предлагает практические инструменты для прогнозирования успешности прохождения онлайн-курсов, что может быть полезным

для платформ онлайн-образования в оптимизации процесса обучения и поддержки студентов.

## II. ЭКСПЕРИМЕНТАЛЬНАЯ ЧАСТЬ

### A. Исходные данные

В качестве исходных данных для разработки модели прогнозирования авторы нашли датасет Открытого университета Великобритании. Это полностью онлайн-университет, предлагающий аккредитованные степени для дистанционного обучения.

Датасет содержит данные о курсах, студентах и их взаимодействии с виртуальной образовательной средой. Набор данных состоит из таблиц, связанных с помощью уникальных идентификаторов. Все таблицы хранятся в формате csv.

Датасет содержит следующие файлы:

- файл `courses.csv` содержит список всех доступных курсов.
- файл `assessments.csv` содержит информацию об оценках на курсах. Обычно каждый курс имеет ряд оценок, за которыми следует итоговый экзамен.
- файл `vle.csv` содержит информацию о материалах курса, которые используются при прохождении.
- файл `studentInfo.csv` содержит демографическую информацию о студентах вместе с их результатами.
- файл `studentRegistration.csv` содержит информацию о времени регистрации студента на курс. Для студентов, которые отменили регистрацию, также записывается дата отмены регистрации.
- файл `studentAssessment.csv` содержит результаты оценок студентов. Если студент не представил оценку, результат не записывается. Результаты итогового экзамена отсутствуют, если результаты оценок не сохранены в системе.
- файл `studentVle.csv` содержит информацию о взаимодействии каждого студента с материалами в виртуальной образовательной среде.

### В. Подготовка данных

Для дальнейшего прогнозирования успешности прохождения студентом онлайн-курса была собрана в единую таблицу информация о взаимодействии каждого студента с виртуальной средой. Предположение было сделано, что пол студента, его этническая принадлежность и т. п. не сильно влияют на результат. Планируется построить предсказательную модель, основываясь на данных активности студента в течение курса.

Для этого были собраны в единую таблицу следующие данные:

- количество кликов на момент прогнозирования;
- количество дней взаимодействия с материалами курса;
- количество обращений к материалам курса;
- количество выполненных промежуточных тестов;
- средний балл промежуточных работ.

Было определено количество дней с начала курса, когда мы хотим предсказывать успешность прохождения курса. Для этого было определено количество процентов завершенности курса, которое было умножено на минимальную длину курса и поделено на 100.

В файле `studentRegistration` были удалены записи, где студенты ещё не были зарегистрированы на курс на момент предсказания, либо уже отписались от курса.

На основе данных из `studentRegistration` и `studentInfo` была получена таблица с информацией о том, на каком курсе был записан каждый студент и с каким результатом он его окончил: `pass` (пройдено), `withdrawn` (снялся с курса), `fail` (правалено), `distinction` (отлично). Также присутствовала информация о количестве зачетных единиц каждого курса, которая пригодилась в дальнейшем.

В файле `studentVle` были удалены записи, где дата взаимодействия студента с материалом была после момента предсказания. На основе полученных данных было вычислено количество записей в каждой группе, тем самым получив таблицу с количеством обращений студента к материалам в виртуальной образовательной среде на момент предсказания.

Для вычисления общего количества дней, которые студенты потратили на обучение на момент предсказания, использовались данные из файла `studentVle` таким образом, чтобы исключались повторяющиеся записи, когда студент обращался в один день к материалам несколько раз.

Также на основе данных из файла `studentVle` была получена таблица с количеством кликов, которые сделал студент на момент предсказания.

На основе файлов `studentAssessment` и `assessments` была получена таблица, где содержится информация о

том, сколько промежуточных оценок получил студент. Был вычислен средний балл промежуточных оценок.

В конце были объединены все полученные таблицы в одну общую по столбцам. В результате получилась таблица со следующими столбцами: `final_result` (результат прохождения онлайн-курса), `studied_credits` (количество зачетных единиц), `num_clicks` (количество кликов), `daysOfTraining` (количество дней обучения), `refToMaterials` (количество обращений к материалам), `averageScore` (средний балл) и `numOfIntermediateClasses` (количество промежуточных оценок).

Поскольку разные курсы имели разную трудоемкость, то количество усилий, затрачиваемых студентами на их прохождение, отличалось в зависимости от курса. В этом случае невозможно было построить общую модель для разных курсов. Для решения этой проблемы каждое значение в столбце `studied_credits` в итоговой таблице было умножено на 36, чтобы получить примерное количество часов, которое студент должен был потратить на прохождение. Затем значения в столбцах `num_clicks`, `daysOfTraining`, `refToMaterials` и `numOfIntermediateClasses` были разделены на полученное время в каждом случае, тем самым получив значение, относительное к трудоемкости.

### С. Выбранные методы машинного обучения

Для реализации алгоритма прогнозирования успешности прохождения студентом онлайн-курса были выбраны следующие методы машинного обучения:

- k-ближайших соседей: алгоритм классификации, основанный на оценивании сходства объектов [2]. Классифицируемый объект относится к тому классу, которому принадлежат ближайшие к нему объекты обучающей выборки.
- Случайный лес: алгоритм машинного обучения, заключающийся в использовании ансамбля решающих деревьев [3]. Основная концепция заключается в применении обширного ансамбля решающих деревьев, где каждое дерево, взятое по отдельности, может обладать лишь невысокой точностью классификации. Однако, благодаря значительному их количеству, конечный результат демонстрирует высокое качество предсказаний.
- Дерево решений: этот метод машинного обучения направлен на разбиение обширного массива входных данных на компактные группы с целью прогнозирования возникновения событий в зависимости от определенных условий [4]. Если визуализировать этот процесс, то можно представить древовидную структуру, в которой конечные вершины отображают классы или принятые решения, а промежуточные вершины представляют собой условия, подлежащие проверке.
- Градиентный бустинг: это техника машинного обучения для задач классификации и регрессии, которая строит модель предсказания в форме ансамбля предсказывающих моделей, обычно деревьев решений [5].

- Логистическая регрессия: алгоритм разработки линейного классификатора, который способен оценивать апостериорные вероятности принадлежности объектов к различным классам [6].
- Метод опорных векторов: метод машинного обучения, который используется при решении задач классификации и регрессии. Ключевая идея данного метода состоит в том, чтобы построить гиперплоскость, которая разделяла бы объекты выборки наилучшим способом [7]. Алгоритм оперирует идеей, что чем значительнее разрыв между разделяющей гиперплоскостью и элементами классов, тем ниже среднее значение ошибки классификации.
- Линейный дискриминантный анализ: представляет собой важный инструмент в области машинного обучения, нацелен на выявление оптимальной линейной комбинации признаков, способной эффективно описывать или разделять различные классы или события [8]. Этот метод позволяет сформировать уникальную комбинацию, которая может быть использована в качестве линейного классификатора, демонстрируя высокую степень точности и предсказательной силы.
- Наивный байесовский классификатор: простой вероятностный классификатор, основанный на применении теоремы Байеса со строгими (наивными) предположениями о независимости [8].

Для проведения экспериментов с различными значениями параметров моделей была использована функция GridSearchCV из библиотеки scikit-learn. GridSearchCV предоставляет удобный инструмент для систематического поиска оптимальных гиперпараметров модели путем перебора всех возможных комбинаций значений в заданном диапазоне.

Для оценки работы алгоритмов использовались следующие метрики:

- **Assurasy:** измеряет долю правильно классифицированных объектов относительно общего числа объектов.
- **F1-мера:** является гармоническим средним precision и recall. Precision представляет собой метрику, определяющую процент правильно предсказанных положительных классов от общего числа классифицированных положительных классов. Recall, в свою очередь, измеряет долю верно определенных положительных классов от общего количества истинных положительных классов.

#### D. Полученные результаты

Прогнозирование производилось на момент, когда курс был завершен примерно на 50%. В качестве признаков использовались следующие поля из итоговой таблицы: num\_clicks, daysOfTraining, refToMaterials, averageScore, numOfIntermediateClasses. В качестве

откликов были определены два значения: 1 — студент провалит итоговый экзамен, 0 — студент не провалит итоговый экзамен. Под значение 0 были объединены те записи итоговой таблицы, где в столбце final\_result значение равнялось Pass и Distinction, а под 1 значения Fail и Withdrawn.

Для метода k ближайших соседей проводились эксперименты со следующими значениями k: 5, 10, 15, 20, 22, 24, 26, 28, 30. Было выявлено, что при увеличении параметра k, значение assurasy увеличивается. Лучшее значение assurasy удалось достигнуть в 73.9% при k=30, лучшее значение f1-метрики было 55% при k=30, 15, 28.

Для методов случайный лес и дерево решений проводились эксперименты с параметром глубины дерева. Использовались следующие значения: 1, 2, 3, 4, 5, 7, 10, 12, 14, 16, 18. Лучшее значение assurasy для случайного леса достигло в 76% при глубине дерева равной 10, для дерева решений 75,8% при глубине дерева равной 7. Лучшее значение f1-метрики для случайного леса достигло 59% при глубине дерева равной 7, а для дерева решений 62% при глубине дерева равной 2.

Для метода градиентный бустинг проводились эксперименты с параметром глубины. Использовались следующие значения: 1, 2, 3, 4, 5, 6, 7. Лучшее значение assurasy для данного метода достигло в 76,2% при глубине равной 4, лучшее значение f1-метрики достигло в 60,2% при глубине равной 2.

Для наивного байесовского классификатора проводились эксперименты с параметром smoothing, который обозначает часть наибольшего отклонения всех объектов, которая добавляется к отклонениям для стабильности вычислений. Использовались значения данного параметра от  $1e-9$  до  $9e-4$ . Было выявлено, что оптимальным значением параметра smoothing для данного метода является  $1e-6$ , при котором значение assurasy равно 69.8%, а f1-метрика равна 60.9%, хотя при больших значениях параметра значение assurasy увеличивается, но значение f1-метрики стремительно уменьшается.

Для метода опорных векторов производились эксперименты с различными типами ядра: линейным, полиномиальным, ядром Гаусса. Лучшим вариантом оказалась модель с полиномиальным ядром со значением assurasy 73.4% и f1-метрикой 53.3%.

Для линейного дискриминантного анализа проводились эксперименты с параметром solver (метод для решения задачи оптимизации): svd (использование сингулярного разложения для вычислений), lsqr (использование метода наименьших квадратов), eigen (разложение на собственные векторы). Лучшие показатели достигли модели при параметрах svd и eigen со значениями assurasy 73.6% и f1-метрикой 52.4%.

Для логистической регрессии также производились эксперименты с параметром solver: lbfgs (метод Ньютона с ограниченной памятью), liblinear (линейный метод), newton-cg (метод Ньютона), sag (стохастический средний градиент). Лучшие показатели были достигнуты для модели с параметром lbfgs со значениями assurasy 73.4% и f1-метрикой 51.8%.

Исходя из проведенных экспериментов видно, что для прогнозирования успешности прохождения онлайн-курса наиболее рекомендуемыми методами будут являться случайный лес, дерево решений, градиентный бустинг. Эти методы достигли наилучших показателей по метрикам *accuracy* и *f1*. Табл. 1 отражает лучшие значения метрик по всем методам.

ТАБЛИЦА 1. ЛУЧШИЕ ЗНАЧЕНИЯ МЕТРИК ДЛЯ РАЗНЫХ МЕТОДОВ

	Метрики	
	<i>accuracy</i>	<i>f1</i>
К-ближайших соседей	73.9%	55%
Случайный лес	76%	59%
Дерево решений	75.8%	62%
Градиентный бустинг	76.2%	60.2%
Наивный байесовский классификатор	69.8%	60.9%
Метод опорных векторов	73.4%	53.3%
Линейный дискриминантный анализ	73.6%	52.4%
Логистическая регрессия	73.4%	51.8%

### III. ЗАКЛЮЧЕНИЕ

В данной работе были представлены результаты исследования по повышению эффективности онлайн-курсов через разработку прогностической модели на основе данных взаимодействия студентов с виртуальной образовательной средой.

Эксперименты с различными методами машинного обучения показали, что для прогнозирования успешности прохождения онлайн-курсов наиболее эффективными являются случайный лес, дерево решений и градиентный бустинг. Эти методы достигли наилучших показателей по метрикам *accuracy* и *f1*. При этом были рассмотрены и другие методы, такие как *k*-ближайших соседей, логистическая регрессия, метод

опорных векторов, линейный дискриминантный анализ и наивный байесовский классификатор, однако их эффективность оказалась ниже в сравнении с рассмотренными методами.

В целом, данная работа демонстрирует потенциал использования машинного обучения для улучшения онлайн-образования. Прогностическая модель, разработанная на основе данного исследования, может быть полезным инструментом для платформ онлайн-образования, помогая предлагать персонализированные подходы к обучению и повышая результативность курсов для студентов.

### СПИСОК ЛИТЕРАТУРЫ

- [1] Слепое пятно EdTech. Почему ученики не достигают успеха после прохождения онлайн-курсов [Электронный ресурс]. Режим доступа: URL: <https://incrussia.ru/understand/edtech-problems/> (дата обращения 07.04.2023)
- [2] Басар А.А. Методы искусственного интеллекта: учебное пособие. Сибирский государственный университет геосистем и технологий, 2024. 164 с.
- [3] Монарх Р. Машинное обучение с участием человека. Издательство "ДМК Пресс", 2022. 498 с.
- [4] Кугаевских А.В., Муромцев Д.И., Кирсанова О.В. Классические методы машинного обучения: Учебное пособие. Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, 2022. 53 с.
- [5] Митина О.А., Жаров В.В. Технологии и инструментарий машинного обучения: Учебное пособие. МИРЭА - Российский технологический университет, 2023. 203 с.
- [6] Коэльо Л.П., Ричарт В. Построение систем машинного обучения на языке Python. Издательство "ДМК Пресс", 2016. 302 с.
- [7] Флах П. Машинное обучение. Наука и искусство построения алгоритмов, которые извлекают знания из данных Издательство "ДМК Пресс", 2015. 400 с.
- [8] Шалев-Шварц Ш., Бен-Давид Ш. Идеи машинного обучения: Учебное пособие. Издательство "ДМК Пресс", 2019. 436 с.

# Генерация трехмерной сцены на основе одного двумерного изображения с помощью глубоких нейронных сетей

А. С. Сырых

Национальный исследовательский  
университет ИТМО

alyoshca.syryh@mail.ru

Г. О. Бондаренко

Санкт-Петербургский государственный  
электротехнический университет  
«ЛЭТИ» им. В.И. Ульянова (Ленина)

Olakola9@gmail.com

**Аннотация.** В данной статье рассматривается актуальная проблема генерации трехмерной сцены на основе одного двумерного изображения. Изучается широкий спектр современных методов и подходов, применяемых для автоматической реконструкции трехмерных сцен из одного двумерного изображения. Анализируются преимущества и ограничения каждого подхода, а также их применимость в различных областях.

**Ключевые слова:** 3D сцена, компьютерное зрение, 3D меш

## I. ВВЕДЕНИЕ

Человек способен обрабатывать огромное количество информации в секунду, и мозг использует различные стратегии для понимания того, что мы видим. Одна из таких стратегий заключается в том, что наш мозг интерпретирует трехмерный объем, хотя мы воспринимаем мир через двумерное изображение. Люди могут легко определить трехмерную геометрию и текстуру объекта по одному двумерному изображению. Мы можем рассуждать о форме и внешнем виде объекта, рассматриваемого с одной точки зрения, потому что мы видели много похожих объектов с разных точек зрения. Опыт позволяет нам сделать обоснованный вывод о форме и внешнем виде нового объекта из знакомой категории, даже несмотря на различия внутри категории [1, 2].

Существующие алгоритмы могут реконструировать трехмерную форму и текстуру по одному изображению, но существуют серьезные проблемы с обобщением: восстановленное представление, созданное нейронной сетью, кажется нереалистичным. Либо восстановленная 3D-сетка сглаживается при просмотре с других точек зрения, либо восстановленная текстура выглядит грязной и нерезкой с отсутствующими деталями [3, 4].

Определение трехмерной формы, текстуры, параметров камеры и освещения на основе отдельных двумерных изображений по своей сути является сложной задачей. Чтобы обойти эту проблему, используются наборы данных с несколькими видами изображений, особенно синтетические [4, 5, 6]. Изображения с несколькими видами недоступны для многих наборов данных, и обучение прогнозированию формы и текстуры с несколькими видами на основе наборов данных с одним изображением является важной областью исследований.

Извлечение трехмерной информации из одного изображения имеет множество применений в компьютерном зрении, робототехнике и анализе сцен, особенно на мобильных устройствах дополненной/виртуальной реальности. Это также может быть применено для ускорения создания 3D-моделей, персонажей, сцен, анимации, кино и, что особенно важно, для упрощения задач создания уровней в играх, архитектурных сцен и многого другого.

## II. ОБЗОР И АНАЛИЗ МЕТОДОВ ПРЕОБРАЗОВАНИЯ ИЗОБРАЖЕНИЯ В СЦЕНУ

Получение 3D-сцены из одного изображения - сложная задача, поскольку она может включать в себя несколько подготовительных этапов перед созданием 3D-модели, не говоря уже о 3D-сцене.

Эти этапы могут включать классификацию объектов, определение их правильного расположения и размеров относительно сцены и только затем создание реалистичных 3D-моделей и текстур. Классификация может быть полезна для выбора подходящего шаблона 3D-модели, если такой метод используется для повышения реалистичности результата.

Обнаружение особенно важно для получения 3D-сцены, поскольку для точного определения размеров и форм объектов необходимо точно очертить границы этих объектов. Если цель состоит в том, чтобы получить 3D-сцену, которой можно впоследствии манипулировать, например, перемещать или масштабировать определенные объекты, важно, чтобы эти объекты были разделены и не пересекались друг с другом. Такие проблемы можно решить с помощью первоначального обнаружения изображения.

Трехмерная информация может быть представлена в различных формах: трехмерные ограничивающие рамки [2, 5, 6], облака точек [6], сетки [3] и воксели [4] (рис. 1).



Рис. 1. Типы представления информации о трехмерных объектах

Недавно модели на основе GAN достигли реалистичного синтеза изображений различных

объектов. Однако для использования этой технологии в играх, симуляторах и виртуальных приложениях необходимо, чтобы они были полностью управляемыми. Недавние работы управляют генерацией изображений путем обработки выходных данных различными типами входных данных, такими как естественные и синтетические изображения, ориентиры и семантические карты.

Современные методы выполняют реконструкцию нескольких объектов путем независимой обработки данных, полученных с помощью современных детекторов объектов, или путем совместного прогнозирования множества объектов в плотной воксельной сетке [4]. Однако это может быть связано с большими вычислительными затратами из-за проблем масштабирования. Вместо этого в Image2Objects [5] предлагается использовать детекторы ключевых точек для определения местоположения объектов в виде разреженных центральных точек и прямого прогнозирования ограничивающих рамок совместно для всех объектов в сцене.

Первым важным аспектом, который необходимо учитывать при создании метода генерации 3D-сцен, является то, что 3D-объекты внутри сцены не должны пересекаться или накладываться друг на друга.

Устранение коллизий способствует появлению непересекающихся объектов, что еще больше повышает реалистичность реконструкции. Учитывая RGB-изображение, представленный подход выполняет легкую реконструкцию в один этап, он способен работать в реальном времени, полностью дифференцируем, и поддается сквозному обучению. Для реалистичной и физически правдоподобной реконструкции объекты должны быть правильно расположены и не пересекаться друг с другом. Для достижения корректного отображения объектов без пересечения одного в другой необходимо ввести потери от столкновений, которые поддерживают правдоподобные реконструкции так, чтобы реконструированные объекты не пересекались.

3D-RCNN оценивает 3D-форму каждого экземпляра объекта на изображении с помощью подхода обучения «рендеринг и сравнение», где форма представлена в виде линейного базиса из базы данных 3D-моделей. Такое представление формы является точным для классов с низкой внутриклассовой изменчивостью, таких как автомобили и люди.

В Image2Objects [5] основная идея заключается в совместной оптимизации обнаружения, выравнивания и формы всех объектов на изображении RGB с акцентом на реалистичные и физически правдоподобные реконструкции.

Для достижения этой цели в Image2Objects используется детектор ключевых точек, который локализует объекты как центральные точки и напрямую предсказывает все свойства объекта, включая ограничивающие рамки и 3D-фигуры, и все это за один прямой проход. Этот подход формулирует 3D-реконструкцию фигуры как задачу выбора формы, выбирая ее из базы данных шаблонов фигур. Это делает его независимым от представлений формы, позволяя ему легко восстанавливать реалистичные и визуально приятные формы на основе моделей САПР, в то время

как задача изучения формы формулируется на основе облаков точек и воксельных представлений.

Для реалистичной и физически правдоподобной реконструкции объектов важно, чтобы объекты были расположены правильно и не пересекались друг с другом. Чтобы добиться правильного отображения объектов без наложения друг на друга, необходимо учитывать потери при столкновении. Эти потери при столкновении поддерживают реалистичную реконструкцию, гарантируя, что реконструированные объекты не пересекаются друг с другом.

Еще одним ключевым фактором является выбор наилучшего представления формы объектов. Хотя были предложены различные представления, такие как знаковые функции расстояния (SDF), сетки, воксельные сетки, облака точек и даже гибридные подходы [3], все они имеют свои преимущества и недостатки в зависимости от задачи. В Image2Objects введен механизм выбора фигур, не зависящий от представления. Это означает, что шаблоны фигур выбираются из заданной базы данных фигур, которая может содержать различные (или множественные) представления. Наиболее подходящее представление выбирается в зависимости от задачи, будь то для целевых функций или для целей визуализации.

Архитектура CenterNet является модульной и может быть легко расширена для решения различных задач, таких как 2D-обнаружение, 3D-обнаружение, оценка позы человеческого тела и отслеживание [6]. В работе CenterNet показана необходимость полной и согласованной 3D-реконструкции нескольких объектов, где каждый пиксель голосует за метку класса, 3D-ограничительную рамку и образец 3D-формы для размещения объектов в мировой системе координат.

Получив одно изображение RGB, модель из [5] определяет центры объектов как ключевые точки на тепловой карте. Сеть непосредственно предсказывает образцы формы и 9-DoF ограничивающие рамки совместно для всех объектов в сцене

В случае Mesh R-CNN [3] процесс начинается с обнаружения и классификации объектов на изображении (рис. 2). Затем эти объекты представляются в виде вокселей, за которыми следует своего рода преобразование из воксельного представления в трехмерные сетки.

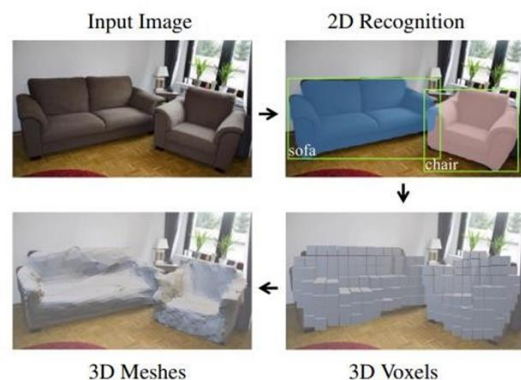


Рис. 2. Алгоритм Mesh R-CNN

CoReNet выполняет плотное предсказание формы в фиксированной сетке из  $128^3$  вокселей, которая не масштабируется в соответствии с размерами реконструируемого мира [4]. Кроме того, во время обучения он инкапсулирует всю информацию о сцене в единую.

CoreNet восстанавливает все объекты в их правильном пространственном положении в общей системе координат, обеспечивая устранение искажений и дефектов формы. Справа показана реконструкция объекта в реальной сцене. На верхнем изображении показана реконструкция, наложенная на входной RGB-сигнал. В нижней строке показаны входные данные наряду с двумя другими типами реконструкций (рис. 3).

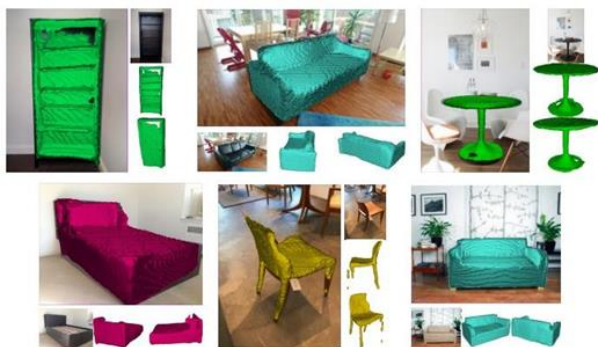


Рис. 3. Примеры результатов архитектуры CoReNet

Упомянутые выше работы основаны на сложной двухэтапной архитектуре, при которой сначала обнаруживаются объекты, а затем оценивается их форма. В отличие от этого, метод Image2Objects является более модульным, способным обнаруживать и реконструировать различное количество объектов, включая новые комбинации классов, которые не наблюдались во время обучения. На рис. 4 приведено сравнение между CoReNet и Image2Objects.



Рис. 4. Сравнение результатов CoReNet и Image2Objects

Для получения качественного результата необходимо решить проблему проклятия размерности, что обусловлено экспоненциальным ростом количества возможных результатов и приводит к необходимости использования большого объема памяти, ухудшению качества результатов обучения нейронной сети и повышению требований к вычислительной мощности.

Основная идея решения проклятия размерности состоит в снижении размерности. При уменьшении размерности также экспоненциально уменьшается кол-во вычислений и предположений, что делает модель, тем самым увеличивается скорость обучения и требуется меньше места для хранения данных.

Т.к. задача получения 3D-сцены из одного изображения является весьма трудоемкой для нейронной сети и требует огромного кол-ва данных для обучающей и тренирующей выборки, а также необходимая размерность с развитием возможностей нейронных сетей будет только возрастать для получения более качественного результата, то важным фактором для её решения будет использование такого метода понижения размерности, который будет выбирать отличительные признаки качественно, и ускорять процесс обучения и тестирования нейронной сети. Если сравнивать новый метод UMAP [7] с t-SNE, то первый работает в среднем в 5 раз быстрее.

Хоть автоэнкодеры в последнее время активно вытесняют остальные методы понижения размерности, но эффективны только на тех данных, которые похожи на обучаемые.

### III. ЗАКЛЮЧЕНИЕ

Был проведен обзор существующих методов устранения проклятия размерности, анализ существующих алгоритмов моделирования 3D-объектов по одному изображению с использованием нейронных сетей и изучение существующих алгоритмов моделирования пространства по одному изображению с использованием нейронных сетей.

За последние 5 лет появились различные подходы к получению 3D-моделей, как из нескольких наборов данных, так и из одного изображения. На данном этапе разработки преимущественно используется сетчатое представление, поскольку нейронные сети предоставляют больше информации о форме по сравнению с облаками точек и воксельными представлениями, что приводит к более точной форме. Однако это приводит к увеличению требований к данным, поступающим от нейронной сети, как для обучения, так и для тестирования, а также увеличивает сложность подготовки наборов данных для обучения и время обучения, поскольку размерность увеличивается, а вычисление ошибок становится более дорогостоящим с точки зрения вычислений.

Из представленного обзора существующих методов можно сделать вывод, что существующие методы получения 3D-модели по изображению позволяют предсказать общую форму объектов, но все же имеют различные проблемы с отображением трехмерной сцены. Кроме того, количество объектов, которые они могут надежно создать в одной сцене из одного изображения, ограничено и обычно составляет от 1 до 3 объектов.

При рассмотрении создания трехмерного пространства из одного изображения особое внимание следует уделить таким подходам, как Mesh R-CNN, CoreNet и Image2Objects. На данном этапе развития данной области существующие методы генерируют объекты с большим количеством дефектов и не пригодны для использования в проектах, поэтому есть необходимость в создании метода, который можно будет использовать для ускорения процессов создания сцен в игровой индустрии и кинематографе.



Каждый метод имеет свои преимущества с точки зрения решения разных подзадач в генерации трехмерной сцены. В случае с Image2Objects очевидно, что для достижения хороших результатов важно заключить все обнаруженные объекты в ограничивающие рамки 9-DoF, отражающие предполагаемый размер объекта, его точное местоположение в сцене и его центр. Это помогает устранить коллизии с другими объектами, что приводит к созданию 3D-сцены с отдельными объектами, которые можно перемещать и масштабировать.

Однако, как видно из рис. 4, такой подход может привести к большому количеству неправильных интерпретаций формы, потому что используются подготовленные 3D объекты по классам и масштабируются по ограничивающей рамке. Даже при наличии возможности выбирать представления из базы данных и вносить незначительные коррективы в Image2Objects, это не является полным успехом, поскольку этот подход в основном работает с примитивами и не учитывает все тонкости форм, которые пользователь может захотеть зафиксировать с помощью алгоритма.

В Mesh-RCNN используется двумерная детекция, которой недостаточно для решения проблемы коллизии, но с помощью генерации в воксельное представление трехмерной сцены в 3D меш получается достигнуть правдоподобной формы, но данный метод чувствителен к перспективным изменениям, к теням и требует больше вычислительных мощностей, чем выше представленные методы.

В планах на будущее – создание синтетических данных (входных изображений и желаемых результатов в виде 3D-сцен) для обучения и тестирования нейронной сети. Кроме того, цель состоит в том, чтобы объединить лучшие качества существующих подходов для получения 3D-сцены из одного изображения с использованием нейронных сетей. Будут использованы

идеи для предварительной обработки, такие как определение центра объекта и создание рамок из Image2Objects, а также постобработка для уточнения формы на основе вокселей и сеток с использованием нейронной сети после обнаружения объекта в пределах рамок из CoreNet и Mesh R-CNN.

Чтобы гарантировать, что этот алгоритм сможет конкурировать с текущими успешными проектами, важно посвятить часть исследований решению проблемы уменьшения размерности (выбору наилучшего метода уменьшения размерности и его гиперпараметров с точки зрения скорости обучения нейронной сети и результатов) и определению взаимного расположения объектов по отношению друг к другу и к другим параметрам, решению проблемы потенциальных пересечений объектов. Это должно привести к увеличению количества объектов, снимаемых в сцене, и улучшению их качества по сравнению с существующими алгоритмами.

#### СПИСОК ЛИТЕРАТУРЫ

- [1] G. Gkioxari, J. Malik, J. Johnson. "Mesh R-CNN". In Conference on Computer Vision and Pattern Recognition (CVPR), 2019.
- [2] S. Popov, P. Bauszat, V. Ferrari. "CoReNet: Coherent 3D scene reconstruction from a single RGB image". In European Conference on Computer Vision (ECCV), 2020.
- [3] F. Engelmann, K. Rematas, B. Leibe, V. Ferrari. "From Points to Multi-Object 3D Reconstruction". In Conference on Computer Vision and Pattern Recognition (CVPR), 2021.
- [4] X. Zhou, D. Wang, P. Krahenbuhl. "Objects as Points". In arXiv preprint arXiv:1904.07850, 2019.
- [5] N. Wang, Y. Zhang, Z. Li, Y. Fu, W. Liu, Y. Gang Jiang. "Pixel2Mesh: Generating 3D mesh models from single RGB images." In ECCV, 2018.
- [6] A. Kundu, Y. Li, J. M. Rehg. "3D-RCNN: Instance-level 3D Object Reconstruction via Render-and-Compare", 2018.
- [7] Leland McInnes, John Healy, James Melville. "UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction", 2020.

# Оценка различных методов машинного обучения для индивидуальной классификации отдельных проб вызванных потенциалов при восприятии правильных и неправильных решений математических примеров

Ж. В. Нагорнова

*Институт эволюционной физиологии и биохимии им. И.М. Сеченова РАН*  
Санкт-Петербург, Россия  
nagornova\_zh@mail.ru

А. С. Сивцева

*Национальный исследовательский университет ИТМО*  
Санкт-Петербург, Россия  
sivalien@mail.ru

Н. В. Шемякина

*Институт эволюционной физиологии и биохимии им. И.М. Сеченова РАН*  
Санкт-Петербург, Россия  
shemyakina\_n@mail.ru

**Аннотация.** В работе представлены результаты классификации вызванных потенциалов при восприятии правильного и неправильного решения математических примеров. Классификация проводилась индивидуально для каждого испытуемого (20 человек). Для классификации использовались показатели амплитуд ЭЭГ во временном окне 0–400 мс после предъявления ответа от лобно-центрального и теменного отведений. Для классификации использовали метод опорных векторов (SVM), логистическую регрессию (LR) и три различных архитектуры сверточных нейронных сетей, используемых для анализа данных ЭЭГ/ВП: *ShallowConvNet*, *EEGInception*, *EEGResNet*. каждая модель была обучена 10 раз. Для каждой модели и каждого запуска была вычислена средняя точность классификации на тестовой выборке. Нейронные сети показали лучшие результаты классификации, по сравнению с линейными методами (SVM, LR). Из трех архитектур нейросетей лучшим результат классификации в среднем по группе получен при помощи *EEGResNet* и *EEGInception*: медиана по группе 66.3 %. Лучшая индивидуальная точность классификации составила 87.5 % (*EEGInception*), однако для некоторых испытуемых точность классификации была выше при использовании *EEGResNet* и *ShallowConvNet*. При оценке информативности коротких окон (200 мс с шагом 100 мс от начала предъявления стимула) было показано, что лучшие точности классификации наблюдаются либо на всем интервале 400 мс, либо в интервале 200–400 мс от начала предъявления стимула.

**Ключевые слова:** ЭЭГ, ВП, классификация отдельных проб, *ShallowConvNet*, *EEGInception*, *EEGResNet*, SVM, LR

## I. ВВЕДЕНИЕ

Вызванные потенциалы (ВП) мозга используются в качестве управляющих сигналов для интерфейсов мозг-компьютер (ИМК) [1]. Вызванные потенциалы — это суммарная реакция головного мозга на внешние стимулы (слуховые, зрительные и соматосенсорные), когнитивные и моторные события [2]. Для выделения

вызванных потенциалов в ЭЭГ требуется усреднение однотипных событий (проб). Классификация однократных событий представляет собой сложную задачу из-за высокой вариабельности от пробы к пробе и низкого соотношения сигнал/шум [3]. В данном исследовании мы использовали несколько архитектур глубокого обучения – *ShallowConvNet*, *EEGInception*, *EEGResNet* в сравнении с линейными методами классификации (методом опорных векторов, SVM и логистической регрессией, LR) для двухклассовой дискриминации реакции на правильный и неправильный ответ математического примера. Выбранные архитектуры нейронных сетей разработаны и широко используются для классификации электроэнцефалографических данных.

*ShallowConvNet* – это легковесная нейросеть для работы с данными ЭЭГ, вдохновленная пайпланом *Filter Bank Common Spatial Patterns (FBCSP)* [4]. По аналогии с *FBCSP* первый слой в *ShallowConvNet* – сверточный слой, который выполняет частотную фильтрацию. Второй слой тоже сверточный, для пространственной фильтрации. Свертки, отвечающие за разные типы фильтраций, отличаются размерностью: в свертке для пространственной фильтрации ядро имеет размерность  $1 \times \text{количество каналов}$ , а для полосовой фильтрации:  $\text{длина фильтра} \times 1$ . Далее идет квадратичная нелинейность, слой пулинга и логарифмическая функция активации, эти слои необходимы для выбора признаков и классификации [4]. *ResNet* – это остаточная нейронная сеть. *EEGResNet* – это адаптация классической *ResNet* для анализа сырого сигнала ЭЭГ. Она отличается от классической реализации *ResNet* тем, что перед остаточными блоками добавляются два легких слоя: один для полосовой фильтрации, другой для пространственной фильтрации [из 4]. *EEGInception* – архитектура нейросети, в которой идея *Inception* блоков адаптирована для работы с данными ЭЭГ/ВП [5]. В классическом *Inception* блоке к входным применяются три сверточных слоя с ядрами (1, 1), (3, 3) и (5, 5) и слой

Max Pooling, затем результаты применения всех этих четырех слоев конкатенируются в одну матрицу. Но такая архитектура приводит к большим вычислительным затратам, поэтому к предложенной блоку добавляют сверточные слои с ядром (1, 1) после каждого слоя (кроме свертки (1, 1)) [6]. Также, в EEGInception используется Depthwise сверточный слой, в котором тензор разделяется на каналы, для каждого канала применяется свой фильтр, и результат сверток соединяется в один тензор.

Цель работы заключалась в разработке подходов к классификации сырых данных ВП в одной пробе без накопления с использованием сверточных нейронных сетей и линейных методов классификации.

## II. МЕТОДЫ

### A. Регистрация и предобработка ЭЭГ/ВП

Были классифицированы данные 20 испытуемых (подростки 14–18 лет, 17 мужчин, 3 женщины). ВП исследование проводилось с использованием парадигмы отсроченной проверки ответа [7]. Схема одной пробы представлена на рис. 1. В каждой пробе через 300 мс испытуемым предъявлялся математический пример (S1, 400 мс), а затем, на 1400 мс – решение (ответ) – правильное или неправильное (S2, 200 мс). Испытуемых просили нажимать кнопку при предъявлении правильного ответа и игнорировать неправильный ответ. Всего предъявлялось 100 проб с правильным ответом и 100 проб с неправильным ответом. ЭЭГ/ЭРП регистрировали от 19 электродов (smartBCI, ООО «Мицар», СПб) монополярно с частотой дискретизации 250 Гц, анализ данных проводили в диапазоне 0.53–30 Гц (коррекция артефактов проводилась при помощи анализа независимых компонент, ICA). Регистрация и препроцессинг данных проводились с помощью программного комплекса WinEEG (Пономарев В.А.,

Кропотов Ю.Д., № государственной регистрации 2001610516 от 08.05.2001).

### B. Извлечение признаков и классификация

Для классификации использовался сырой сигнал ЭЭГ во временном интервале после предъявления ответа (S2). Для выявления наиболее информативного временного интервала рассматривались несколько временных окон после предъявления ответа: 0–200 мс, 100–300 мс, 200–400 мс и весь интервал 0–400 мс. Обработка и классификация данных проводилась с использованием библиотек python. Данные каждого испытуемого были разделены на обучающую (80 %) и тестовую (20 %) части. В связи с индивидуальными различиями в ЭЭГ-сигнале каждая модель обучалась отдельно для каждого испытуемого. Каждая модель обучалась 10 раз с использованием кросс-валидации. Для классификации линейными методами (SVM, LR) в качестве входных данных использовались двумерные массивы с размерностью (количество проб, количество каналов\* количество временных отчетов). В качестве предварительной обработки данных использовался полосовой FIR-фильтр с полосой пропускания 0.5–15 Гц (библиотека MNE). Далее данные были нормализованы (от 0 до 1). Реализация линейных моделей была взята из библиотеки sk-learn. Для классификации с помощью CNN данные представлялись в виде 3D-массива следующей размерности (количество проб, количество каналов, количество временных отчетов). Для каждой нейронной сети подбирались learning rate и оптимизационный алгоритм из следующего списка: SGD, RMSprop, Adam или AdamW. Подбор происходил с помощью алгоритма GridSearch. Каждая модель обучалась на данных одного испытуемого 10 раз, после чего находилась медиана для 10 полученных значений. Статистическое сравнение результатов классификации проводилось с помощью ANOVA по Фридману и коэффициента согласия Кендалла в программном пакете STATISTICA10.

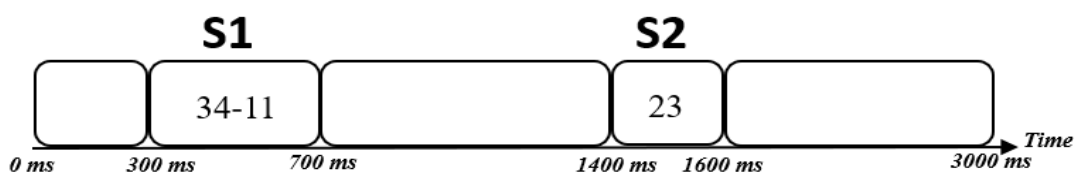


Рис. 1. Схема одной пробы

## III. РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

Средние точности классификации рассмотренных моделей (медиана для каждого человека) для двухклассовых классификаций ВП при предъявлении правильного и неправильного ответа (в интервале времени 0–400 мс после предъявления ответа) представлены в табл. I. Все сверточные нейронные сети продемонстрировали лучшую точность классификации по сравнению с линейными методами: ANOVA Chi Sq. (N = 20, df = 4)=32.6 p<0.0001, коэффициент согласия = 0.4. Среди линейных моделей, SVM показало лучшую точность классификации по сравнению с LR (тест Вилкоксона: Z= 2.9, p<0.004).

ТАБЛИЦА I. Точность классификации ВП в интервале 0–400 мс от предъявления ответа математического примера при использовании различных моделей классификаторов

Subjects	Classification accuracy (median), %				
	LR	SVM	Shallow ConvNet	EEG-Inception	EEG-ResNet
Subj.1	60.8	59.1	76.3	<b>87.5</b>	75
Subj.2	51.9	53.4	51.3	62.5	<b>67.5</b>
Subj.3	<b>57.5</b>	57.1	44.9	48.7	55.1
Subj.4	55.5	56.2	<b>67.5</b>	65	65
Subj.5	50	54.4	<b>61.5</b>	58.8	57.5
Subj.6	52.2	54.5	<b>62.8</b>	53.8	61.5
Subj.7	57.5	56.9	61.3	67.5	<b>70</b>
Subj.8	53.5	56.1	59	64.1	<b>78.2</b>

Subjects	Classification accuracy (median), %				
	LR	SVM	Shallow ConvNet	EEG-Inception	EEG-ResNet
Subj.9	56.4	56.4	<b>73.1</b>	53.8	61.5
Subj.10	56.1	55.8	56.4	59	<b>64.1</b>
Subj.11	58.3	59.9	71.8	<b>78.2</b>	69.2
Subj.12	54.8	58	69.2	<b>74.4</b>	70.5
Subj.13	52.2	54.8	<b>57.5</b>	51.3	52.5
Subj.14	52.6	62.2	71.8	<b>76.9</b>	64.1
Subj.15	55.6	59.8	71.1	68.4	<b>77.6</b>
Subj.16	53.2	59.5	72.5	<b>77.5</b>	72.5
Subj.17	53.7	56.7	<b>72.5</b>	71.3	70
Subj.18	55.3	53.1	52.5	<b>71.3</b>	55
Subj.19	52.4	58.6	71.3	<b>77.5</b>	76.3
Subj.20	50.9	52.9	61.5	64.1	64.1
Median	54.3	56.6	65.2	66.3	66.3

<sup>a</sup>. Для каждого испытуемого (Subj.) приводится медиана для 10 запусков обучения. В среднем по группе межквартильный размах составил 3.4% и не приводится в таблице. Жирным шрифтом выделена лучшая точность классификации для каждого испытуемого.

Между различными типами CNN (ShallowConvNet, EEGInception, EEGResNet) статистических различий точности классификации не обнаружено. Можно отметить, что линейные методы не смогли удовлетворительно классифицировать данные ни одного испытуемого, так как точность классификации LR и SVM не превышает 60 % (при теоретическом уровне вероятности 50 %) для всех испытуемых, кроме испытуемых №1, 14. Однако и для этих испытуемых CNN (EEGInception) дал более высокую точность классификации. Данные испытуемого №3 не были классифицированы ни одной моделью.

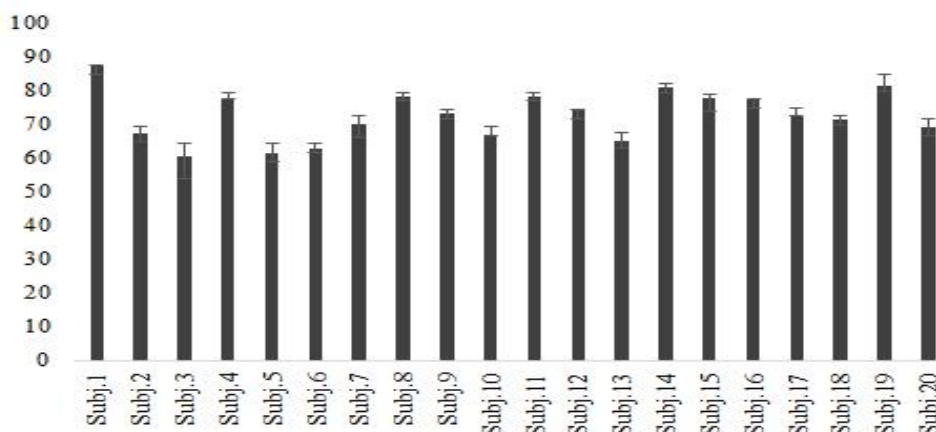


Рис. 2. Лучшая медианная точность классификации тестовой выборки для каждого испытуемого

Для большей части испытуемых наилучшая точность была достигнута нейронной сетью EEGInception, но для испытуемых №№ 2, 7, 8, 15 наивысшая точность классификации была достигнута EEGResNet, а для испытуемых №№ 4, 5, 6, 9, 13, 17 – ShallowNet. Однако все значения точности классификации, превышающие 80 %, были получены с помощью модели EEGInception. Для большинства испытуемых наиболее информативными интервалами оказались интервал 200–400 мс после предъявления стимула и общий интервал 0–400 мс, но для испытуемых № 3, 13 самая высокая точность была получена для интервала 0–200 мс, а для испытуемого № 10 – для интервала 100–300 мс.

При оценке точности классификации в коротких временных интервалах после предъявления стимула можно отметить, что точность классификации в интервалах 200–400 мс и 0–400 мс была в среднем выше, чем точность в интервалах 0–200 мс и 100–300 мс. Однако были испытуемые, для которых наилучшая точность классификации была достигнута в других интервалах: 100–300 мс были наиболее информативным интервалом для испытуемых № 10 и № 20, а 0–200 мс – для испытуемых № 3, № 9, № 13.

Для некоторых испытуемых выбор подходящего интервала значительно повысил медианную точность классификации: для испытуемого № 4 точность увеличилась на 12.5 % по сравнению с общим интервалом, а для испытуемого № 13 – на 13.7 %. Для трех испытуемых нам удалось достичь точности выше 80 %: 87.5 % на общем интервале для испытуемого № 1, 80.8 % на интервале 200–400 мс для испытуемого № 14 и 81.3% на интервале 200–400 мс для испытуемого № 20.

Выбор наилучшей модели CNN и информативного временного интервала для классификации позволил нам превзойти порог точности классификации в 60 % для каждого испытуемого. Наилучшая точность классификации для каждого испытуемого представлена на рис. 2.

#### IV. ЗАКЛЮЧЕНИЕ

Показана принципиальная возможность различения отдельных реакций (ВП) мозга на правильную и неправильную обратную связь (на примере восприятия правильного и неправильного ответа математического примера), что может быть использовано при разработке модулей обратной связи в интерфейсах мозг-компьютер.

Использование CNNs предпочтительнее для классификации амплитуд ВП по одной пробе по сравнению с линейными методами классификации. Статистически результаты классификации с использованием трех архитектур CNN (ShallowConvNet, EEGInception, EEGResNet) не отличаются при сравнении

групповых данных, но по индивидуальным данным предпочтительными моделями являются EEGInception и EEGResNet.

#### СПИСОК ЛИТЕРАТУРЫ

- [1] X. Xiao, M. Xu, J. Jin, Y. Wang, T.P. Jung, D. Ming, "Discriminative Canonical Pattern Matching for Single-Trial Classification of ERP Components", *IEEE Trans Biomed Eng.*, vol. 67, pp. 2266-2275, August 2020.
- [2] S.J. Luck. *An Introduction to the Event-Related Potential Technique*, The MIT Press, 2005.
- [3] B. Blankertz, S. Lemm, M. Treder, S. Haufe, K.R. Müller, "Single-trial analysis and classification of ERP components—a tutorial", *Neuroimage*, vol. 56, pp. 814-825, May 2011.
- [4] R.T. Schirrmester, J.T. Springenberg, L.D.J. Fiederer, M. Glasstetter, K. Eggersperger, M. Tangermann, F. Hutter, W. Burgard, T. Ball, "Deep learning with convolutional neural networks for EEG decoding and visualization", *Hum. Brain. Mapp.*, vol. 38, pp. 5391-5420, November 2017.
- [5] E. Santamaria-Vazquez, V. Martinez-Cagigal, F. Vaquerizo-Villar, R. Hornero, "EEG-Inception: A Novel Deep Convolutional Neural Network for Assistive ERP-Based Brain-Computer Interfaces", *IEEE Trans. Neural. Sys.t Rehabil. Eng.*, vol. 28, pp. 2773-2782, December 2020.
- [6] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S.E. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, A. Rabinovich, "Going deeper with convolutions" In *IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2015, Boston, MA, USA, June 7-12, 2015*, pp. 1-9.
- [7] Z.V. Nagornova, N.V. Shemyakina, S.I. Soroko, "Cognitive Event-Related Potentials in Solving Arithmetic Tasks by Adolescents Living in Different Regions of Northern Russia", *Hum. Physiol.*, vol. 46, pp. 257–263, May 2020.

# Применение больших языковых моделей для выделения намерений в целеориентированных диалоговых системах

А. Е. Шухман, В. Р. Бадиков, Л. В. Легашев

Оренбургский государственный университет

shukhman@gmail.com, sheduele.ai@ya.ru, silentgir@gmail.com

**Аннотация.** Современные методы машинного обучения в домене обработки естественного языка могут использоваться для автоматической генерации сценариев для целеориентированных диалоговых систем. В рамках текущей статьи представлена общая схема исследования автоматической генерации сценариев для целеориентированных диалоговых систем. Актуальным способом является извлечение знаний из современных языковых моделей на основе подготовленных текстовых промптов. В статье проведено исследование пяти актуальных больших языковых моделей и сделаны выводы о целесообразности применения тех или иных моделей для решения конкретных задач целеориентированных диалоговых систем.

**Ключевые слова:** большие языковые модели, промпт инжиниринг, обработка естественного языка, диалоговые системы

## I. ВВЕДЕНИЕ

Важнейшим практически значимым направлением искусственного интеллекта в последнее время стала обработка естественных языков NLP (Natural Language Processing). Одним из наиболее популярных направлений исследований является разработка диалоговых систем, способных вести диалог с пользователями, автоматически решая множество рутинных задач, связанных с уточнением информации, ответами на наиболее часто встречающиеся вопросы, решением конкретных проблем пользователя, и т. д. В настоящее время сценарии для диалоговых систем разрабатываются вручную. В то же время на основе имеющихся диалогов возможна автоматическая генерация сценариев для целеориентированных диалоговых систем с использованием современных методов машинного обучения.

Основная научная проблема исследования заключается в снижении детерминированности и ограниченности диалоговых систем за счет использования подхода к автоматической генерации возможных сценариев реакций на высказывания пользователей на основе размеченных диалоговых данных в определенной предметной области. Общая схема исследования представлена на рис. 1.

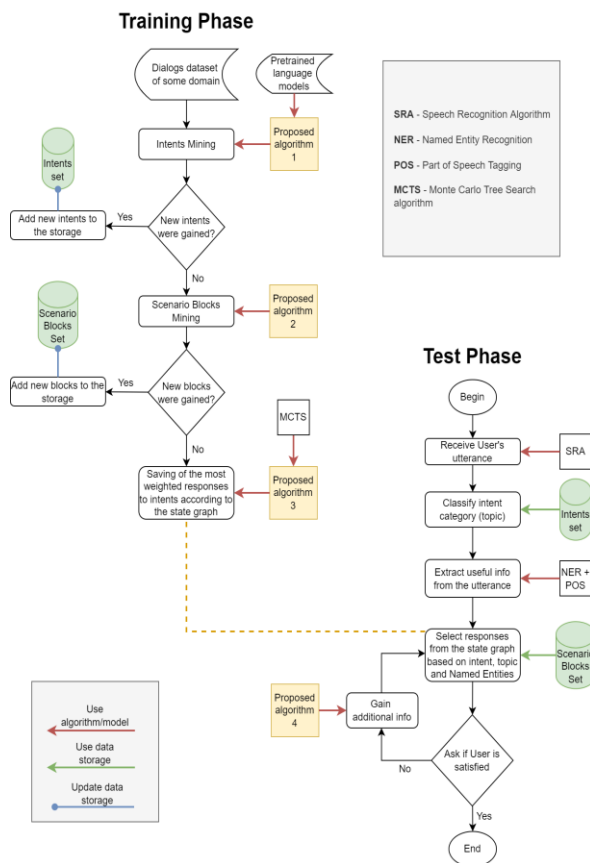


Рис. 1. Общая схема исследования автоматической генерации сценариев для целенаправленных диалоговых систем с поддержкой непрерывного обучения

На первом этапе необходимо выполнить предварительную обработку диалоговых данных. Для заданного набора диалогов строится множество намерений (intents set) (категорий намерений) пользователей, в состав которого добавляются общие (повседневные) намерения, включающие в себя интенты приветствия, прощания, согласия и отказа с репликами оператора, удовлетворенностью работы и т. д. Также строится множество сценарных блоков (scenario blocks set), в состав которого будут входить стандартные и наиболее часто используемые блоки ответов на высказывания пользователей. Будет выполнена разметка данных диалогов в соответствии с множеством намерений и множеством сценарных блоков. Далее необходимо реализовать методы для выделения тематик намерений пользователей по их высказываниям. Для

Исследование выполнено при финансовой поддержке Российского научного фонда (проект № 23-21-00503).

решения задачи автоматической генерации графа сценария будет построено дерево принятия решений, в котором вершинами будут намерения пользователей на каждом этапе диалога, а дугами - возможные вариации ответов оператора, в зависимости от категории намерения и предметной области. Планируется использовать эвристический алгоритм поиска Монте Карло, который позволит выделить несколько взвешенных переходов к ответам операторов на основе обученных диалоговых данных выбранной прикладной области. Для актуализации алгоритмов генерации и преодоления дрейфа модели будет реализовано непрерывное машинное обучение, использующее содержательные метаданные с автоматической переоценкой весов моделей. На основе сохранения и анализа диалогов в процессе эксплуатации системы планируется проводить дообучение моделей, включая выделение новых намерений и сценарных блоков, коррекции графа сценария и модели контекста диалога.

## II. ОБЗОР ИССЛЕДОВАНИЙ

На текущий момент времени проблема автоматического формирования сценариев для целеориентированных диалоговых систем является новой для исследований в области обработки естественных языков. При этом множество исследований связаны с генерацией ответов на вопросы пользователя в форме осмысленного диалога. В работе [1] представлен условный вариативный фреймворк для генерации диалогов, на основе полуавтоматической глубокой генеративной модели. В работе [2] описывается бенчмарк генерации естественного языка для симуляции настроек обучения на малом количестве данных в проблемно-ориентированных диалоговых системах. С целью улучшения генерации ответов в диалоговых системах авторы статьи [3] предлагают фреймворк, который использует один и тот же кодировщик для извлечения общих и не зависящих от проблемы признаков с помощью разных декодеров для изучения признаков, специфичных для проблемы. В статье [4] предлагается новый фреймворк, который использует обучение с подкреплением для улучшения качества генерации диалогов в проблемно-ориентированных диалоговых системах. Авторы исследования [5] генерируют ответы на основе извлечения знаний из неструктурированных документов для проблемно-ориентированных диалоговых систем. В работе [6] представлен фреймворк многоцелевой аугментации данных, который позволяет генерировать множество разнообразных подходящих ответов диалога, которые приводят к достижению цели диалога.

Эвристический алгоритм поиска Монте-Карло (Monte Carlo Tree Search, MCTS) используется для принятия решений в классах задач, для которых на каждом ходу существует множество возможных исходов. Указанный алгоритм можно адаптировать для решения задачи построения диалоговой системы, в которой для каждого намерения пользователя можно подобрать один или несколько взвешенных переходов к ответу системы. Небольшая часть исследований посвящена этому вопросу. В статье [7] алгоритм MCTS используется в связке с моделями глубокого обучения

для решения задачи построения диалога при бронировании билетов в кино. В работе [8] представлен Байес-адаптивный алгоритм планирования для целеориентированных диалоговых систем с использованием рекуррентных нейронных сетей и алгоритма MCTS. В публикации [9] алгоритм MCTS используется для оценки качества сгенерированной последовательности диалогов в состязательных сетях, генерирующих последовательности диалоговых систем. В статье [10] автор выполняет поиск по пространству состояний ответов в диалогах малой и средней длины с использованием алгоритма MCTS.

Совсем недавно появилось большое количество статей о многообещающем будущем промпт инжиниринга для решения прикладных задач. Выходят обзоры по применению больших языковых моделей [11, 12], рассматриваются практические случаи для классификации материалов [13], автоматизации воспроизведения багов из отчетов об ошибках [14], социального анализа прогнозирования рыночных тенденций [15] и др.

## III. МЕТОДЫ ВЫДЕЛЕНИЯ НАМЕРЕНИЙ

Классический способ выделения намерений пользователей заключается в подготовке размеченного датасета на основе диалогов пользователя и оператора и обучении классификатора с целью мульти-классовой классификации. Более актуальным способом является использование больших языковых моделей LLM для извлечения информации из неструктурированного текста посредством технологии промпт инжиниринга. Рассмотрим оба подхода к решению поставленной задачи.

### A. Методы классификации намерений для целеориентированных диалоговых систем

В качестве практического примера мы рассмотрели англоязычный датасет MultiWOZ 2.2 (Multi-Domain Wizard-of-Oz) [16], содержащий письменные разговоры между людьми на тематику поиска и бронирования отелей, ресторанов, поездов, достопримечательностей и т.д. Представленные данные содержат поочередный диалог между пользователем (USER) и системой (SYSTEM), каждый шаг диалога (turn) представляет собой одно высказывание (utterance) пользователя или системы (например, *'I'm looking for a local place to dine in the centre that serves chinese food.'*). Всего в наборе данных MultiWOZ 2.2 8 438 диалогов. Полученный после выполнения операций датасет содержит два признака – высказывание и намерение ('sentence', 'intent'). Удалим из итогового датасета записи с интенциями поиска госпиталя (find\_hospital) и поиска автобуса (find\_bus) по причине их малого количества. Итоговый датасет содержит девять классов. В категорию NONE полученного датасета попали все промежуточные реплики пользователей во время диалогов.

Для решения задачи многоклассовой классификации будут рассмотрены два подхода. Первый подход заключается в использовании логистической регрессии, которая основана на линейном разделении объектов и преобразует текст в его векторное представление.

Второй подход заключается в использовании англоязычных моделей трансформеров BERT, который основан на дообучении предварительно обученной нейронной сети с дополненными слоями классификатора на размеченном наборе данных. Для логистической регрессии на первом этапе выполняется числовое кодирование целевой переменной – названия десяти категорий высказываний. Задается словарь стоп-слов из англоязычного корпуса библиотеки nltk и задается минимальная и максимальная длина n-грам от 1 до 3. На следующем этапе объект CountVectorizer использует модель мешка слов (bag-of-words), формируя словарь n-грамм длины m, и каждый текст представляется вектором длины m, в котором каждый элемент соответствует количеству вхождений соответствующей n-граммы в текст. Объект TfidfVectorizer вместо количества вхождений для n-граммы сохраняет коэффициент TF-IDF. Для модели трансформера BERT на первом этапе также выполняется числовое кодирование целевой переменной. Задается максимальный размер словаря num\_words = 15000 и максимальная длина высказывания пользователя max\_len = 200 в токенах, происходит выравнивание предложений исходного датасета до одинаковой длины (padding='post'). На следующем этапе выполняется токенизация обучающей выборки с помощью моделей bert-base-uncased и xlm-roberta-base. Векторные представления формируются с помощью входного слоя нейронной сети на основе списка словарных номеров текстовых токенов. Функция softmax библиотеки torch используется для получения предсказанной вероятности принадлежности выборки к одной из девяти категорий высказываний. Результаты сравнения двух исследуемых подходов классификации высказываний пользователей по метрикам Precision, F1-score и Matthews correlation coefficient (MCC) представлены в табл. 1.

ТАБЛИЦА 1. СРАВНЕНИЕ МЕТРИК

Approach	Metrics values		
	Precision	F1-score	MCC
Logistic Regression and CountVectorizer	0.7440	0.7393	0.6806
Logistic Regression and TfidfVectorizer	0.7535	0.7508	0.6951
BERT with bert-base-uncased model	0.7962	0.7798	0.7392
BERT with xlm-roberta-base model	0.7716	0.7602	0.7159

Подход на основе трансформеров BERT с использованием модели bert-base-uncased показал лучшие результаты по трём метрикам Precision (0,80), F1-score (0,78) и Matthews correlation coefficient (0,74) в сравнении с другими методами.

**B. Выделение намерений с помощью промпт-инжиниринга и больших языковых моделей**

Языковая модель ChatGPT 3.5 разработана компаний Open AI на основе дообучения модели GPT 3.0, содержащей 175 миллиардов параметров. Языковая модель mistral-Saiga базируется на большой языковой модели (LLM) Mistral-7B-v0.1, которая представляет собой предварительно обученную генеративную текстовую модель с 7 миллиардами параметров.

Русскоязычная языковая модель GigaChat базируется на большой языковой модели ruGPT-3.5-13B с 13 миллиардами параметров. Языковая модель Yandex GPT (YaLM 100B) использует 100 миллиардов параметров. Языковая модель Google Bard получена дообучением модели PaLM, содержащей 540 миллиардов параметров. Примеры выделения интенгов и именованных сущностей для модели ChatGPT представлены на рис. 2.

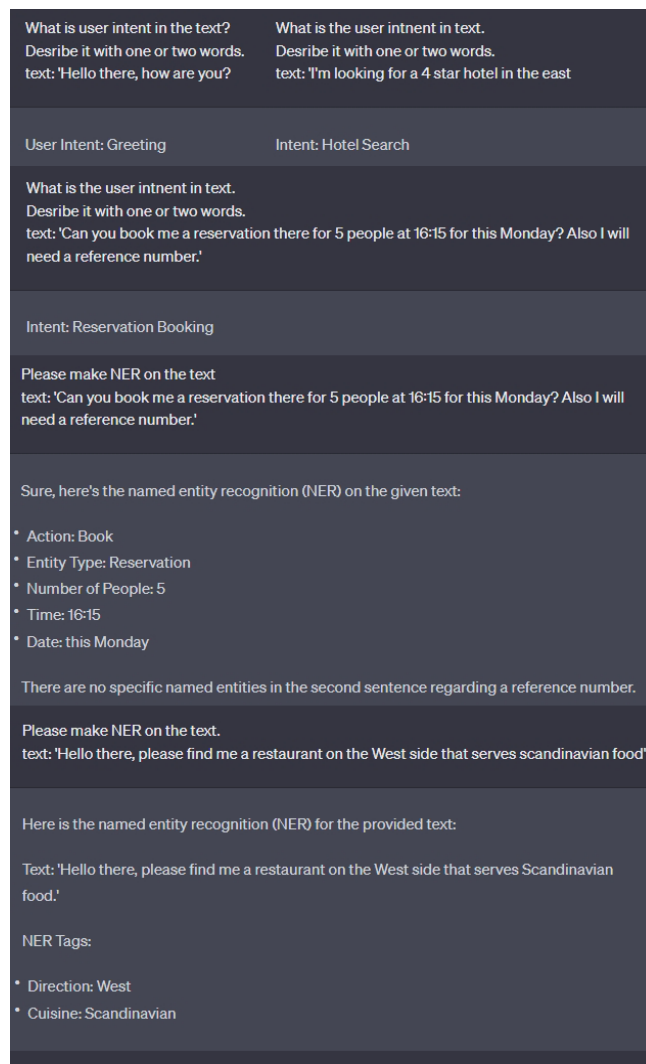


Рис. 2. Примеры извлечения намерений и именованных сущностей моделью ChatGPT

В текущей статье для сравнения результатов работы LLM мы использовали следующие промпты для сравнения англоязычных и русскоязычных запросов к большим языковым моделям GhatGPT, Mistral, GigaChat, Yandex GPT и Google BARD:

- Промпт выделения намерений пользователя/сценарного блока ответа оператора на английском языке: *What is the user intent in the text? Describe it with one or two words. text: "[input]"*.
- Промпт выделения именованных сущностей в реплике пользователя/оператора на английском языке: *Please, make NER on the text: text: "[input]"*.



- Промпт выделения намерений пользователя/сценарного блока ответа оператора на русском языке: *Укажи намерение пользователя в тексте. Опиши намерение пользователя одним или двумя словами text: “[input]”*
- Промпт выделения именованных сущностей в реплике пользователя/оператора на русском языке: *Выполни распознавание именованных сущностей в тексте text: “[input]”*.

Выполним сравнение функционала англоязычных и русскоязычных моделей для решения поставленной задачи. Будем выполнять оценку того, насколько четко модель выделяет намерения пользователей/сценарные блоки ответа оператора, насколько четко выделяет именованные сущности, в каком виде представляется ответ для последующего парсинга, насколько ответ языковой модели устойчив к опечаткам в тексте промпта, а также можно ли развернуть модель локально. Результаты сравнений проводились на 150 подготовленных интентах на русском и английском языках. Обращение к языковым моделям выполнялось в декабре 2023 г.

Субъективное сравнение языковых моделей на тестовых промптах показывает, что для выделения намерений пользователей/сценарных блоков ответа оператора для русского языка лучшие результаты показывает языковая модель Yandex GPT, для английского языка лучшие результаты показывает языковая модель BARD. Для выделения именованных сущностей для русского и английского языков лучшие результаты показывает языковая модель ChatGPT. У языковой модели BARD наиболее четко структурированный ответ, большинство исследуемых языковых моделей показывают хорошую устойчивость к опечаткам и перестановкам токенов в промпте. Модель Mistral-Saiga может быть развернута локально, что является преимуществом с точки зрения обработки персональных и конфиденциальных данных. Наилучшие результаты попарного сравнения по метрике SBS показала языковая модель ChatGPT. Для детерминированности результата вывода языковой модели можно использовать различные методы синтаксической близости предложений, например, расширение python-Levenshtein.

#### IV. ЗАКЛЮЧЕНИЕ

В статье выполнено исследование двух подходов выделения интенгов и сценарных блоков ответов оператора для целеориентированных диалоговых систем на основе обучения моделей трансформеров и извлечения знаний из современных языковых моделей на основе подготовленных текстовых промптов. Подход на основе мультиклассовой классификации оказался неэффективным для решения поставленных задач по причине ограниченного количества классов для мультиклассовой классификации, а также необходимости обучения/дообучения датасета для каждой новой диалоговой тематики. Подход на основе

использования больших языковых моделей обладает большим потенциалом и может применяться для структурированного выделения намерений пользователей, сценарных блоков ответа оператора, а также именованных сущностей в произвольно структурированных неразмеченных текстах.

#### СПИСОК ИСТОЧНИКОВ

- [1] X. Shen, H. Su, Y. Li, W. Li, S. Niu, Y. Zhao, A. Aizawa and G. Long, “A conditional variational framework for dialog generation,” Association for Computational Linguistics (ACL), 2017, pp. 504–509.
- [2] B. Peng, C. Zhu, C. Li, X. Li, J. Li, M. Zeng and J. Gao, “Few-shot Natural Language Generation for Task-Oriented Dialog”, Dialog, vol. 1., 2020, pp.1–12.
- [3] X. Chen, Z. Cui, J. Zhang, C. Wei, J. Cui, B. Wang, D. Zhao and R. Yan, “Reasoning in Dialog: Improving Response Generation by Context Reading Comprehension” Proceedings of the AAAI Conference on Artificial Intelligence, vol. 35, № 14, 2021, pp. 12683–12691.
- [4] M. Zhou, J. Arnold and Z. Yu, “Building task-oriented visual dialog systems through alternative optimization between dialog policy and language generation” Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), 2019, pp. 143–153.
- [5] D. Thulke, N. Daheim, C. Dugast and H. Ney, “Efficient Retrieval Augmented Generation from Unstructured Knowledge for Task-Oriented Dialog” arXiv preprint arXiv:2102.04643, 2021.
- [6] Y. Zhang, Z. Ou and Z. Yu, “Task-oriented dialog systems that consider multiple appropriate responses under the same context” Proceedings of the AAAI Conference on Artificial Intelligence, vol. 34, № 5, 2020, pp. 9604–9611.
- [7] S. Wang, K. Zhou, K. Lai and J. Shen “Task-completion dialogue policy learning via Monte Carlo tree search with dueling network” Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP), 2020, pp. 3461–3471.
- [8] Y. Jang, J. Lee and K.E. Kim “Bayes-adaptive monte-carlo planning and learning for goal-oriented dialogues” Proceedings of the AAAI Conference on Artificial Intelligence, vol. 34, № 5, 2020, pp. 7994–8001.
- [9] Y.L. Tuan and H.Y. Lee “Improving conditional sequence generative adversarial networks by stepwise evaluation” IEEE/ACM Transactions on Audio, Speech, and Language Processing, 2019, vol. 27, № 4, pp. 788–798.
- [10] S. Mukherjee “An Unsupervised Approach to Automatic Response Generation for Conversational e-Commerce Agents using Monte Carlo Tree Search”, unpublished.
- [11] Sahoo P., Singh A.K., Saha S., Jain V., Mondal S., & Chadha A. (2024). A Systematic Survey of Prompt Engineering in Large Language Models: Techniques and Applications. arXiv preprint arXiv:2402.07927.
- [12] Li H., Leung J., & Shen Z. (2024). Towards Goal-oriented Large Language Model Prompting: A Survey. arXiv preprint arXiv:2401.14043.
- [13] Liu S., Wen T., Pattamatta A. S. L., & Srolovitz D. J. (2024). A Prompt-Engineered Large Language Model, Deep Learning Workflow for Materials Classification. arXiv preprint arXiv:2401.17788.
- [14] Feng S., & Chen C. (2024, February). Prompting Is All You Need: Automated Android Bug Replay with Large Language Models. In Proceedings of the 46th IEEE/ACM International Conference on Software Engineering (pp. 1-13).
- [15] Wahidur R.S., Tashdeed I., Kaur M., & Lee H.N. (2024). Enhancing Zero-Shot Crypto Sentiment with Fine-tuned Language Model and Prompt Engineering. IEEE Access.
- [16] Zang X. et al. Multiwoz 2.2: A dialogue dataset with additional annotation corrections and state tracking base-lines // arXiv preprint arXiv:2007.12720. 2020.

# Объединение глубокого обучения и объяснимого ИИ для неинвазивного прогнозирования мутаций EGFR и KRAS в NSCLC: новый радиогеномный подход

Фаридоддин Шариати<sup>1</sup>, В. А. Павлов<sup>1</sup>, С. В. Федяшина<sup>1</sup>, Н. А. Серебренников<sup>1,2</sup>

<sup>1</sup>Институт электроники и телекоммуникаций  
Санкт-Петербургский политехнический университет Петра Великого

<sup>2</sup>Санкт-Петербургский научно-исследовательский институт неотложной медицины имени И.И. Джанелидзе

Email: shariaty3@gmail.com, pavlov\_va@spbstu.ru, fedyashina.sv@edu.spbstu.ru

**Аннотация.** Диагностика рака легких основывается на фенотипических и генетических данных, что создает основу для разработки персонализированных стратегий лечения. Интеграция методов объяснимого ИИ (ХАИ) в моделирование предсказаний генных мутаций при раке легких представляет собой важнейшее достижение в области интерпретируемости и достоверности алгоритмов глубокого обучения. В работе Grad-CAM в сочетании с моделью ResNet18 используется для выявления признаков на КТ-изображениях легочных узелков, которые указывают на конкретные генетические мутации, такие как EGFR и KRAS. Результаты ХАИ, в частности тепловые карты Grad-CAM, служат связующим звеном между предсказаниями модели глубокого обучения и клиническим пониманием морфологии опухоли легкого и ее генетической основы. Это исследование не только улучшает объяснимость моделей ИИ, но и способствует более детальному пониманию взаимосвязи между морфологическими признаками и генетическими вариациями, что потенциально может привести к более целенаправленному и персонализированному лечению пациентов с раком легких.

**Ключевые слова:** рак лёгких, ХАИ, Deep Learning, KRAS, EGFR

## I. ВВЕДЕНИЕ

Рак легких остается основной причиной смертности от онкологических заболеваний во всем мире, при этом немелкоклеточный рак легкого (НМРЛ) составляет около 85 % всех случаев заболевания [1]. Гетерогенность НМРЛ на молекулярном уровне представляет значительные трудности для диагностики, прогноза и лечения, что подчеркивает необходимость применения подходов персонализированной медицины. В последние годы выявление генных мутаций, способствующих распространению рака, таких как рецептор эпидермального фактора роста EGFR и онкоген KRAS, произвело революцию в стратегиях лечения [2]. Однако инвазивный характер традиционных методов биопсии ограничивает их повторное применение, подчеркивая необходимость неинвазивных методов диагностики.

Радиогеномика, пересечение радиологии и геномики, предлагает многообещающий путь, связывая фенотипы визуализации с экспрессией генов, что потенциально позволяет неинвазивно охарактеризовать опухоль.

Предыдущие исследования в области радиогеномики в основном были направлены на соотнесение специфических признаков визуализации с генетическими изменениями [3]. В исследованиях использовались различные модели машинного обучения для предсказания генных мутаций на основе рентгенографических изображений, демонстрируя потенциал искусственного интеллекта в повышении точности диагностики и персонализации лечения. Например, работы с использованием сверточных нейронных сетей (СНС) показали значительные перспективы в определении подтипов рака легких и предсказании генетических мутаций по КТ [4]–[6]. Более того, появление более сложных моделей, таких как Vision Transformer (ViT) [7], открыло новые пути для анализа медицинских изображений, предлагая преимущества в захвате глобальных признаков изображения. Однако, несмотря на эти достижения, остается критическая проблема интерпретируемости моделей искусственного интеллекта (ИИ), которая имеет решающее значение для клинического принятия и понимания биологической значимости предсказаний, сделанных этими моделями [8], [9].

Хотя существующие исследования достигли значительного прогресса в предсказании генных мутаций по данным визуализации, остается заметный пробел в интеграции методов объяснимого ИИ (explainable AI, ХАИ). Большинство предыдущих работ дают ограниченное представление о том, как модели делают прогнозы, что является препятствием для их внедрения в клиническую практику. Кроме того, не хватает комплексных исследований, использующих последние достижения в архитектуре глубокого обучения и ХАИ для повышения интерпретируемости прогнозов генных мутаций на основе КТ-изображений легких. Данная работа направлена на восполнение этого пробела путем использования Градиентно-взвешенное отображения активаций классов (Gradient-weighted Class Activation Mapping, Grad-CAM) и ResNet18, фокусируясь

Исследование выполнено за счет гранта Российского научного фонда № 24-25-00204, <https://rscf.ru/project/24-25-00204/>.

на интерпретируемости прогнозов и предоставляя понимание процесса принятия решений моделью [10]–[12].

В этой статье представлен новый подход к прогнозированию мутаций EGFR и KRAS у пациентов с НСКЛ на основе радиогеномного набора данных, который объединяет модели глубокого обучения с ХАИ для повышения точности и интерпретируемости прогнозов. В работе используется архитектура глубокого обучения ResNet18 [13], известная своей эффективностью в задачах классификации изображений, в сочетании с Grad-CAM, методом ХАИ, который предоставляет визуальные объяснения для прогнозов модели. Таким образом, достигается высокая эффективность прогнозов и раскрывается зависимость модели от конкретных признаков изображения, предлагая идеи о биологическом и клиническом значении этих признаков. Предложенная комплексная методология, включающая сбор данных, реализацию модели и применение методов ХАИ, открывает путь к более глубокому пониманию связи между фенотипами изображений и генетическими мутациями, что позволит достигнуть новых результатов в области персонализированной медицины при лечении рака легкого.

## II. МЕТОДОЛОГИЯ

### A. Получение данных

В исследовании использован радиогеномный набор данных больных НМРЛ, состоящий из 211 человек. Этот набор объединяет различные формы диагностических и прогностических данных, включая компьютерную томографию (КТ), позитронно-эмиссионную томографию (ПЭТ)/КТ-изображения, семантические аннотации с использованием контролируемого словаря и карты сегментации опухолей на КТ-изображениях. Кроме того, количественные показатели, полученные с помощью ПЭТ/КТ-изображений, сочетаются с подробным геномным анализом, включающим данные о мутациях генов и секвенировании РНК, которые были получены из опухолевой ткани. Этот набор данных дополнительно обогащен соответствующими клиническими данными, такими как результаты выживания, что создает надежную основу для изучения корреляции между геномными признаками и медицинской визуализацией [14].

### B. Теоретическая основа и ХАИ

Интеграция методов ХАИ в моделирование предсказания генных мутаций при раке легких представляет собой важнейшее достижение в области интерпретируемости и достоверности алгоритмов глубокого обучения. В частности, в данной работе используется Grad-CAM в сочетании с моделью ResNet18 для выявления признаков КТ-изображений легочных узелков, которые указывают на специфические генные мутации, такие как EGFR и KRAS.

1) *ResNet18 для классификации мутаций*: ResNet18, входящая в семейство Residual Networks, известна своей способностью эффективно обучаться на изображениях благодаря использованию остаточных связей, которые смягчают проблему исчезающего градиента в глубоких

архитектурах. Для классификации мутаций ResNet18 была обучена на наборе данных КТ-изображений легких, причем последние слои были адаптированы для вывода прогнозов, касающихся именно наличием мутаций EGFR и KRAS. Архитектура модели позволяет извлекать из КТ-изображений сложные паттерны и признаки, которые свидетельствуют о мутациях [15]:

$$Y = f_{ResNet18}(X; \theta), \quad (1)$$

где  $Y$  – результат классификации мутаций,  $X$  – входное КТ-изображение, а  $\theta$  – параметры модели ResNet18.

2) *Grad-CAM для интерпретации*: Чтобы повысить интерпретируемость прогнозов модели ResNet18, после классификации используется Grad-CAM. Grad-CAM формирует тепловую карту, визуализирующую области на входном КТ-изображении, которые существенно влияют на прогноз модели. Это достигается путем вычисления градиентов целевого класса (т. е. конкретной генной мутации) по отношению к картам признаков последнего свёрточного слоя модели ResNet18, таким образом, выделяя области интереса, которые приводят к определенному классификационному решению [16].

Тепловая карта Grad-CAM  $H$  для класса  $c$  вычисляется следующим образом:

$$H^c = ReLU \left( \sum_k \alpha_k^c \cdot A^k \right) \quad (2)$$

где  $A^k$  представляет собой карты признаков последнего свёрточного слоя,  $\alpha_k^c$  – веса, вычисленные путем глобального усреднения градиентов оценки класса относительно  $A^k$ , а  $ReLU$  применяется для фокусировки на признаках, которые оказывают положительное влияние на интересующий класс.

3) *Применение Grad-CAM для интерпретации классификации мутаций*: Применяя Grad-CAM к задаче классификации мутаций, можно визуально интерпретировать, какие анатомические особенности и паттерны внутри легочного узелка вносят наиболее существенный вклад в прогноз модели. Это не только дает представление о процессе принятия решения моделью, но и помогает радиологам и онкологам соотнести специфические особенности изображения с генетическими мутациями, способствуя более глубокому пониманию взаимосвязи между фенотипическими проявлениями и генотипическими вариациями при раке легкого.

Реализация Grad-CAM в сочетании с ResNet18 для классификации генных мутаций воплощает в себе значительный шаг на пути к объяснимому и интерпретируемому ИИ в медицинской области. Она гарантирует, что предсказания, сделанные моделями глубокого обучения, будут не только точными, но и понятными, что повысит доверие к диагностике с помощью ИИ и ее полезность в персонализированной медицине.

### III. РЕЗУЛЬТАТЫ

#### A. Прогнозирование генных мутаций

В этом разделе описывается прогностический анализ, проведенный для выявления генных мутаций, связанных с раком легкого, а именно мутаций EGFR и KRAS, с помощью анализа изображений, полученных с помощью КТ. Для этого использовались две модели глубокого обучения: Vision Transformer (ViT) [7] и ResNet18 [13]. Эти модели были выбраны благодаря их эффективности в задачах классификации изображений, предлагая новый подход к предсказанию генных мутаций в контексте рака легких.

Набор данных для анализа был разделен на обучающий и тестовый наборы, а распределение мутантных и диких случаев для мутаций EGFR и KRAS представлено в табл. 1 и 2, соответственно.

ТАБЛИЦА I. РАСПРЕДЕЛЕНИЕ ДАННЫХ О МУТАЦИЯХ EGFR

Класс	Train	Test
Mutant	370	159
Wildtype	1495	641

ТАБЛИЦА II. РАСПРЕДЕЛЕНИЕ ДАННЫХ О МУТАЦИЯХ KRAS

Класс	Train	Test
Mutant	407	175
Wildtype	1486	637

Результаты классификации ViT и ResNet18 в задачах предсказания мутаций EGFR и KRAS показаны в табл. 3 и 4, соответственно. В этих таблицах представлены показатели *Accuracy*, *Precision*, *Recall*, *F1-score* и площадь под кривой (*AUC*), достигнутые каждой моделью, что подчеркивает их возможности в классификации генных мутаций по КТ-изображениям.

ТАБЛИЦА III. ПОКАЗАТЕЛИ ЭФФЕКТИВНОСТИ ДЛЯ ПРЕДСКАЗАНИЯ МУТАЦИЙ EGFR

Method	Accuracy	Precision	Recall	F1-score	AUC
ViT	0.96	0.95	0.92	0.93	0.92
ResNet18	0.97	0.98	0.93	0.95	0.93

TABLE I. ПОКАЗАТЕЛИ ЭФФЕКТИВНОСТИ ДЛЯ ПРЕДСКАЗАНИЯ МУТАЦИЙ KRAS

Method	Accuracy	Precision	Recall	F1-score	AUC
ViT	0.97	0.95	0.92	0.95	0.95
ResNet18	0.98	0.98	0.97	0.97	0.97

Анализ показал, что ResNet18 превосходит ViT в прогнозировании мутаций EGFR и KRAS, получая более высокие показатели по всем оцениваемым параметрам. Это подчеркивает потенциал использования передовых архитектур глубокого обучения для точной идентификации генетических мутаций по результатам КТ легких, что вносит значительный вклад в развитие персонализированной медицины в лечении рака легких.

#### B. Объяснимый искусственный интеллект (XAI)

Применение методов XAI, в частности Grad-CAM в сочетании с моделью ResNet18, позволило получить убедительные сведения об интерпретационных механизмах, лежащих в основе классификации легочных узелков на основе генных мутаций. Визуализация Grad-CAM сыграла ключевую роль в раскрытии этих идей,

выявив специфические особенности изображения, которые модель считает значимыми для своих прогнозов.

Тепловые карты Grad-CAM, созданные после классификации ResNet18, ярко выделяют края легочных узелков красным цветом, указывая на то, что эти области имеют большое влияние на процесс принятия решения моделью. Это наблюдение совпадает со многими проанализированными КТ-изображениями, что позволяет предположить наличие закономерности в том, как модель различает различные генные мутации (EGFR и KRAS) в легочных узелках.

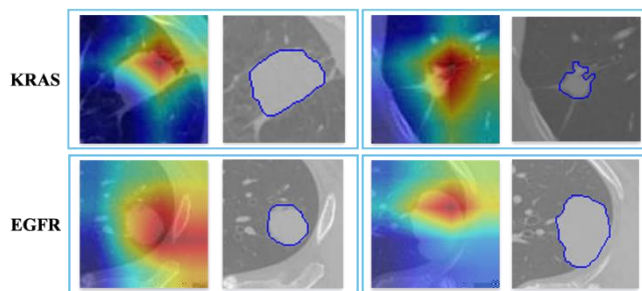


Рис. 1. Визуализация Grad-CAM, выделяющая края узлов красным цветом для классификации генных мутаций. Эти тепловые карты подчеркивают, что модель опирается на краевые особенности для различения мутантных и диких генотипов

Результаты XAI, в частности тепловые карты Grad-CAM, служат связующим звеном между предсказаниями модели глубокого обучения и клиническим пониманием характеристик легочных узелков, связанных с генными мутациями. Сосредоточив внимание на краях узелков, модель отражает клиническую практику, в которой рентгенологи часто тщательно изучают форму, неровности границ и текстуру узелков на предмет признаков, указывающих на злокачественность или специфические генетические изменения.

Анализ этих XAI-результатов подчеркивает важность включения объяснимости в решения для здравоохранения, основанные на ИИ, способствуя прозрачному и интерпретируемому подходу к диагностике рака легких и предсказанию генных мутаций. Детальный анализ визуализаций Grad-CAM открывает путь к более глубокой совместной работе между технологиями ИИ и клиническим опытом, повышая точность и персонализацию стратегий лечения рака.

Выводы, сделанные на основе визуализации Grad-CAM, особенно акцент модели на краях легочных узелков, позволяют выдвинуть еще одну гипотезу, которая заслуживает детального изучения. Выделенные края, идентифицированные моделью ResNet18, могут не только очерчивать границы узелков, но и указывать на наличие сосудистых структур, критически важных для понимания микроокружения опухоли. Это наблюдение позволяет предположить, что сосуды по краям узелков могут играть важную роль в процессе классификации модели, потенциально коррелируя с агрессивностью генных мутаций (EGFR и KRAS) или ангиогенной активностью опухоли.

Потенциальная корреляция между сосудами, подсвеченными по краям, и генными мутациями может предложить новое измерение в диагностической оценке рака легких. Можно предположить, что определенные сосудистые особенности, заметные по краям легочных узелков, свидетельствуют о специфических генетических изменениях, предоставляя тем самым неинвазивный метод определения биологии опухоли. Этот подход соответствует принципам радиогеномики, которая стремится соотнести фенотипы визуализации с генетическими данными, повышая точность диагностики рака легкого.

Изучение взаимосвязи между характеристиками краев узелков, в частности сосудистыми структурами, и генными мутациями позволяет еще больше сократить разрыв между анализом на основе ИИ и клиническими исследованиями. Такое исследование не только повышает объяснимость моделей ИИ, но и способствует более глубокому пониманию морфологии легочных узелков и их генетических основ, что потенциально может привести к разработке более целенаправленных и персонализированных стратегий лечения пациентов с раком легкого.

#### IV. ЗАКЛЮЧЕНИЕ

В этой статье представлен подход для прогнозирования мутаций EGFR и KRAS у пациентов с НМРЛ, использующий интеграцию моделей глубокого обучения с объяснимыми методами искусственного интеллекта. Благодаря использованию ResNet18 в сочетании с Grad-CAM была достигнута высокая эффективность при идентификации мутаций генов по КТ-изображениям, при этом модели достигают Accurasy до 0,98, Precision 0,98 и Recall 0,97 для прогнозирования мутаций KRAS. и столь же высокие показатели для прогнозирования мутаций EGFR. Эти результаты не только демонстрируют возможность использования глубокого обучения для идентификации мутаций генов, но также подчеркивают важность интерпретируемости этих прогнозов с помощью объяснимых методов искусственного интеллекта.

Интеграция ХАИ дала информацию о процессах принятия решений в наших моделях, в частности подчеркнув значимость определенных признаков визуализации для прогнозирования генных мутаций. Это достижение устраняет разрыв между вычислительной мощностью моделей ИИ и клинической необходимостью того, чтобы их решения были понятными и заслуживающими доверия. Кроме того, результаты этого исследования открывают путь для будущих исследований в области радиогеномики, предлагая направления для разработки более совершенных моделей ИИ, способных более эффективно обрабатывать сложные и неоднородные данные о раке.

По мере продвижения вперед крайне важно продолжать изучение и совершенствование этих методологий ИИ и объяснимого ИИ, обеспечивая их применимость и эффективность в реальных клинических условиях. Сотрудничество специалистов по обработке данных, радиологов, онкологов и генетиков будет иметь решающее значение для развития этой области и реализации всего потенциала ИИ и радиогеномики в

преобразовании диагностики и лечения рака легких. Эта статья представляет собой шаг на пути к этому будущему, предлагая взглянуть на возможности, которые лежат на стыке ИИ, радиологии и геномики для улучшения лечения пациентов с раком легких.

#### СПИСОК ЛИТЕРАТУРЫ

- [1] F. Shariaty, S. Shariati, S. Navvabi, M. N. Oshnari, and B. Novikov, "Application of artificial intelligence for rapid prevention of epidemic diseases (covid-19)," 2022.
- [2] M. Skalunova, F. Shariaty, S. Rozov, and A. R. Radmard, "Personalized chemotherapy selection for lung cancer patients using machine learning and computed tomography," in 2023 International Conference on Electrical Engineering and Photonics (EExPolytech), pp. 128–131, IEEE, 2023.
- [3] F. Shariaty, L. Duan, V. Pavlov, M. Mousavi, and T. Pervunina, "A novel gene assay combined with medical imaging for accurate prognosis and prediction of cancer type," in 2022 International Conference on Electrical Engineering and Photonics (EExPolytech), pp. 118–121, IEEE, 2022.
- [4] V. A. Pavlov, F. Shariaty, M. Orooji, and E. N. Velichko, "Application of deep learning techniques for detection of covid-19 using lung ct scans: model development and validation," in International Youth Conference on Electronics, Telecommunications and Information Technologies: Proceedings of the YETI 2021, St. Petersburg, Russia, pp. 85–96, Springer, 2022.
- [5] D. Taranova and F. Shariaty, "Radiomic analysis for prediction of t stage parameter (t1-t2) in lung cancer patients," in , pp. 77–80, 2022.
- [6] F. Shariaty, H. Caiqin, V. A. Pavlov, L. Duan, S. V. Zavyalov, T. M. Pervunina, and W. Ying, "Integrating quantitative and convolutional features to enhance the efficiency of pathology classification in CT imaging," , vol. 16, no. 4, pp. 60–69, 2023.
- [7] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, and N. Houlsby, "An image is worth 16x16 words: Transformers for image recognition at scale," ArXiv, vol. abs/2010.11929, 2020.
- [8] Z. Chen, S. Gao, C. Ding, T. Luo, J. Xu, S. Xu, and S. Li, "CT-based non-invasive identification of the most common gene mutation status in patients with non-small cell lung cancer," Medical Physics, 2023.
- [9] L. Yang, Z. Li, D. W. Binzel, P. Guo, and T. M. Williams, "Targeting oncogenic kras in non-small cell lung cancer with EGFR aptamer-conjugated multifunctional rna nanoparticles," Molecular Therapy-Nucleic Acids, vol. 33, pp. 559–571, 2023.
- [10] X. Tang, Y. Li, L.-T. Shen, W.-F. Yan, W.-L. Qian, and Z.-G. Yang, "Ct radiomics predict egfr-t790m resistance mutation in advanced non-small cell lung cancer patients after progression on first-line egfr-tki," Academic Radiology, vol. 30, no. 11, pp. 2574–2587, 2023.
- [11] X. Lv, Y. Li, B. Wang, Y. Wang, Z. Xu, and D. Hou, "Multisequence mri-based radiomics signature as potential biomarkers for differentiating kras mutations in non-small cell lung cancer with brain metastases," European Journal of Radiology Open, vol. 12, p. 100548, 2024.
- [12] C. Ladbury, A. Amini, A. Govindarajan, I. Mambetsariyev, D. J. Raz, E. Massarelli, T. Williams, A. Rodin, and R. Salgia, "Integration of artificial intelligence in lung cancer: Rise of the machine," Cell Reports Medicine, vol. 4, no. 2, 2023.
- [13] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 770–778, 2016.
- [14] S. Bakr, O. Gevaert, S. Echegaray, K. Ayers, M. Zhou, M. Shafiq, H. Zheng, J. A. Benson, W. Zhang, A. N. Leung, et al., "A radiogenomic dataset of non-small cell lung cancer," Scientific data, vol. 5, no. 1, pp. 1–9, 2018.
- [15] X. Yu and S.-H. Wang, "Abnormality diagnosis in mammograms by transfer learning based on resnet18," Fundamenta Informaticae, vol. 168, no. 2-4, pp. 219–230, 2019.
- [16] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-cam: Visual explanations from deep networks via gradient-based localization," in Proceedings of the IEEE international conference on computer vision, pp. 618–626, 2017.

# Разработка метода выявления дипфейков: применение частотного анализа и уменьшения цветового пространства изображения для повышения точности классификации

В. Роговой, В. М., Коржук, О. А. Кокорина

Национальный исследовательский университет ИТМО

v\_rogovoi@itmo.ru, vmkorzhuk@itmo.ru

**Аннотация.** С развитием генеративных и диффузионных моделей нейросетей появилась возможность создавать высококачественные реалистичные изображения, визуально неотличимые от реальных фотографий. Это расширяет горизонты использования технологии для синтеза данных и медиа контента, однако с помощью данной технологии появляется возможность генерации или подмены лиц реальных людей, дипфейков. Несмотря на то, что современные методы распознавания дипфейков показывают высокие результаты на синтетических данных, они страдают от проблемы ложных срабатываний при работе с обработанными изображениями реальных и синтетических лиц. Более того, изменение размера изображения в соответствии с входным слоем классификаторов, основанных на сверточных нейронных сетях, может привести к дополнительным искажениям, что может повлиять на процесс классификации. В данной работе предлагается метод, основанный на редукции цветовой палитры и спектральном анализе изображений для повышения точности классификации дипфейков. Предлагаемый метод позволил повысить точность распознавания модифицированных дипфейков до 99.4 %.

**Ключевые слова:** цифровая обработка изображений; глубокое обучение; дипфейк; подделка личности

## I. ВВЕДЕНИЕ

Дипфейк – это метод синтеза контента, основанный на использовании генеративных моделей, например, генеративно-состязательных нейросетей (GAN) [1] и методов устойчивой диффузии (Stable diffusion) [2], с целью подмены лица человека в медиаконтенте.

В последние годы данная технология стала неотъемлемой частью цифровой среды, вызывая широкий общественный интерес и растущую озабоченность из-за своего потенциального влияния на общественное мнение и безопасность данных. Отмечается, что первые упоминания о дипфейках появились к концу 2017 года, и с тех пор термин активно используется в различных контекстах. Пики активности поисковых запросов, связанные с дипфейками, соотносятся с определенными событиями, такими как запрет видео на сайте «PornHub» и обнаружение реалистичных дипфейков с участием известных личностей, например, Тома Круза.

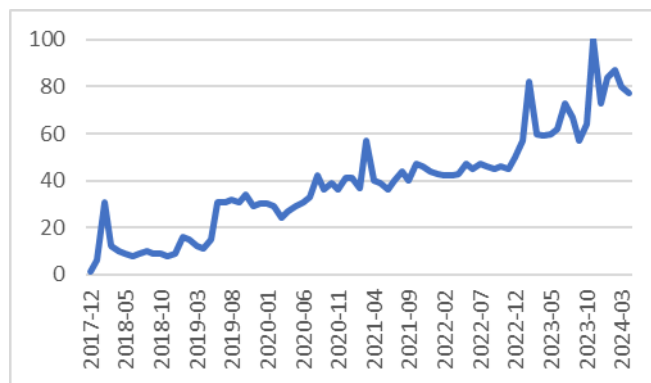


Рис. 1. Активность по поисковому запросу «deepfake» в Google Trends

Однако рост популярности дипфейков также сопровождается увеличением случаев цифрового мошенничества. В 2023, по данным Onfido, году количество подобных случаев выросло в 30 раз [3], а в феврале 2024 года дипфейки были использованы для кражи значительной суммы денег у крупной компании [4].

Это подчеркивает необходимость разработки и совершенствования методов выявления дипфейков с целью защиты от негативных последствий их использования.

## II. ОБЗОР ЛИТЕРАТУРЫ

Существует множество работ, в которых для классификации применяются специфические артефакты, получаемые в ходе создания дипфейков. Например, в работе «Detecting and simulating artifacts in GAN fake images» [5] представили метод, который воспроизводит артефакты присущие изображениям, созданные с помощью GAN и подает их на вход классификатора для идентификации дипфейков. В работе «Not made for each other: audio-visual dissonance-based deepfake detection and localization» [6] предложено схожее, но более обобщенное решение.

Кроме того, в работах [7, 8] исследователи предложили новую систему обнаружения, основанную на физиологических измерениях, например, сердцебиении.

Также велись и исследования зависимости точности от размера обучаемой модели. Так, например, в работе «FaceForensics++: Learning to detect manipulated facial images» [9] показывают, что CNN с большим количеством гиперпараметров [10] превосходят малоразмерные CNN. В некоторых методах применяются техники извлечения из видеок кадров текстурных признаков [11], пространственно-временных признаков [12, 13], текстур [14, 15], используя при этом 68 ориентиров на лице [16, 17] с визуальными артефактами (например, движение глаз, зубов, губ и т. д.). Эти признаки использовались в качестве входных данных для обучения классификаторов.

Дальнейшие успехи были достигнуты за счет внедрения механизма внимания, как в работе «On the detection of digital face manipulation» [18], а в [19, 20] были получены хорошие результаты за счет использования архитектуры, названной капсульной сетью (CN). Для обучения CN требуется меньшее количество параметров, чем для глубоких сетей. Для повышения производительности таких структур применяется техника ансамблевого обучения [21, 22], которая позволяет достичь точности более 99 %.

В большинстве исследований применяется пок кадровый анализ видео или изображений с целью анализа лица и отслеживания его движения для достижения лучшей производительности. Например, в работах «Deepfake video detection using recurrent neural networks» и «Recurrent convolutional structures for audio spoof and video deepfake detection» [23–27] предлагаются сети на основе RNN для извлечения признаков на различных микро- и макроскопических уровнях для обнаружения дипфейков.

Несмотря на столь впечатляющие результаты в обнаружении, большинство методов склоняются к чрезмерной подгонке. Для решения этих проблем применяются архитектуры на основе автоэнкодеров [28–31]. На различные модели накладывается пиксельная маска [32], чтобы получить область изображения лица. В [33] применили подходы к обучению на основе состязательности, а затем механизмы, основанные на внимании как в [34]. В работе [35] исследователи предложили метод кластеризации, включив в функцию потерь при классификации, основанную на маргинальном триплетном вкраплении. В [36, 37] предложили технику предварительной обработки данных для обнаружения глубоких подделок с помощью методов свертки изображений.

Научная задача данной работы – разработать научно-методический аппарат по обнаружению дипфейков.

### III. ОСНОВНАЯ ЧАСТЬ

В данном разделе приводится обоснование компонент и архитектуры, которые будут использоваться для классификации.

#### A. Выбор признаков для анализа

Согласно исследованию [38] для классификации обычно используются следующие компоненты:

- визуальные артефакты;
- частотные компоненты;

- шумовые артефакты;
- артефакты GAN
- мезоскопия;
- биологические артефакты.

И др.

Однако большая часть методов, исследуемых в предложенной работе, в случаях, когда изображение или видео прошли постобработку, сильно теряют в точности. И это является большой проблемой, поскольку в реальном мире дипфейки плохого качества можно определить на глаз. По этой причине было принято решение остановиться на выделении шумовых и спектральных компонент изображения.

Выбор спектральных компонент обусловлен работами [39, 40], в которых рассматривался факт того, что генеративные модели, из-за особенности их работы, оставляют дискретные шумы в спектре изображения.

Их использование ограничено использованием предобработки для фотографий, например, наложением различных фильтров или обработкой в фоторедакторе. В данном случае обработка может влиять на дискретные шумы в спектре, и модель будет чаще ошибаться. На рис. 2 представлена матрица классов обработанных, реальных и дипфейк изображений.

Реальное изображение	Реальное обработанное
Дипфейк	Обработанный дипфейк

Рис. 2. Матрица классов для задачи распознавания дипфейков

На основании рис. 2 мы расширяем задачу классификации real/fake до задачи классификации real, real processed и fake, fake processed, однако в данном исследовании будет проводиться только классификация real/fake.

По этой причине было принято решение рассмотрения методов детекции обработки изображений. [41]

Каналов RGB и спектрального представления недостаточно, чтобы справиться со всеми различными случаями манипулирования. В частности, искаженные изображения, которые были тщательно обработаны для сокрытия границ срачивания и уменьшения различий в контрастности, создают проблемы для потока RGB.

В этом случае используется локальное распределение шума на изображении для получения дополнительных данных. В отличие от RGB-потока, шумовые компоненты предназначены для того, чтобы уделять больше внимания шуму, а не семантическому содержанию изображения.

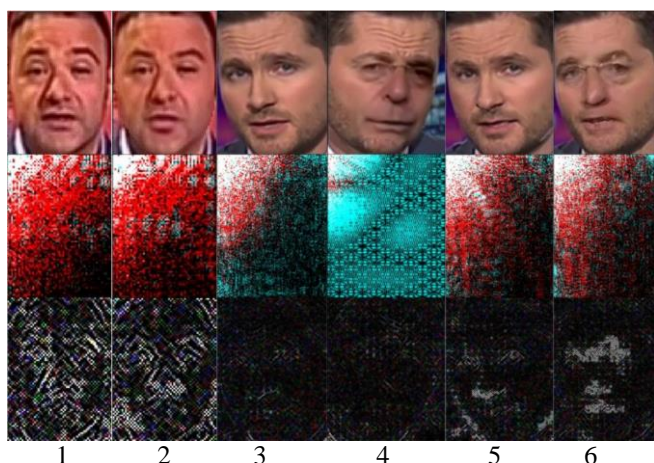


Рис. 3. Спектральные и шумовые компоненты для оригинального изображения лица человека (1) и различных методов подмены (2–6)

На рис. 3 представлен пример спектральных и шумовых компонент изображения различных методов генерации изображений, взятых из набора данных DFDC++ [42].

Для извлечения шумовых компонент изображения рассматривались следующие фильтры:

$$\frac{1}{4} \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 2 & -1 & 0 \\ 0 & 2 & -4 & 2 & 0 \\ 0 & -1 & 2 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \quad \frac{1}{12} \begin{bmatrix} -1 & 2 & -2 & 2 & -1 \\ 2 & -6 & 8 & -6 & 2 \\ -2 & 8 & -12 & 8 & -2 \\ 2 & -6 & 8 & -6 & 2 \\ -1 & 2 & -2 & 2 & -1 \end{bmatrix}$$

Рис. 4. Ядра SRM фильтров, которые были использованы для предобработки изображений

### B. Описание используемой архитектуры

В качестве классификатора была использована ResNet [43] подобная архитектура сети. Ее основная идея заключается в использовании «скачков» (residual connections), которые позволяют эффективно обучать глубокие нейронные сети без проблемы затухания градиента.

В данном исследовании мы использовали архитектуру ResNet [3, 3, 2, 1] с большим количеством блоков в начале нейронной сети, для анализа более простых признаков, таких как границы и текстуры, выделенных нами компонент. При этом использование не слишком большого количества блоков в каждом этапе помогает ускорить обучение и улучшить обобщающую способность модели. Слишком глубокие сети могут страдать от переобучения и требуют больше вычислительных ресурсов для обучения.

Это особенно важно, поскольку мы выделяем 2 типа признаков для обучения, и, соответственно, увеличиваем сложность нейросети. Итоговая архитектура представлена на рис. 5.

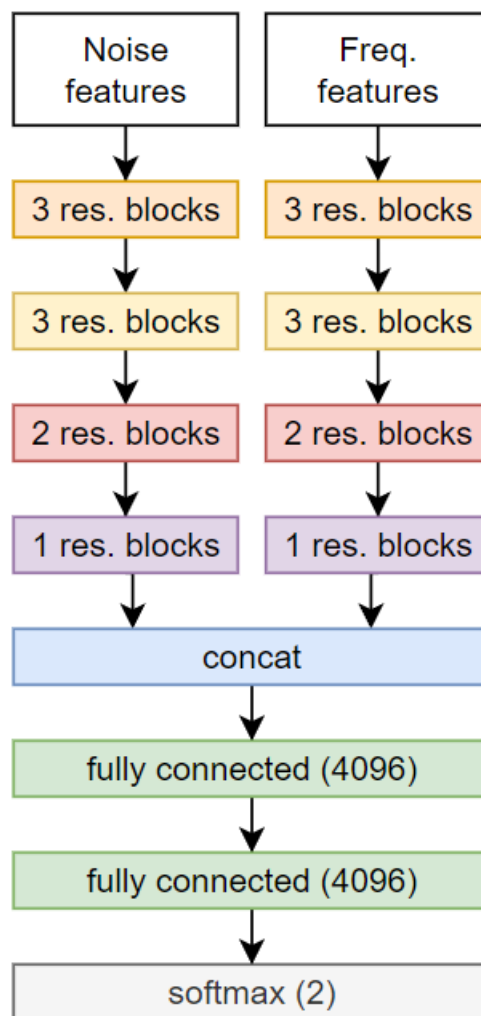


Рис. 5. Предлагаемая архитектура классификатора на основе ResNet

### C. Описание набора данных

В данном исследовании мы использовали 2 набора данных для проведения экспериментов: FF++ [44] и CELEB-DF [45], а также синтезировали с помощью Stable Diffusion 5000 изображений лиц людей. Для сегментации боксов лиц из видеозаписей была выбрана нейросеть YOLOv8 [46]. В табл. 1 представлено распределение количества лиц для тренировочного и тестового набора данных.

ТАБЛИЦА I. ОПИСАНИЕ НАБОРА ДАННЫХ

	Real	Forged
Train	100 000	100 000
Test	50 000	50 000

Первые тесты проводились на наборе данных DeeperForensics [47], однако точность классификации на данном наборе уже на первой эпохе была равна 1, в то время как на тестах на этом же наборе данных точность составляла 72 %, а на других наборах данных ~52 %. Это было по той причине, что распределение размеров лиц для реальных и поддельных фото очень сильно отличались, ввиду чего происходило переобучение.



Ввиду этого было принято решение составлять набор данных таким образом, чтобы распределение размеров лиц было приближено к равномерному.

#### D. Проведение экспериментов

Для оценки качества обученной модели были использованы метрики точность (1), где TP – количество истинно положительных примеров, а FP – количество ложноположительных примеров и квадратичная функция потерь (2) модели, где C – константа, y – истинное значение выхода модели (которое должно быть получено в идеальном случае), y' – фактический выход модели.

$$Accuracy = (TP + TN) / (TP + FP + TN + FN) \quad (1)$$

$$\lambda(y) = C(y - y') \quad (2)$$

Процесс обучения представлен на рис. 6.

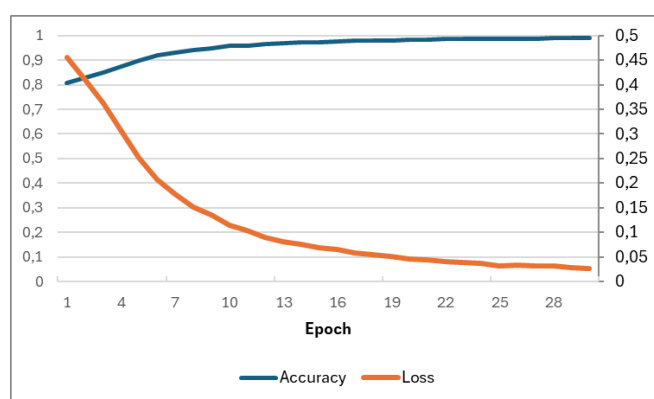


Рис. 6. Точность и функция потерь модели в процессе обучения

На тестах модель показала точность 99.01 %. Сравнение с существующими моделями представлено в табл. 2 и 3. Основной акцент был сделан на сравнении предлагаемого метода с существующими в условиях, когда изображение было размыто или зашумлено.

ТАБЛИЦА II. СРАВНЕНИЕ ПРЕДЛАГАЕМОГО МЕТОДА С СУЩЕСТВУЮЩИМИ РЕШЕНИЯМИ

Метод	Точность, %
Rahmouni	97.03
MesoNet	98.6
XceptionNet	<b>99.26</b>
CNN-DCT	99.07
Предлагаемый метод	99.01

ТАБЛИЦА III. СРАВНЕНИЕ ТОЧНОСТИ ПРЕДЛАГАЕМОГО МЕТОДА С СУЩЕСТВУЮЩИМИ РЕШЕНИЯМИ ПРИ УСЛОВИИ ОБРАБОТКИ ИЗОБРАЖЕНИЯ

	Размытие	Шум	Комбинация
CNN-DCT	93.61%	89.56%	92.17%
Предлагаемый метод	<b>97.13%</b>	<b>95.31%</b>	<b>96.57%</b>

#### IV. ЗАКЛЮЧЕНИЕ

За последние годы проблема подделки личности с помощью дипфейков становится все более актуальной. В связи с этим возрастает потребность в надежных методах и инструментах для их обнаружения. В данной работе мы рассмотрели использование комбинированного подхода к извлечению группы признаков изображения, а

также ResNet подобной архитектуры для повышения точности классификации.

Предлагаемый метод не уступает существующим в точности, а также превосходит лучший из них в условиях постобработки изображений.

Однако при проведении тестов на реальных роликах, взятых из интернета, где подмена лиц проводилась более качественно, чем в представленных наборах данных, точность классификации составила всего 78 %. По этой причине дальнейшая работа будет заключаться как в расширении выборки, так и в исследовании других комбинаций признаков, на основе которых будет производиться классификация.

#### СПИСОК ЛИТЕРАТУРЫ

- [1] Swathi P., Sk S. Deepfake creation and detection: A survey //2021 Third International Conference on Inventive Research in Computing Applications (ICIRCA). – IEEE, 2021. – С. 584-588.
- [2] Borji A. Generated faces in the wild: Quantitative comparison of stable diffusion, midjourney and dall-e 2 //arXiv preprint arXiv:2210.00586. – 2022.
- [3] Onfido [Электронный ресурс] / URL: <https://onfido.com/use-cases/customer-acquisition/> (дата обращения: 15.04.24)
- [4] Хакер [Электронный ресурс] / URL: <https://хакер.ru/2024/02/06/deepfake-scammers/> (дата обращения: 01.04.24)
- [5] 75. R. Durall, M. Keuper, F.-J. Pfrendt, and J. Keuper, “Unmasking DeepFakes with simple features,” 2019, arXiv:1911.00686.
- [6] K. Ghugh, P. Gupta, A. Dhall, and R. Subramanian, “Not made for each other- audio-visual dissonance-based deepfake detection and localization,” 2020, arXiv:2005.14405.
- [7] K. Songsri-in and S. Zafeiriou, “Complement face forensic detection and localization with facial landmarks,” 2019, arXiv:1910.05455.
- [8] A. Kumar and A. Bhavsar, “Detecting deepfakes with metric learning,” 2020, arXiv:2003.08645.
- [9] A. Rossler, D. Cozzolino, L. Verdoliva, C. Riess, J. Thies, and M. Niessner, “FaceForensics++: Learning to detect manipulated facial images,” in Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV), Oct. 2019, pp. 1–11
- [10] A. Khodabakhsh and C. Busch, “A generalizable deepfake detector based on neural conditional distribution modelling,” in Proc. Int. Conf. Biometrics Special Interest Group (BIOSIG), Darmstadt, Germany, Sep. 2020, pp. 1–5.
- [11] Karasavva V., Noorbhai A. The real threat of deepfake pornography: A review of Canadian policy //Cyberpsychology, Behavior, and Social Networking. – 2021. – Т. 24. – №. 3. – С. 203-209
- [12] A. Khodabakhsh and C. Busch, “A generalizable deepfake detector based on neural conditional distribution modelling,” in Proc. Int. Conf. Biometrics Special Interest Group (BIOSIG), Darmstadt, Germany, Sep. 2020, pp. 1–5.
- [13] Y. Li, M.-C. Chang, and S. Lyu, “In ictu oculi: Exposing AI created fake videos by detecting eye blinking,” in Proc. IEEE Int. Workshop Inf. Forensics Secur. (WIFS), Dec. 2018, pp. 1–7.
- [14] Y. Li and S. Lyu, “Exposing deepfake videos by detecting face warping artifacts,” in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR) Workshops, 2019, pp. 46–52
- [15] I. Ganiyusufoglu, L. Minh Ngô, N. Savov, S. Karaoglu, and T. Gevers, “Spatio-temporal features for generalized detection of deepfake videos,” 2020, arXiv:2010.11844.
- [16] I. Kukanov, J. Karttunen, H. Sillanpää, and V. Hautamäki, “Cost sensitive optimization of deepfake detector,” 2020, arXiv:2012.04199.
- [17] A. Haliassos, K. Vougioukas, S. Petridis, and M. Pantic, “Lips don’t lie: A generalisable and robust approach to face forgery detection,” 2020, arXiv:2012.07657. 25510 VOLUME 10, 2022 M. S. Rana et al.: Deepfake Detection: Systematic Literature Review
- [18] H. Dang, F. Liu, J. Stehouwer, X. Liu, and A. K. Jain, “On the detection of digital face manipulation,” in Proc. IEEE/CVF Conf.

- Comput. Vis. Pattern Recognit. (CVPR), Seattle, WA, USA, Jun. 2020, pp. 5780–5789, doi: 10.1109/CVPR42600.2020.00582.
- [19] H. H. Nguyen, J. Yamagishi, and I. Echizen, “Capsule-forensics: Using capsule networks to detect forged images and videos,” in Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP), Brighton, U.K., May 2019, pp. 2307–2311, doi: 10.1109/ICASSP.2019.8682602.
- [20] H. H. Nguyen, J. Yamagishi, and I. Echizen, “Use of a capsule network to detect fake images and videos,” 2019, arXiv:1910.12467.
- [21] N. Bonettini, E. Daniele Cannas, S. Mandelli, L. Bondi, P. Bestagini, and S. Tubaro, “Video face manipulation detection through ensemble of CNNs,” 2020, arXiv:2004.07676.
- [22] M. S. Rana and A. H. Sung, “DeepfakeStack: A deep ensemble-based learning technique for deepfake detection,” in Proc. 7th IEEE Int. Conf. Cyber Secur. Cloud Comput. (CSCloud)/6th IEEE Int. Conf. Edge Comput. Scalable Cloud (EdgeCom), New York, NY, USA, Aug. 2020, pp. 70–75, doi: 10.1109/CSCloud-EdgeCom49738.2020.00021.
- [23] X. Ding, Z. Raziei, E. C. Larson, E. V. Olinick, P. Krueger, and M. Hahsler, “Swapped face detection using deep learning and subjective assessment,” EURASIP J. Inf. Secur., vol. 2020, no. 1, pp. 1–12, Dec. 2020, doi: 10.1186/s13635-020-00109-8.
- [24] Z. Guo, G. Yang, J. Chen, and X. Sun, “Fake face detection via adaptive manipulation traces extraction network,” 2020, arXiv:2005.04945.
- [25] D. Mas Montserrat, H. Hao, S. K. Yarlagadda, S. Baireddy, R. Shao, J. Horváth, E. Bartusiak, J. Yang, D. Güera, F. Zhu, and E. J. Delp, “Deepfakes detection with automatic face weighting,” 2020, arXiv:2004.12027.
- [26] L. M. Dang, S. I. Hassan, S. Im, and H. Moon, “Face image manipulation detection based on a convolutional neural network,” Expert Syst. Appl., vol. 129, pp. 156–168, Sep. 2019.
- [27] Z. Liu, X. Qi, J. Jia, and P. H. S. Torr, “Real or fake: An empirical study and improved model for fake face detection,” in Proc. 8th Int. Conf. Learn. Represent. (ICLR), Apr. 2020, pp. 1–12.
- [28] X. Ding, Z. Raziei, E. C. Larson, E. V. Olinick, P. Krueger, and M. Hahsler, “Swapped face detection using deep learning and subjective assessment,” EURASIP J. Inf. Secur., vol. 2020, no. 1, pp. 1–12, Dec. 2020, doi: 10.1186/s13635-020-00109-8.
- [29] Z. Guo, G. Yang, J. Chen, and X. Sun, “Fake face detection via adaptive manipulation traces extraction network,” 2020, arXiv:2005.04945.
- [30] D. Mas Montserrat, H. Hao, S. K. Yarlagadda, S. Baireddy, R. Shao, J. Horváth, E. Bartusiak, J. Yang, D. Güera, F. Zhu, and E. J. Delp, “Deepfakes detection with automatic face weighting,” 2020, arXiv:2004.12027.
- [31] L. M. Dang, S. I. Hassan, S. Im, and H. Moon, “Face image manipulation detection based on a convolutional neural network,” Expert Syst. Appl., vol. 129, pp. 156–168, Sep. 2019.
- [32] Z. Liu, X. Qi, J. Jia, and P. H. S. Torr, “Real or fake: An empirical study and improved model for fake face detection,” in Proc. 8th Int. Conf. Learn. Represent. (ICLR), Apr. 2020, pp. 1–12.
- [33] I. Amerini, L. Galteri, R. Caldelli, and A. Del Bimbo, “Deepfake video detection through optical flow based CNN,” in Proc. IEEE/CVF Int. Conf. Comput. Vis. Workshop (ICCVW), Oct. 2019, pp. 1205–1207.
- [34] D. Cozzolino, J. Thies, A. Rössler, C. Riess, M. Nießner, and L. Verdoliva, “Forensic Transfer: Weakly-supervised domain adaptation for forgery detection,” 2018, arXiv:1812.02510.
- [35] M. A. S. Habeeba, A. Lijiya, and A. M. Chacko, “Detection of deepfakes using visual artifacts and neural network classifier,” in Innovations in Electrical and Electronic Engineering (Lecture Notes in Electrical Engineering), vol. 661, M. Favorskaya, S. Mekhilef, R. Pandey, and N. Singh, Eds. Singapore: Springer, 2020, pp. 411–422, doi: 10.1007/978-981-15-4692-1\_31.
- [36] P. Korshunov and S. Marcel, “DeepFakes: A new threat to face recognition? Assessment and detection,” 2018, arXiv:1812.08685.
- [37] A. Gandhi and S. Jain, “Adversarial perturbations fool deepfake detectors,” in Proc. Int. Joint Conf. Neural Netw. (IJCNN), Jul. 2020, pp. 1–8.
- [38] Rana M. S. et al. Deepfake detection: A systematic literature review //IEEE access. 2022. T. 10. C. 25494-25513.
- [39] Frank J. et al. Leveraging frequency analysis for deep fake image recognition //International conference on machine learning. PMLR, 2020. C. 3247-3258.
- [40] Corvi R. et al. Intriguing properties of synthetic images: from generative adversarial networks to diffusion models //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2023. C. 973-982.
- [41] Zhou P. et al. Learning rich features for image manipulation detection //Proceedings of the IEEE conference on computer vision and pattern recognition. 2018. C. 1053-1061.
- [42] benpflaum, Brian G, djdj, Irina Kofman, JE Tester, JLElliott, Joshua Metherd, Julia Elliott, Mozaic, Phil Culliton, Sohler Dane, Woo Kim. (2019). Deepfake Detection Challenge. Kaggle. <https://kaggle.com/competitions/deepfake-detection-challenge>
- [43] Koonce B., Koonce B. ResNet 50 //Convolutional neural networks with swift for tensorflow: image recognition and dataset categorization. 2021. C. 63-72.
- [44] Andreas Rossler and Davide Cozzolino and Luisa Verdoliva and Christian Riess and Justus Thies and Matthias Niener. FaceForensics++: Learning to Detect Manipulated Facial Images., International Conference on Computer Vision (ICCV), 2019.
- [45] Yuezun Li, Xin Yang, Pu Sun, Honggang Qi and Siwei Lyu., CelebDF: A Large-scale Challenging Dataset for DeepFake Forensics, IEEE Conference on Computer Vision and Patten Recognition (CVPR), 2020
- [46] YOLOv8 [электронный ресурс] / URL: <https://blog.roboflow.com/whats-new-in-yolov8/#what-is-yolov8> (дата обращения: 13.02.24)
- [47] Jiang L. et al. Deeperforensics-1.0: A large-scale dataset for real-world face forgery detection //Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020. C. 2889-2898.

# Распознавание и локализация объектов посредством компьютерного зрения

В. В. Ковалев, Е. В. Илатовская, Д. М. Филатов,  
А. О. Федоркова, Д. П. Михайлов

Санкт-Петербургский государственный электротехнический университет  
«ЛЭТИ» им. В.И. Ульянова (Ленина)

vladimir.kovalev.2012@bk.ru, evilatovskaia@etu.ru, dmfilatov@etu.ru,  
anastasiikanata@gmail.com, dpmikhailov@etu.ru

**Аннотация.** Доклад рассматривает вопросы распознавания и локализации объекта наблюдателем посредством компьютерного зрения, в условиях прямой видимости самого объекта. Доклад призван расширить понимание проблем управления автономными объектами в условиях прямой видимости.

**Ключевые слова:** обнаружение объектов, компьютерное зрение, отслеживание объектов

## I. ВВЕДЕНИЕ

С момента, когда человечество освоило полёты на воздушных шарах, появилась идея создания беспилотных летательных аппаратов, в первую очередь для аэрофотосъемки. В наши дни же, благодаря прорывам в области микро- и нанoeлектроники, а также современных компонентов, стало возможным создавать автономные устройства, способные передвигаться как в воздухе, так и под водой. Такие объекты могут действовать автономно как в помещениях, так и на открытом пространстве, однако ключевой задачей остаётся их точное позиционирование.

## II. МЕТОДЫ ЛОКАЛИЗАЦИИ АВТОНОМНЫХ ОБЪЕКТОВ

Для управления автономным объектом в условиях прямой видимости достаточно стандартного протокола передачи данных между пультом управления и оператором. Однако основная цель создания беспилотного объекта – его способность к автоматическому перемещению. Для этого объект должен точно определять своё местоположение, отслеживать перемещения во времени и корректно управлять используемыми актуаторами.

Сегодня системы локализации автономного объекта можно поделить на две большие группы: основанные на GNSS (англ. Global Navigation Satellite System – глобальная навигационная спутниковая система) и основанные на окружающей среде. Системы GNSS считаются более традиционными. С другой стороны, системы, основанные на окружающей среде, более современные и точные. [1]

### A. Глобальная навигация

Глобальная навигация подразумевает под собой использование GNSS модуля, который получает данные со спутников, расположенных на околоземной орбите. По полученным данным можно вычислить координаты объекта и скорость перемещения. В идеальных условиях

данная технология обеспечивает точность в пределах 2–5 метров. Однако в большинстве реальных ситуаций точность снижается до 10 метров или более. Это происходит из-за медленного обновления данных, наличия помех и препятствий, таких как здания, туннели и другие металлические конструкции, которые мешают прямой передаче сигналов. [1]

Использование GNSS навигации сопряжено с рядом проблем, особенно когда число доступных спутников сокращается до четырех или менее. В таких условиях теряется необходимый минимум для точного позиционирования, что приводит к потере точности в определении перемещения объекта. В некоторых случаях грубое знание трехмерной среды может быть использовано для прогнозирования и предотвращения аварийно-опасных ситуаций. Тем не менее, в сложных условиях полета избежать использование GNSS бывает невозможно, и уменьшение количества работающих спутников может серьезно снизить точность или даже сделать невозможным использование данных о местоположении. Кроме того, проблемы усугубляются при использовании инерциальных датчиков MEMS потребительского класса на малых беспилотных летательных аппаратах. Эти датчики имеют относительно низкую производительность, что приводит к быстрому увеличению ошибок в отсутствие спутниковых данных о местоположении. [2]

### B. Локализация на основе данных GPS RTK

Под локализацией на основе GPS RTK необходимо понимать глобальную навигацию, но с использованием дополнительного маяка – базовой станции. Саму базовую станцию устанавливают с высокой точностью, что позволяет спутниковой навигации получать более точные данные о местоположении быстрее, чем при использовании только спутников. Точность в зоне действия базовой станции может достигать до 1 сантиметра. [3]

Основной принцип, лежащий в основе RTK GPS, прост: система использует локальную базовую станцию вместе со спутниковыми сигналами для коррекции возможных ошибок в данных GPS. Коррекция происходит за счет сравнения данных, полученных базовой станцией, с данными, полученными мобильным устройством. Любые ошибки, присутствующие в данных, затем исправляются, что приводит к получению высокоточных данных о местоположении.

Основными компонентами системы RTK GPS являются GPS-приемник, локальная базовая станция и канал передачи данных между ними. GPS-приемник обычно устанавливается на транспортном средстве или другой мобильной платформе и принимает данные со спутников GPS. Базовая станция обычно расположена на устойчивой поверхности, такой как здание или башня. Она получает те же GPS данные, что и мобильное устройство. Кроме того, базовая станция получает данные от близлежащей опорной станции, использующейся для коррекции возможных ошибок в данных GPS. Такая схема улучшает точность и надежность системы позиционирования.

Одной из наиболее важных особенностей RTK GPS является его способность предоставлять высокоточные данные о местоположении. Точность достигается за счет использования местной базовой станции в дополнение к спутниковым сигналам, что позволяет системе исправлять любые ошибки, которые могут присутствовать в данных GPS. Такой подход позволяет получать данные о местоположении с точностью до сантиметров, что делает его идеальным для широкого спектра применений.

Другой ключевой особенностью RTK GPS является его способность предоставлять данные в режиме реального времени. Поскольку система использует локальную базовую станцию в дополнение к спутниковым сигналам, она может предоставлять данные в режиме реального времени с точностью до миллисекунд. Данная особенность делает его идеальным для приложений, требующих навигации в режиме реального времени, таких как автономные транспортные средства и беспилотные летательные аппараты.

В дополнение к высокой точности и возможностям работы в режиме реального времени, RTK GPS также предлагает ряд других функций, которые делают его идеальным для широкого спектра применений. К ним относятся:

- длительное время автономной работы: система использует местную базовую станцию в дополнение к спутниковым сигналам, она может работать в течение продолжительного времени без подзарядки.
- надежность: системы RTK GPS спроектированы таким образом, чтобы быть надежными даже в сложных условиях, что делает их идеальными для использования на пересеченной местности и в других сложных условиях.
- низкая стоимость: по сравнению с другими типами высокоточных GPS-систем, RTK GPS стоит относительно недорого, оставаясь доступным для широкого круга пользователей, от индивидуальных геодезистов до крупных строительных компаний.

### C. Локализация на основе окружающей среды

Локализация на основе окружающей среды включает использование различных датчиков, установленных непосредственно на управляемом объекте, таких как датчики дальности, камеры, лидары, а также инерционные измерительные устройства, включая гироскопы и акселерометры. Примером такой технологии является SLAM (Simultaneous Localization and Mapping – Одновременная локализация и картографирование).

### D. Технология одновременной локализации и построения карты SLAM

SLAM – это метод навигации, применяемый для определения местоположения и ориентации автономных роботов в условиях заранее неизвестной территории, а также, для обновления или дополнения уже известных карт окружающего пространства. [4]

SLAM представляет собой метод локализации на основе данных окружающей среды, при котором автономный объект должен не только запоминать своё местоположение, но и собирать, обновлять данные с датчиков, а также строить карту окружения на основе собранных данных. В основе этой технологии лежат методы одометрии, позволяющие определить перемещение объекта на основе анализа предыдущего движения колёс, данных с камер или инерциального измерительного юнита (IMU), а также комбинируя все эти данные. SLAM позволяет не только строить и ориентироваться по карте, но и передавать её, а также визуализировать, создавая трёхмерную модель окружения, как показано на рисунке ниже:

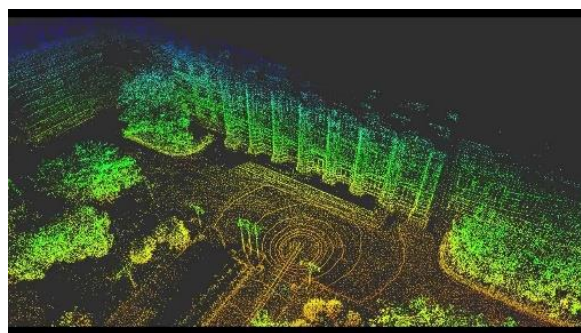


Рис. 1. Облака точек, полученных с технологии SLAM в работе

В вышеуказанных примерах не рассматриваются ситуации, где автономный объект может перемещаться в пространстве и управляться удаленно наблюдателем. Такие системы обычно требуют только наличия камеры и самого автономного объекта. Рассмотрим подробнее, как функционирует распознавание в таком автономном объекте.

### III. РАСПОЗНАВАНИЕ АВТОНОМНОГО ОБЪЕКТА

С помощью библиотеки компьютерного зрения openCV и языка программирования Python рассмотрим ситуацию, представленную на рисунке ниже, когда в кадре присутствует объект управления:



Рис. 2. Внешний вид автономного объекта

Из рис. 2 видно, что автономный объект оснащен двумя красными и двумя синими светодиодами. Для распознавания такого объекта на камере можно использовать метод поиска красных точек. Нахождение центра между двумя красными светодиодами позволит определить центральную точку самого объекта.

Ниже представлен результат программной реализации фильтра поиска такого объекта на изображении:

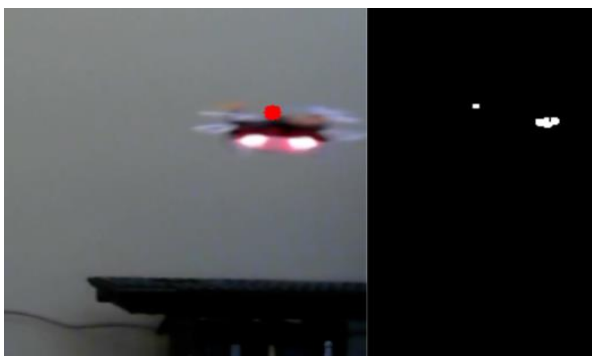


Рис. 3. Результат поиска автономного объекта на изображении с помощью маски цвета, слева – представлен сам объект и найденный центр объекта, справа – маска изображения для поиска объекта по цвету

Из рис. 4 отметим, что при поиске самого объекта сначала применяется поиск подходящего цвета объекта. Затем, когда выявлены два характерных пятна этого цвета, можно сделать вывод, что центр объекта располагается примерно между этими двумя пятнами.

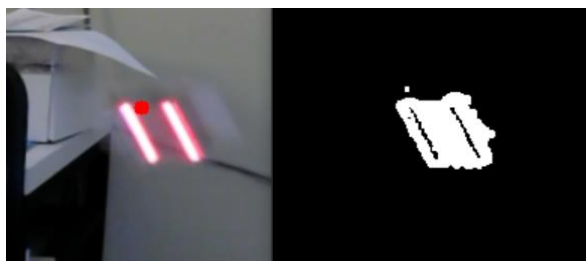


Рис. 4. Результат поиска автономного объекта на изображении с помощью маски цвета, слева – объект и найденный центр объекта, справа – маска изображения для поиска объекта по цвету

Невооружённым глазом видно, что центр объекта, который отмечен красным на рис. 3 и 4, не совпадает с реальным центром объекта. Кроме того, такая простая в реализации и прямолинейная в исполнении программа может давать сбои. Результат работы программы зависит от условий освещения, и она не всегда способна обеспечить достаточно точные результаты, как в определении расстояния между светодиодами, так и в определении расстояния от объекта до камеры.

Выходом из данной ситуации может быть обучение нейронной сети на распознавание и слежение за одним классом – автономным объектом. Рассмотрим этот подход далее.

Возьмем нейронную сеть YOLOv5 модель S, и проведем обучение на выборке в 1 минуту нахождения объекта в кадре. На рисунках ниже показаны графики

точности определения класса при обучении и результат работы программы нейронной сети по распознаванию объекта:

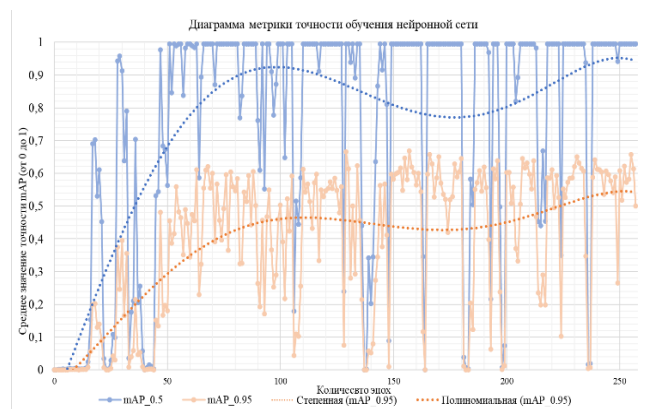


Рис. 5. График обучения нейронной сети YOLOv5 mS, который демонстрирует точность обучения при выборке в 270 эпох и единственном классе объекта

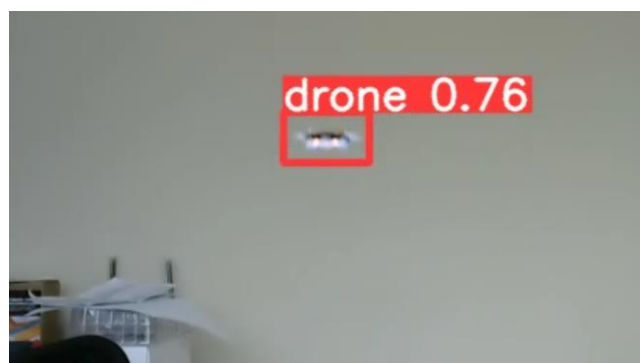


Рис. 6. Распознанный объект управления, точность распознавания 76 %

Из рис. 6 видно, что выборка для обучения сети слабая, а количество эпох составляет всего чуть больше 250, тем не менее, на рис. 7 объект, который требуется находить на изображении, обведён красным и находится ровно в центре рамки, с вероятностью 76 % это искомый объект.

Если объект управления находится в кадре, вероятность его обнаружения составляет около 60 %, когда он расположен в центре кадра. Если объект частично перекрывается препятствиями, вероятность обнаружения снижается до 40 %.

#### IV. ЛОКАЛИЗАЦИЯ АВТОНОМНОГО ОБЪЕКТА

Реализуя корректное распознавание объекта на изображении, можно попытаться рассчитать его местоположение по простой формуле:

$$S = \frac{FH}{h}, \quad (1)$$

где  $S$  – это расстояние до объекта наблюдения,  $F$  – фокусное расстояние объектива,  $H$  – горизонтальный размер объекта,  $h$  – размер матрицы по горизонту.

Так, при наличии параметров исходных размеров объекта можно узнать, насколько далеко от камеры он находится и существует ли смещение влево или вправо относительно центра камеры.

До этого момента подразумевалось, что при работе алгоритма распознавания объекта задействована только одна камера. Однако использование нескольких камер может значительно улучшить результаты, так как это позволяет наблюдать объект с разных ракурсов, увеличивая точность и надежность распознавания.

Усовершенствовать алгоритм можно, установив две камеры так, чтобы их углы обзора были перпендикулярны друг другу и их поля зрения перекрывались. Используя такую конфигурацию, можно применять формулу (1) для определения расстояния до объекта отдельно для каждой камеры. Комбинируя полученные данные, можно точно определить трехмерное положение объекта и отслеживать его перемещения в пространстве. Данный подход значительно повышает точность и эффективность системы мониторинга.

Отдельного внимания заслуживает технология стерео-зрения. Она также реализуется с использованием двух камер, что позволяет создавать карту глубины изображения. Это, в свою очередь, облегчает определение расстояния до объектов, используя визуальную информацию с одного ракурса и зная расстояние между камерами. Как правило, чем больше расстояние между камерами, тем точнее и детализированнее получается карта глубины и понимание трехмерного пространства. Однако с увеличением расстояния между камерами уменьшается область пересечения их полей зрения, что может усложнить получение согласованной карты.

## V. ЗАКЛЮЧЕНИЕ

Существует множество методов локализации автономного объекта, среди которых наиболее доступным является использование GPS-навигации на основе спутников. Однако для достижения высокой точности позиционирования часто применяют технологии GPS RTK. Тем не менее, в некоторых ситуациях глобальные системы локализации могут оказаться неэффективными, например, в закрытых или плохо доступных для спутниковых сигналов местах. В таких случаях целесообразно использовать системы SLAM или системы, основанные на наблюдении за объектом.

Системы позиционирования с наблюдателем обычно включают в себя камеры наблюдения. Однако использование таких систем сталкивается с определёнными трудностями. Основная проблема – что делать, если объект выходит из зоны видимости камеры.

## СПИСОК ЛИТЕРАТУРЫ

- [1] Определение местоположения в Беспилотных автомобилях (localization) [Электронный ресурс]. – Режим доступа: <https://habr.com/ru/articles/747634/> (дата обращения: 14.03.2024).
- [2] Causa, F., Vetrella, A. R., Fasano, G., & Accardo, D. (2018). Multi-UAV formation geometries for cooperative navigation in GNSS-challenging environments. 2018 IEEE/ION Position, Location and Navigation Symposium (PLANS).
- [3] RTK GPS: Understanding Real-Time Kinematic GPS Technology [Электронный ресурс]. – Режим доступа: <https://globalgpsystems.com/gnss/rtk-gps-understanding-real-time-kinematic-gps-technology/#:~:text=Real-time%20kinematic%20GPS%20is%20a,accuracy%20of%20the%20GPS%20data> (дата обращения: 14.03.2024).
- [4] Исследование методов SLAM для навигации мобильного робота внутри помещений. Опыт исследования R2 Robotics [Электронный ресурс]. – Режим доступа: <https://habr.com/ru/articles/560856/> (дата обращения: 14.03.2024).

# Влияние использования искусственного интеллекта на когнитивные способности школьников: субъективная оценка

М. А. Сорочинский<sup>1</sup>, С. Г. Прохорова<sup>2</sup>, К. А. Базанова<sup>1</sup>

<sup>1</sup> Северо-Восточный федеральный университет имени М.К. Аммосова

<sup>2</sup> ГАНОУ «Арктическая школа» Республики Саха (Якутия)

ma.sorochinskiy@s-vfu.ru

**Аннотация.** На сегодняшний день, до конца не изучено влияние искусственного интеллекта (ИИ) на сферы нашей жизнедеятельности. Крупные компании, организации, внедряют ИИ в свою работу для увеличения прибыли и сокращения издержек, ведутся работы по внедрению ИИ в медицину, образование, культуру и другие направления, что может повысить качество оказания медицинской помощи, увеличить эффективность образования и т.д. Однако, мало кто говорит об этических вопросах его использования, а также о влиянии ИИ на когнитивные функции человека. В статье приводятся результаты анализа положительных и отрицательных факторов влияния искусственного интеллекта на развитие человека, а также результаты опроса среди школьников г. Якутск (Республика Саха (Якутия), Россия) по самооценке влияния использования искусственного интеллекта на собственные когнитивные способности.

**Ключевые слова:** когнитивная деятельность; искусственный интеллект; когнитивные способности; школьники; самооценка

## I. ВВЕДЕНИЕ

Проблему снижения когнитивных функций человека поднимали еще в эпоху начала информационного века и развития цифровизации. Этой проблеме и сейчас уделяется особое внимание. Так, например, рассматривают то, как Интернет может изменить наше познание: «как уникальные особенности онлайн-мира могут влиять на внимание, процессы памяти, социальное познание. В целом, имеющиеся данные указывают на то, что Интернет может производить как острые, так и устойчивые изменения в каждой из этих областей познания, которые могут быть отражены в изменениях в мозге» [1]. Также проводятся лонгитюдные исследования, которые доказывают влияние частоты использования Интернета на развитие мозговых структур и вербального интеллекта [2].

Однако сейчас, проблема спада когнитивных функций переходит в следующий этап развития в связи с активным внедрением ИИ, т.к. основная задача ИИ – это попытка воспроизвести когнитивные функции человека, дополнить их, но очень важно посмотреть, не приведет ли это к обратному эффекту – деградации когнитивных функций. Одна из проблем заключается в том, что люди порой чрезмерно полагаются на ИИ, что приводит к снижению навыков решения проблем и анализа

ситуаций. Например, «если бездумно использовать программное обеспечение для автоматического перевода, то снизится способность к изучению иностранных языков и переводу текстов и речи без применения технических средств» [3].

Цель работы: изучить, какое влияние искусственный интеллект может оказывать на когнитивную деятельность человека и провести оценку влияния использования искусственного интеллекта на когнитивные функции школьников.

## II. ТРЕНДЫ РАЗВИТИЯ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

Основной тренд – массовое использование ИИ, внедрение в производство, совершенствование и ускорение процессов. Такой эффект будет возможен в том числе за счет снижения стоимости внедрения в деятельность организаций моделей ИИ, а также повышения доступности данных для обучения таких моделей.

Ключевой технологический тренд (рис. 1) – развитие и широкое распространение генеративного ИИ, в том числе больших языковых моделей. Генеративный ИИ используется для создания контента, включая текст, аудио, изображения, видео.

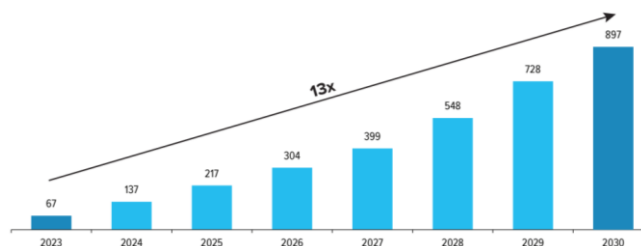


Рис. 1. Оценка объема рынка генеративного ИИ (млрд долл США)

По мнению экспертов, «в ближайшие годы генеративный ИИ станет одним из самых динамично растущих направлений ИИ. По оценке Bloomberg Intelligence, глобальный рынок решений на основе генеративного ИИ в 2023–2030 годах увеличится более чем в 13 раз – с 67 млрд до 897 млрд долларов США. Рост публичного интереса к сфере ИИ в последние годы во многом связан с развитием моделей генеративного ИИ. Благодаря стремительному росту производительности модели генеративного ИИ имеют

широкий потенциал применения в различных сферах в части оптимизации процесса взаимодействия с пользователями, улучшения клиентского опыта и других» [4].

### III. ОБЗОР ЛИТЕРАТУРЫ

#### A. Основные факторы, влияющие на когнитивные способности человека

Когда мы говорим о когнитивных способностях, то мы говорим не только о способности думать, рассуждать, но и понимать и обрабатывать информацию. Когнитивные способности человека – это ментальные процессы и навыки, которые используются для восприятия, понимания, анализа и обработки информации. Они включают в себя такие аспекты, как внимание, память, мышление, воображение, решение проблем, языковые навыки и другие когнитивные функции. Когнитивные способности играют важную роль в понимании окружающего мира, обучении, принятии решений и выполнении различных задач. Они также являются основой для развития интеллекта у человека.

Можно выделить следующие факторы, которые оказывают влияние на когнитивные способности: генетическая предрасположенность; уровень образования; физическое и психическое здоровье; социальное окружение; возраст. Важно отметить, что каждый человек уникален, и влияние этих факторов может быть различным для каждого.

#### B. Влияние искусственного интеллекта на когнитивные способности человека

Когда стало понятно, какой потенциал есть у систем ИИ многими странами и международными организациями принято решение о начале регулирования этой деятельности. Однако, с точки зрения использования ИИ человеком, фактически нет никаких ограничений. Давайте рассмотрим, какое влияние он может оказывать на жизнь людей.

В 2022 году вышла большая работа «Когнитивные процессы человека и искусственный интеллект в контексте цифровой цивилизации» (И. М. Дзялошинский), в которой автор очень подробно рассматривает когнитивные процессы сквозь призму ИИ. Выполненный автором анализ показал, что человеческий интеллект представляет собой «продукт многотысячелетней истории развития и взаимодействия трех взаимосвязанных процессов: 1) становления и развития человеческой личности; 2) формирования сложных сетевых отношений между членами социального сообщества; 3) коллективной деятельности как основы существования и развития сообществ и индивидов. Искусственный интеллект представляет собой комплекс технологических решений, имитирующих когнитивные процессы человека. В силу этого он всегда будет связан с деятельностью человека. Другими словами, только люди (а не машины) являются конечным источником и определителем ценностей, от которых зависит любой искусственный интеллект. За все, что делает искусственный интеллект, отвечают люди» [5].

Татьяна Черниговская говорит о том, что человек слишком упростился и продолжает упрощаться, в то время как мир все время усложняется. Многие человеческие навыки уже не нужны, их доверили технологиям.

С точки зрения И. М. Дзялошинского, «искусственный интеллект – это всего лишь ресурс, который может быть использован эффективно или не очень, во благо или во зло. Как и любой ресурс, он используется и будет использоваться в зависимости от особенностей экономической и политической системы, которую конкретное сообщество использует для адаптации к изменяющимся условиям жизни». [5]

В подтверждении этих слов, можно найти множество исследований, где ИИ позиционируется как инструмент для образования, «способствующий развитию познавательных и педагогических навыков. Это перспективная технология, которую можно использовать для достижения образовательных целей. Но каждый учитель должен принять участие в изучении и адаптации новых технологических изменений, чтобы улучшить процесс преподавания». [6]

### IV. МАТЕРАИЛЫ И МЕТОДЫ

В качестве инструмента определения уровня влияния использования искусственного интеллекта на когнитивные способности школьников был проведен опрос. Он состоял из следующих вопросов:

1. Пользуетесь ли вы искусственным интеллектом? (да/нет).
2. Если да, то какими именно? (перечислить варианты).
3. Как вы считаете, на какую вашу когнитивную функцию ИИ повлиял больше всего? (внимание, память, способность анализировать и делать выводы, планирование, ориентация в пространстве и времени).
4. Опишите, как ИИ повлиял на вашу способность планировать (от 1 до 5, где 1 – не повлиял, 5 – повлиял в полной мере).
5. Опишите, как ИИ повлиял на вашу способность ориентироваться на местности и во времени (от 1 до 5, где 1 – не повлиял, 5 – повлиял в полной мере).
6. Опишите, как ИИ повлиял на ваше внимание (от 1 до 5, где 1 – не повлиял, 5 – повлиял в полной мере).
7. Опишите, как ИИ повлиял на вашу память (от 1 до 5, где 1 – не повлиял, 5 – повлиял в полной мере).
8. Опишите, как ИИ повлиял на вашу способность анализировать и делать выводы (от 1 до 5, где 1 – не повлиял, 5 – повлиял в полной мере).

Опрос проводился методом стихийной выборки и распространялся с помощью социальных сетей. Количество респондентов – 66 человек, школьники г. Якутска (Республика Саха (Якутия), Россия) с 5 по 11 класс. При генеральной совокупности 30000 человек, размер выборки соответствует доверительной вероятности в 90 % с доверительным интервалом в 10 %.



## V. РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

По результатам опроса, 87,7 % респондентов в той или иной степени пользовались технологией искусственного интеллекта.

Наибольшее количество респондентов использовало сервис Chat GPT (69,2 % или 45 человек), голосовыми помощниками, от различных производителей, пользуются 50,8 % (33 человека), 40,6 % (30 человек) используют чат-боты на основе искусственного интеллекта в различных приложениях.

Большинство респондентов, отвечая на вопрос «На какую вашу когнитивную функцию ИИ повлиял больше всего» выделили – способность анализировать и делать выводы (66,7 %). Значительно меньшее влияние наблюдается на способность планирования (10,6 %), память (9,1 %), внимание (7,6 %) и ориентацию в пространстве и времени (6 %). Данный факт также подтверждается при детальном рассмотрении каждой рассматриваемой когнитивной функции.

Из 5 рассматриваемых когнитивных функций (рис. 2), по четырем, большинство респондентов поставили уровень влияния как средний (3 из 5). Однако у показателя «способность анализировать и делать выводы» – преобладание оценки 4 (выше среднего) отмечается у 23 человек (34,8 %), еще 11 человек (16,7 %) отмечают этот показатель как высокий уровень влияния (оценка 5).

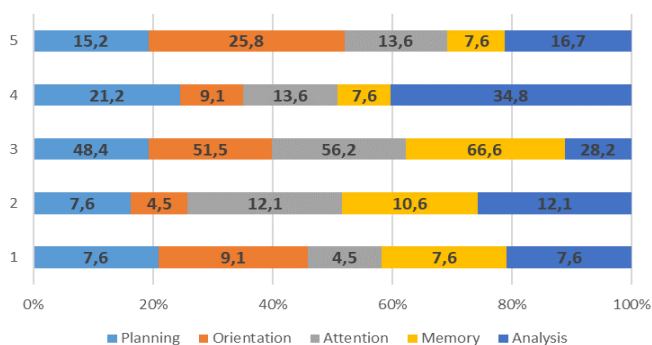


Рис. 2. Оценка уровня влияния ИИ на собственные когнитивные функции респондента по 5-и бальной шкале

Полученные результаты исследования говорят о том, что большинство школьников уже сейчас понимают, что ИИ влияет на их когнитивные способности. Однако субъективная оценка нуждается в дальнейшей объективной проверке и сравнении полученных данных.

Кроме этого, полученные результаты соотносятся и с другими исследованиями. Данный факт подтверждает анализ исследований (139 работ с 1998 по 2018 гг.) по воздействию электронных устройств и Интернета на состояние когнитивных функций молодого поколения, где делается вывод о том, «что ситуация погружения подрастающего поколения в цифровую электронную среду влияет, прежде всего, на изменение мозга, выступающего в качестве биоплатформы для формирования когнитивных функций у школьников и студентов. А изменение познавательных функций не может не влиять на когнитивные процессы внимания, памяти, мышления, а также на навыки, необходимые для

процесса познания (содержание мыследеятельности, умственную работоспособность, умение критически оценивать факты и потоки информации и пр.)» [7].

На наш взгляд, применение технологии ИИ в образовании должно быть четко регламентировано и обосновано, в то же время исследовано. Потому что, необходимо четко понимать к каким последствиям это может привести. Несмотря на преимущества использования ИИ в образовании, как «персонализированное обучение, экономия времени, помощь в обучении, и др.» [8], проблемам влияния ИИ на когнитивные функции человека уделяется не достаточное внимание, все больше говорят о таких проблемах как «этические соображения, технологические ограничения, отсутствие возможности интерпретации, доступность». [8]

Все больше исследователей сходятся во мнениях, что есть большие перспективы в использовании когнитивных технологий, которые «сочетают в себе методы искусственного интеллекта с элементами человеческих когнитивных процессов, что приводит к более продвинутому решению проблем, пониманию естественного языка и возможностям принятия решений» [9]. Данный факт говорит о том, чтобы использовать одновременно преимущества ИИ и когнитивных способностей человека.

## VI. ЗАКЛЮЧЕНИЕ

Таким образом, в рамках данной работы положено начало исследования влияния искусственного интеллекта на развитие когнитивных функций школьников. Получены данные о том, что с субъективной точки зрения школьников, ИИ влияет на их когнитивные функции. Исходя из полученных данных, школьники отмечают, что уровень воздействия искусственного интеллекта на их когнитивные способности достаточно весомый. Однако эти данные нуждаются в объективной проверке и сравнении с полученной субъективной оценкой.

В настоящее время использование искусственного интеллекта становится все более актуальным и широко применяется в различных сферах нашей жизни. Одной из таких сфер является образование, где применение искусственного интеллекта может значительно оптимизировать и улучшить процесс обучения. Однако, для полного понимания эффективности и проблем использования искусственного интеллекта в образовании, необходимо более детально изучить его влияние на когнитивные аспекты.

## БЛАГОДАРНОСТИ

Благодарим Всероссийскую образовательную инициатива по поиску и реализации научно-технологических проектов «Сириус.Лето: начни свой проект».

## СПИСОК ЛИТЕРАТУРЫ

- [1] Firth J, Torous J, Stubbs B, et al. "Internet-brain": how the Internet can change our perception. *World Psychiatry*. 2019;18(2):119-29. <https://doi.org/10.1002/wps.20617>.
- [2] Takeuchi H, Taki Y, Asano K, et al. Impact of frequency of internet use on development of brain structures and verbal intelligence:

- longitudinal analyses. *Psychology, Medicine. Human Brain Mapping*. 2018;39:4471-9. <https://doi.org/10.1002/hbm.24286>.
- [3] Дзен: Как использование ИИ влияет на когнитивные способности людей. URL: <https://dzen.ru/a/ZlJvGc5tF0eVfKху> (дата обращения: 20.10.2023).
- [4] Банк России: Применение искусственного интеллекта на финансовом рынке (доклад для общественных консультаций). URL: [https://www.cbr.ru/Content/Document/File/156061/Consultation\\_Paper\\_03112023.pdf](https://www.cbr.ru/Content/Document/File/156061/Consultation_Paper_03112023.pdf) (дата обращения: 11.10.2023).
- [5] Дзялошинский И.М. Когнитивные процессы человека и искусственный интеллект в контексте цифровой цивилизации: Монография. Москва : Ай Пи Ар Медиа, 2022. 583 с. ISBN 978-5-4497-1596-8.
- [6] Graaff, Armand & Smit, Danie & Eybers, Sunet. (2023). Let's Play Games: Using No-Code AI to Reduce Human Cognitive Load During AI Solution Development. 10.1007/978-3-031-49002-6\_7.
- [7] Авдеева Е.А., Корнилова О.А. Влияние цифровой электронной среды на когнитивные функции школьников и студентов // Кардиоваскулярная терапия и профилактика. 2022. Т. 21, № S3. С. 43-50. DOI 10.15829/1728-8800-2022-3331.
- [8] Umar Ismail. (2024). Artificial intelligence (AI) in education. *Education Next*. 1. 1-7.
- [9] Kharb, Latika & Chahal, Deepak & Singh, Vaibhav. (2023). Harnessing the superiority of cognitive technology over artificial intelligence. *European Chemical Bulletin*. 12. 6990-6994. 10.48047/ecb/2023.12.10.505.

# Using an Evolutionary Algorithm such as Genetic Algorithm (Ga) to Optimize Linear Equations

Abdul sttar Ismail Wdaa

College of Education for pure  
Science, University of Anbar  
Ramadi, Iraq

Corresponding Author Email:  
sttarwdaa2019@uoanbar.edu.iq

Shokhan M. Al-Barzinji

Department of Computer Networks  
Systems, College of Computer  
Science and Information  
Technology, University of Anbar  
Ramadi, Iraq

shokhan.albarzinji@uoanbar.edu.iq

Hiba A. Tarish

Civil Engineering Departments,  
University of Technology Baghdad  
Iraq

hiba.a.tarish@uotechnology.edu.iq

**Abstract**— An Evolutionary algorithm is a subset of evolutionary computations. Cloning, mutation, recombination, and selection are some of the methods used by the evolutionary algorithm that were inspired by biological evolution. The problem that was addressed in this study is how to find the coefficient values of the variables in linear equations that contain a set of variables, it must equal the right side. After reviewing previous research looking for optimization, genetic algorithm, which will use in this paper, so how are going to use this strategy to solve our problems? Select the best chromosomes from this population. Perform random chromosomal hybridization. Then, within the genes of our chromosomes, produce random mutations. Will send these chromosomes for selection again. The best chromosomes among them are selected again. It crosses and turns again, and the process is repeated until a certain number of generations have passed or find the best option. After applying the genetic algorithm through the Matlab program to the linear equation, which contains six variables, for example, the results were very excellent, as the parameter values were found in a very short time and with high accuracy, completely matching the right side of the equation.

**Keywords**—Evolutionary Algorithm (EA), Genetic Algorithms (GA); Linear Equations, Optimization, Artificial Intelligence (AI)

## I. INTRODUCTION

The expression “artificial intelligence” (AI) is utilized to depict the capacities of machines like PCs and robots to finish assignments that regularly require human intelligence. The expression is utilized to depict endeavors to make machines that can do mental undertakings like those of people [1]. These errands include the ability to reason, identify, summarize, and gather facts. Since the advent of advanced PCs in the 1940s, it has become clear how they can be instructed to perform unusually complex exercises [1]. Examples of artificial intelligence include transformative algorithms [2]. The Hereditary calculation, a developmental calculation, was utilized to address straight conditions in this exploration. Optimization, whether it is the minimization or maximization. Every person in the population is assessed for their environmental adaption (fitness). Optimization problems and learning problems are correlated with each other. Sometimes the function that needs to be optimized is so complex that cannot afford to calculate the objective value for every solution [3]. As the history of the field suggests, there are many different variants of evolutionary algorithms. The common underlying idea behind all these techniques is the same: given a population of individuals within some

environment that has limited resources, competition for those resources causes natural selection (survival of the fittest). Research problem the ability of a genetic algorithm to find the optimal solution in an ideal time to solve a problem consisting of six variables [4].

Studies that the use of the genetic algorithm is one of the best artificial intelligence algorithms.

The purpose of this research is to use one of the most important evolutionary algorithms, which is the genetic algorithm to solve complex problems in linear equations, where the research methodology includes an introduction to evolutionary algorithms, linear equations, and the mechanism of work of the genetic algorithm, after that the proposed solution was applied, then sensitivity analysis and conclusion. For more information, researchers refer to the following references [5, 6].

## II. THE EVOLUTIONARY ALGORITHMS

Incorporates a gathering of people (chromosomes) that develop through a blend of determination, hybrid, and mutation. every single one of them with regards to maximized execution. The amplified or limited capability is assessed for this situation. Through hybrid and change, the best quality mixes (individuals) are picked through choice. This ought to bring about better arrangements in the following populace. The following Fig. 1 shows the flowchart of the evolutionary algorithm 7.

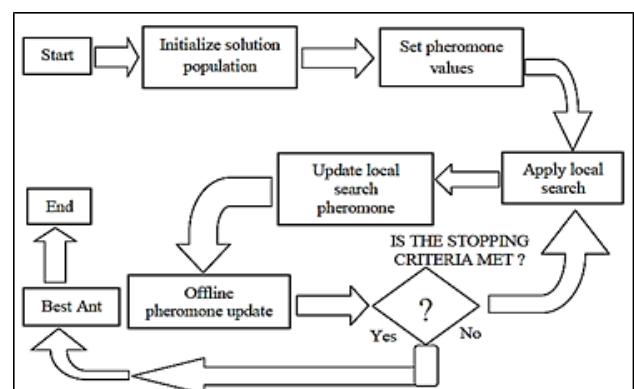


Fig. 1. Evolutionary Algorithms7

The diagram above shows the process of evaluating the solution area to see if the global solution is viable was found. If not, then the current ants will go through an exponential pheromone decrease before starting the next

iteration in the optimization process. This pheromone decay reduces the impact of hysteresis on the solution space and helps prevent premature convergence.

III. LINEAR EQUATIONS (LE)

A mathematical statement with the equal symbol '(=)' between the algebraic expressions is called an equation. Linear equations are equations with a degree of one. The formula for a straight line is this. Replacing the answer of the linear equation with an unknown value makes the equation true. With one variable, there is only one solution. There is just one solution to the equation which is The Cartesian coordinates of a point in the Euclidean plane are obtained as the solutions to the two-variable linear equation. In this study, the focus was on the genetic algorithm, which is characterized by its high ability to solve complex linear equations, by applying an equation that contains several variables, provided that this equation is equal to a certain number 13.

A. How the genetic algorithms works

One of the most well-known machine learning techniques is genetic algorithms (GAs). They are conceptually similar to natural selection. A parallel search is used by GA. Apply crossover to choose persons at random from a pool of prospects (exchange candidate information) and alter the candidates (perturb) information) until the entire system converges (all candidates are equal). If the system satisfies some user-defined requirement [14]. Every a generation is a collection of selection, crossover, and mutation processes. The quantity the number of generations can be used to determine how quickly a GA model converges. In GA, the current generation's candidates are referred to as "the." The population of candidate datasets will be referred to as "the GA population" to distinguish it from the "population" commonly use in statistics. Binary strings are the most popular representation of individuals in GA populations (effectively one-dimensional arrays). There is a scarcity of work on what are known as matrix-GAs (two-dimensional arrays). There are compelling reasons to utilize a matrix structure for synthetic data creation. To begin, Microdata is data gathered directly from individual population units, and the matrix is the most frequent way for displaying it [15]. Point out, matrix GAs can best present data structures of their candidates, allowing them to optimize not only records but also their data structures. To begin, let's define genetic algorithms as a specific subclass of stochastic search algorithms inspired by natural selection. Fig. 3 is a simplified illustration of a basic GA, which uses a binary string format for candidates and a well-defined problem. Generally speaking, a genetic algorithm (GA) for a specific function works will go as follows: 15, 16.

St1: Decide the populace size (N), hybrid recurrence (pc), and transformation rate (pm) for a bunch of chromosomes that addresses the issue's variable domain.

St2: Create a fitness function to measure how well one chromosome does in the target environment. When reproducing, the fitness function is utilized to select the optimal pairing of chromosomes.

St3: Produce an initial population of chromosomes, of size N, using random sampling. (1) shows that the number of variables N and in this study N = 6.

$$x_1+x_2+x_3+\dots+x_n \tag{1}$$

St4: Calculate the fitness of each individual chromosome:

$$f(x_1), f(x_2), f(x_3) \dots \dots \dots f(x_n) \tag{2}$$

St5: Select mates from populations based on chromosomal admixture. Selection of parental chromosomes based on probability of happiness. Highly matched chromosomes are more likely to be selected for mating than poorly matched chromosomes.

St6: Utilizing the hereditary administrator's hybrid and change, make a pair of posterity chromosomes.

St7: Place the made posterity chromosomes in the new populace.

St8: Repeat step 5 until the size of the new chromosome population matches the size of the original population N. St9: Substitute the offspring chromosomal population for the original one.

St10: Continue to Step 4 until the termination requirement is met.

A GA, as can see, is an iterative procedure. A generation is the name given to each iteration. For a simple GA, the number of generations can range from 50 to over 500. The expression "run" is utilized to depict a progression of ages. Toward the finish of each run, desire to track down no less than one exceptionally fit chromosome. Is there a standard method for preventing a hereditary calculation from running? A probabilistic hunting system is used in these calculations, so population health can remain stable for some time until the emergence of superior chromosomes in GA. This makes it even more difficult to use standard termination rules. A typical technique is to stop GA after a given number old enough gatherings and decide the best chromosome in the populace. Fig. 3 shows a flowchart of the hereditary cycle computation utilized expecting deficient outcomes. Is there a standard method for keeping legacy estimations from running? Since these estimations utilize probabilistic following methods, populace wellbeing can remain stable for quite a while until predominant chromosomes arise in GA. This makes utilizing standard check impedes more troublesome. A typical strategy is to stop GA after a given number old enough gatherings and decide the best chromosome in the populace. In the event that no ideal situation is found, run GA again [15].

B. How the genetic algorithms works

If the problem of this research shown in Table 1 is how to get a representation of it with six inputs and one output y. If I had data consisting of one sample, I would have a representation of it with six inputs and one output 13

TABLE I. VALUE OF 6 INPUT ONE OUTPUT 17

	X1	X2	X3	X4	X5	X6	Y
DATA	4	2	7	5	11	1	44.1

Need a search model that ties the inputs and outcomes together. If the model is employed, as defined in (3). This

shows the relationship between the input and the target on the right side.

$$y' = w_1x_1 + w_2x_2 + w_3x_3 + w_4x_4 + w_5x_5 + w_6x_6 \quad (3)$$

So that each value of x is a value corresponding to w in order for the input and output to be successful, need to choose the best values of w from w1.....w6 that satisfies the equation according to the (3).

The initial expected value is calculated according to (3).

Goal is to find the set of parameter (w1:w6) that maps the following input to its output. Can substitute from the list of inputs, 18 which produces the following (4). As can see the diagram above GA represents an iterative process. Each iteration is called a generation. The typical number of generations for GA can range from the entire generational group called the range. In At the end of the run, expect to find one or more highly fit chromosomes. Initial values are weighted as in the following Table 2.

TABLE II. INITIAL VALUES ARE WEIGHTED 18.

Initial weight	W1	W2	W3	W4	W5	W6
	2.4	0.7	8	-2	5	1.1

Substituting the weight values into (4). The equation was previously defined.

$$y' = 4 * 2.4 - 2 * 0.7 + 7 * 8 + 5 * (-2) + 11 * 5 + 1.1$$

$$y' = 110.3$$

Expected output

Calculate the error between the original value and the expected value. Through (5). Calculate Absolute error.

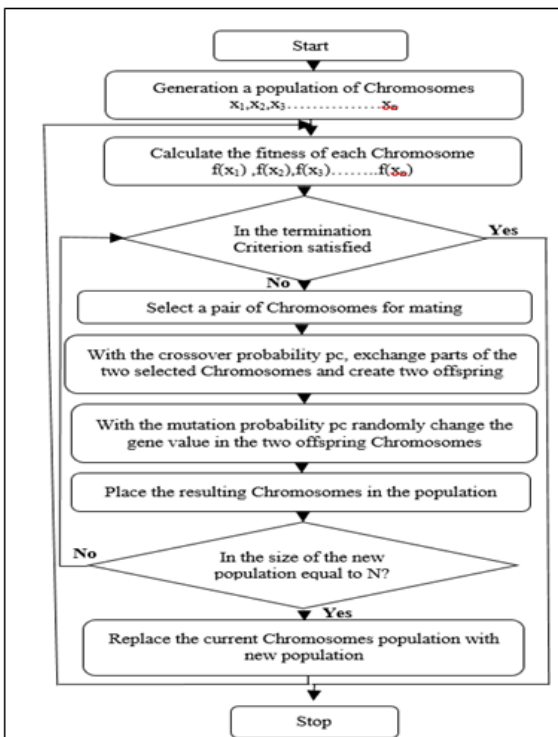


Fig. 2. Cycle Genetic Algorithm [15–16]

$$\begin{aligned} \text{error} &= |y - y'| \\ \text{error} &= |44.1 - 110.3| \\ \text{error} &= 66.2 \end{aligned} \quad (4)$$

This means that there is an error in the search model in the weights values, so the weights values must be modified Table 3.

TABLE III. UPDATE INITIAL VALUES ARE WEIGHTED

Update weight	W1	W2	W3	W4	W5	W6
	-0.4	2.7	5	-1	7	0.1

$$y' = 4x_1 - 2x_2 + 7x_3 + 5x_4 + 11x_5 + x_6$$

$$y' = 4 * 0.269 - 2 * 0.45922 + 7 * 0.78577 + 5 * 2.6107 + 11 * 2.3263 + (-0.031)$$

$$y' = 41.1011$$

Absolute error

$$\begin{aligned} \text{error} &= |44.1 - 100.3| \\ \text{error} &= 56 \end{aligned}$$

There is still an error so the weights values must be adjusted Table 4.

TABLE IV. UPDATE WEIGHTED17

Update weight	W1	W2	W3	W4	W5	W6
	-1	2	2	-3	2	0.9

Absolute error

$$\begin{aligned} \text{error} &= |y - y'| \\ \text{error} &= |44.1 - 13.9| \\ \text{error} &= 30.2 \end{aligned} \quad (4)$$

The loads should be changed so mistake = 0 is reached. It is trying to distinguish the ideal arrangement without depending on an enhancement approach like a hereditary calculation (GA). Abstract of the Natural Setting for Hereditary Calculations Picture 1 Hereditary calculations mirror the manner in which living beings develop through time. A little science boost will aid with (GA) understanding. Picture (5) [19].

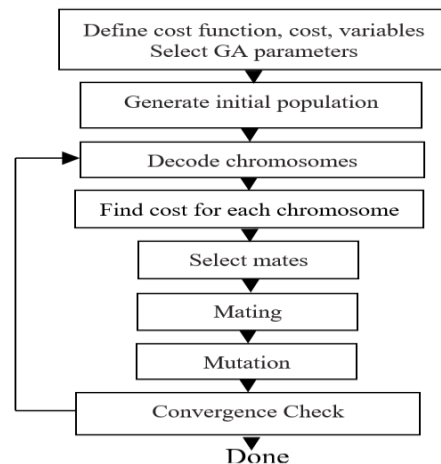


Fig. 3. Flowchart of a binary GA [19]

The figure above GA, like any other optimization algorithm, begins with the optimization identification.

Variables and cost function. Ends like other optimization Algorithms too, by convergence testing. However, this algorithm is completely different. The path through GA components is Appears as a flowchart in the previous figure. In fact all of the methods requiring a gradient of the cost function won't work well with discrete data. The GA has no problem.

#### IV. WHAT ARE THE GENS?

A gene is anything that can improve outcomes when altered. Examining the following model, 6 weights improve the result, so each weight represents a gene in the GA. Table 5 shows the specific gravity of each gene 19-20. The new expected value is calculated from (4). The equation is already defined.

TABLE V. COMPARING GENES WITH WEIGHT

Gene 0	Gene 1	Gene 2	Gene 3	Gene 4	Gene 5
w1	w2	w3	w4	w5	w6

$$y' = w_1x_1 + w_2x_2 + w_3x_3 + w_4x_4 + w_5x_5 + w_6x_6 \quad (3)$$

$$f(c) = \frac{1}{error} = \frac{1}{|y - y'|} \quad (4)$$

#### V. INITIAL POPULATION OF SOLUTION (GENERATION 0) POPULATION SIZE =6

To solve such problems by using the gene algorithm from the first gene that contains (IPS) Initial population of solution this solution is not the best, but it is the initial solution, and when evolution obtains the solution, it can produce the best solution: a population that contains a set of solutions. Genetics is based on Darwin's theory. The best individual advances to the second solutions. In this problem, choose weight values for a total of chromosomes, let it be 6 chromosomes. Table 6 shows the initial weights for each chromosome.

TABLE VI. INITIAL WEIGHTS FOR EACH CHROMOSOME 18

w1	w2	w3	w4	w5	w6
2.4	0.7	8	-2	5	1.1
-0.4	2.7	5	-1	7	0.1
-1	2	2	-3	2	0.9
4	7	12	6.1	1.4	-4
3.1	4	0	2.4	4.8	0
-2	3	-7	6	3	3

Table 7 represents one of the chromosomes and the location of the gene.

TABLE VII. CHROMOSOME AND GENE 18

Chromosomes					
-1	2	2	-3	2	0.9

The following Fig. 4 represents the steps of solving using the algorithm with how to find the fitness functions f(c) and fitness value.

The figure above is Chromosomal Fitness Assessment - It should be taken into account that the function type affects the assessment of chromosomal fitness (vector). However, the important feature of the genetic algorithm is that one feature is also used to calculate the characteristic values of all vectors of the initial set.

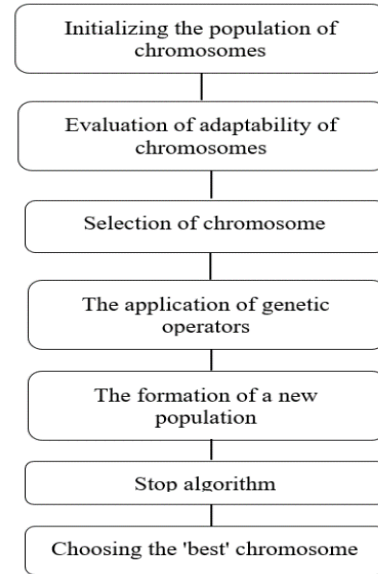


Fig. 4. Stages evaluate fitness value[20]

#### VI. CALCULATED FITNESS FUNCTIONS

Now it is fitness functions calculated through equation number 8

$$f(c) = \frac{1}{error} = \frac{1}{|y - y'|}$$

$$f(c) = \frac{1}{|44.1 - 110.3|} = 0.015 \quad (5)$$

ستة مرات f© من خلال هذه المعادلة يتم حساب جميع دوال

#### VII. INITIAL POPULATION OF SOLUTION (GENERATION 0)

The following Table 8 shows the values of y' and f(c) for each chromosome, From Table 6, the highest available values of fitness value f(c) are chosen after calculating the Expected output

TABLE VIII. CALCULATE EXPECTED OUTPUT AND FITNESS VALUE 17

w1	w2	w3	w4	w5	w6	y'	f(c)
2.4	0.7	8	-2	5	110.3	110.3	0.015
-0.4	2.7	5	-1	7	100.1	100.1	0.018
-1	2	2	-3	2	13.9	13.9	0.033
4	7	12	6.1	1.4	127.9	127.9	0.012
3.1	4	0	2.4	4.8	69.2	69.2	0.04
-2	3	-7	6	3	3	3	0.024

#### VIII. MATING POOL


In evolutionary computation, a mating pool is a notion. Virtual populations are usually handled as a whole. Another option is to isolate only those people who will produce offspring from the rest of the population. These are then thrown into a breeding pool 21. Checking for potential parents is a common practice. After computing the Expected output(y'), the highest existing values of fitness

value  $f(c)$  are picked from Table 7 and the chromosomes that have the greatest value (3-5-6) are selected as shown in Fig. 5.

TABLE IX. HIGHEST VALUE FROM 18

w1	w2	w3	w4	w5	w6	y'	f(c)
2.4	0.7	8	-2	5	110.3	110.3	0.015
-0.4	2.7	5	-1	7	100.1	100.1	0.018
-1	2	2	-3	2	13.9	13.9	0.033
4	7	12	6.1	1.4	127.9	127.9	0.012
3.1	4	0	2.4	4.8	69.2	69.2	0.04
-2	3	-7	6	3	3	3	0.024

Choose the top 3 offspring-bearing individuals from the available pool Fig. 5.



W1	W2	W3	W4	W5	W6
-1	2	2	-3	2	0.9
3.1	4	0	2.4	4.8	0
-2	3	-7	6	3	3

Fig. 5. Best 3 individuals to the mating pool

In the above figure, the top three fitness value is chosen, and then the mating pool process is performed between the first chromosome with the second and third in the first stage, while the second stage is crossover between the second and third chromosomes.

### IX. CROSSOVER

The center idea of GA is hybrid, which helps produce posterity conveying qualities from a few chromosomes. At the point when two chromosomes, called guardians, combine, they produce a third, called the posterity. A chromosome in the populace is picked in light of the expressed wellness capability. Hybrid can happen at a certain point, two focuses, a few focuses, or be uniform. The chromosomes that outcome from the mating system are assessed again for their wellness22. Hybrid is shown in Fig. 6.

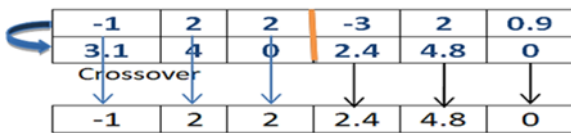


Fig. 6. Crossover

The Figure 6. Above shows the crossover process where the first three elements are selected from the first chromosome with the second three elements from the second chromosome.

### X. MUTATION

GA relies heavily on mutation. The mutation operator introduces random changes to the chromosomes' characteristics. Mutation is frequently considered as a background operator that helps to sustain population genetic diversity. Some of the genetic structure's building pieces are altered by mutation. In GA, mutation is a proven strategy for escape the trap of local minima and maintaining population diversity [18–22].

The operation offspring and mutant showing in Fig. 7.

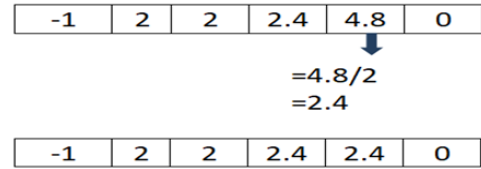


Fig. 7. Offspring & Mutant

Figure above shows the process Offspring & Mutant do pick chromosomes from the parent population, recombine them via mutation and crossover, and then pass on the resulting offspring. Figure out how well the new population is doing; supplant the populace with healthier people; while this has been accomplished to my satisfaction, Fitness value.

In the table below, new population (generation 1) shows that New Individuals is the product of the mating pool operation performed on old Individuals.

Figure 6. Represents the process that occurs in a genetic algorithm. Through this process, six population operations are performed that randomly represent the chromosomes (populations). This happens in other generations as well. At each generation, fitness functions are computed separately for all chromosomes and the best chromosome is selected. Based on the best fit function, then the mating pool process, then the crossover process, and finally based on the previously mentioned mutation process [19, 23].

TABLE X. GENERATION 118

<b>Old Individuals</b>	-1	2	2	-3	2	0.9
	3.1	4	0	2.4	4.8	0
	-2	3	-7	6	3	3
<b>New Individuals</b>	-1	2	2	2.4	2.4	0
	3.1	4	0	6	1.5	3
	-2	3	-7	-3	1	0.9

New population (generation 1 & calculate (y'), f(c))

TABLE XI. EXPECTED OUTPUT & FITNESS VALUE 18

						y'	f(c)
<b>Old Individual</b>	-1	2	2	-3	2	0.9	13.9
	3.1	4	0	2.4	4.8	0	69.2
	-2	3	-7	6	3	3	0.024
<b>New Individual</b>	-1	2	2	2.4	2.4	0	44.4
	3.1	4	0	6	1.5	3	53.9
	-2	3	-7	-3	1	0.9	-66.1

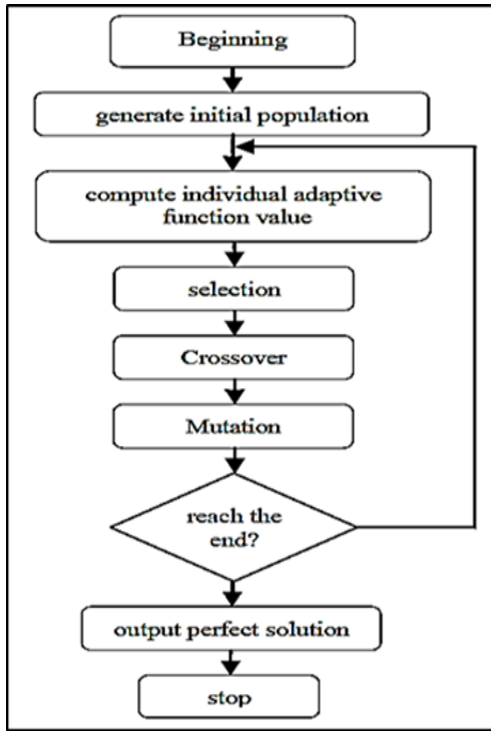


Fig. 8. Genetic Algorithm Operations[24]

XI. RESULTS FOR FUNCTION

The following Table [8] having six available (4). Has been considered as test problem:

$$y = w_1x_1 + w_2x_2 + w_3x_3 + w_4x_4 + w_5x_5 + w_6x_6$$

$$y = 44.1$$

$$(w_1, w_2, w_3, \dots, w_6) \tag{2}$$

$$y' = 4x_1 - 2x_2 + 7x_3 + 5x_4 + 11x_5 + x_6$$

$$y' = 4 * 0.269 - 2 * 0.45922 + 7 * 0.78577 + 5 * 2.6107 + 11 * 2.3263 - 0.031$$

$$y' = 41.1011 \tag{3}$$

Table 8 when applying equation 4 shows that the value of the population where different samples were tested (200-150-300-100) showed that the best sample size when (=100 population) and that the best generation out of a mixed total (300-250-100) where the best = 250.

Generation and that the best Mutation scale among the different values tested (1.2-1.13) is 1.13 = Mutation scale. From the table it is clear that the value of weight has changed until reaching the best value which represents the right side from the equation, this means the best values of weight that achieve the target equation, which is as follows.

Population = 100  
 Generation = 250  
 Mutation scale=1.13

When calculating the objective function and the error rate from (5)

$$y' = 4x_1 - 2x_2 + 7x_3 + 5x_4 + 11x_5 + x_6$$

$$y' = 4 * 0.269 - 2 * 0.45922 + 7 * 0.78577 + 5 * 2.6107 + 11 * 2.3263 + (-0.031)$$

$$y' = 41.1011 \tag{3}$$

Find that the value of the absolute error is very small, as follows:

Absolute error

$$\text{error} = |y - y'|$$

$$\text{error} = |44.1 - 41.1011|$$

$$\text{error} = 0.0011 \approx 0 \tag{5}$$

TABLE XII. MUTATION SCALE, GENERATION, POPULATION (SIZE) FITNESS FUNCTION AND WEIGHT

Mutation scale	Generation	Population (size)	w1	w2
1.2	150	200	1.0447	1.0779
1.2	150	200	0.85707	0.89348
1.2	150	200	-0.46347	-0.32055
1.2	150	150	0.082905	0.029876
1.2	300	150	1.3354	-0.32101
1.13	300	300	2.7719	0.83674
1.13	300	300	0.42377	-0.91714
1.13	300	300	1.7164	1.358
1.13	300	150	0.27447	1.0162
1.13	250	100	1.0621	1.4574
1.13	250	100	0.12534	1.4262
1.13	250	100	1.8071	1.5546
1.13	250	100	0.2269	0.45922

Figure 10. Shows the best fitness value for each chromosome and Fig. 10: show search of the best weight.

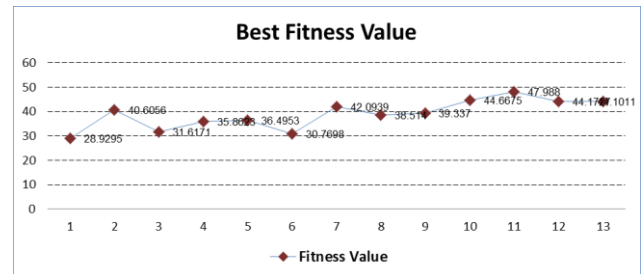


Fig. 9. Best fitness value

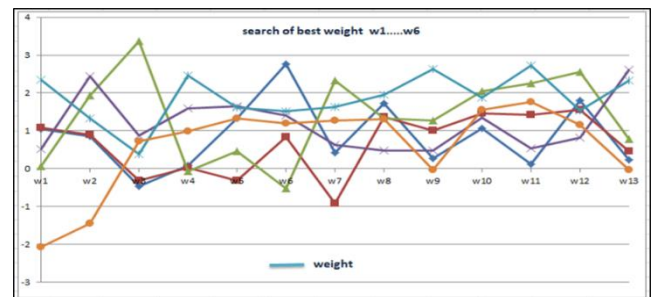


Fig. 10. Search of the best weight w1.....w6

XII. SENSITIVITY ANALYSIS

Union in GA depends vigorously on the upsides of its boundaries. The proposed technique was tried with a few upsides of the scaling and hybrid capabilities. Subsequently,



this part gives a responsiveness examination of the proposed strategy by introducing a change of the hybrid capabilities and scaling capability to inspect the boundaries of the recommended calculation.

Figure 11. Is a diagram portraying the development of GA settings? The littlest conceivable figure was utilized to decide the populace size and number of ages. It is essential to see that for each situation, the roulette wheel determination capability and the Gaussian change capability combine to their ideal qualities. Changing the loads to get however close as conceivable to the goal capability worth might have been done when different hybrid capabilities were used, like single point, twofold point, and dispersed.

### XIII. RESULTS AND DISCUSSION

The accompanying thoughts can be gotten from the first segment's tests and information: Without knowing the beginning conjecture, the ideal arrangement consolidating supernatural capabilities can be found utilizing a transformative procedure in light of GA. The handiness of the proposed technique is affirmed by near investigations with past examination. You can see the results of the proposed calculation in Table 12. The best outcome I could get was 44.1727, and, surprisingly, that was off by a unimportant 0.0011. The mistake was decreased by 0.0011 thanks to the calculation's endeavors. The proposed calculation assumed a part in improving the end result. Can likewise see that the upsides of the loads created by refreshing the chromosomes containing qualities were improved, yet worse than the genuine worth accomplished by the proposed approach.

From the results, the genetic algorithm has proven to be one of the best evolutionary algorithms, especially in complex issues such as those discussed in this article. The genetic algorithm has proven is more accurate and has impressive results when the number of possibilities is very large. This examination gives a technique to tackling troublesome frameworks of conditions that can be applied across disciplines. In this review, GA was utilized to ask about how such issues could be settled. Both straight and supernatural issues have been settled utilizing GA here. The GA boundaries control the pace of assembly. The wellness capability, a fundamental component, was recently overlooked and was at first taken care of as the amount of the modulus of the singular capability

### XIV. CONCLUSION

This study presents a method for dealing with complex systems of equations in many areas. GA was utilized to investigate the solution of such problems in this study. GA has been used to solve transcendental and linear equations in the current study. GA's convergence is determined by its parameters. Initially, the fitness function, which is an important element, was treated as the sum of the modulus of the individual function, which was not taken into account in previous studies. Following that, empirical analysis was used to determine population size, mutation function, crossover function, and generation number. Sensitivity analysis was also performed to validate the methodology by adjusting the crossover and scaling functions. The experimentation took place.

### ACKNOWLEDGEMENT

The author would like to thank Alanbar University for providing the necessary approval and support, which made this research possible.

### REFERENCES

- [1] [Wang P. On defining artificial intelligence. *Journal of Artificial Intelligence*. 2019 May 1;10(2):1-37. <https://doi.org/10.2478/jagi-2019-0002>
- [2] Reier Forradellas RF, Garay Gallastegui LM. Digital transformation and artificial intelligence applied to business: Legal regulations, economic impact and perspective. *Laws*. 2021 Aug 27;10(3):70. <https://doi.org/10.3390/laws10030070>
- [3] Mitchell M. *An introduction to genetic algorithms*. MIT press; 1998 Mar 2. <https://doi.org/10.7551/mitpress/3927.001.0001>
- [4] Jin YF, Yin ZY, Shen SL, Zhang DM. A new hybrid real-coded genetic algorithm and its application to parameters identification of soils. *Inverse Problems in Science and Engineering*. 2017 Sep 2; 25(9):1343-66. <https://doi.org/10.1080/17415977.2016.1259315>
- [5] Wdaa, AS. Using differential evolution with neural networks forecasting model creating for pipeline corrosion. *Journal of Engineering and Applied Sciences*. (2018) Jan13(23):9908-9913. <https://doi.org/10.3923/jeasci.2018.9908.9913>
- [6] Abdulazeez FA, Ismail AS, Abdulaziz RS. Using Gradient Descent to an Optimisation Algorithm that uses the Optimal Value of Parameters (Coefficients) for a Differentiable Function. *International Journal of Communication Networks and Information Security*. 2023Apr1; 15(1):75-83. <https://doi.org/10.17762/ijcnis.v15i1.5718>
- [7] Sands T. *Advances in Spacecraft Attitude Control*. IntechOpen; 2020. <https://doi.org/10.5772/intechopen.77574>
- [8] Abduljabbar IA, Abdullah SM. An evolutionary algorithm for solving academic courses timetable scheduling problem. *Baghdad Science Journal*. 2022 Apr1; 19(2):0399. <https://doi.org/10.21123/bsj.2022.19.2.0399>
- [9] Abduljabbar IA, Abdullah SM. An evolutionary algorithm for solving academic courses timetable scheduling problem. *Baghdad Science Journal*. 2022 Apr 1;19(2):0399 <https://doi.org/10.21123/bsj.2022.19.2.0399>
- [10] Villarrubia G, De Paz JF, Chamoso P, De la Prieta F. Artificial neural networks used in optimization problems. *Neurocomputing*. 2018 Jan 10;272:10-6. <https://doi.org/10.1016/j.neucom.2017.04.075>
- [11] Green B, Tao T. Linear equations in primes. *Annals of mathematics*. 2010 May 1:1753-850. <http://doi.org/10.4007/annals.2010.171.1753>
- [12] <http://doi.org/10.4007/annals.2010.171.1753>
- [13] Wdaa AS. Hybrid artificial neural network and fuzzy logic for function approximation. *Journal of Theoretical and Applied Information Technology*. 2018 Nov15; 96(21). [https://www.researchgate.net/publication/329376415\\_Hybrid\\_artificial\\_neural\\_network\\_and\\_fuzzy\\_logic\\_for\\_function\\_approximation](https://www.researchgate.net/publication/329376415_Hybrid_artificial_neural_network_and_fuzzy_logic_for_function_approximation)
- [14] Jawad, R. N., & Ali, F. H. (2020). Using evolving algorithms to cryptanalysis nonlinear cryptosystems. *Baghdad Science Journal*, 17(2 (SI)), 0682-0682.
- [15] Lambora A, Gupta K, Chopra K. Genetic algorithm-A literature review. In 2019 international conference on machine learning, big data, cloud and parallel computing (COMIT Con) 2019 Feb 14 (pp. 380-384). IEEE. <https://doi.org/10.1109/COMITCon.2019.8862255>
- [16] Davis L. *Handbook of Genetic Algorithms*. Van Nostrand Reinhold, pp.222, New York, 1991. <https://Handbook of genetic algorithms by Lawrence Davis | Open Library>
- [17] Mitchell M. *An introduction to genetic algorithms*. MIT press; 1998 Mar 2. <https://doi.org/10.7551/mitpress/3927.001.0001>
- [18] Lee W, Kim HY. Genetic algorithm implementation in Python. In *Fourth Annual ACIS International Conference on Computer and Information Science (ICIS'05)* 2005 Jul 14 (pp. 8-11). IEEE. <https://doi.org/10.1109/ICIS.2005.69>

- [19] Gad AF, Gad AF, John S. Practical computer vision applications using deep learning with CNNs. Berkeley:Apress; 2018. <https://link.springer.com/book/10.1007/978-1-4842-4167-7>
- [20] Haupt RL, Haupt SE. Practical genetic algorithms. John Wiley & Sons; 2004 Jul 30. <https://doi.org/10.1002/0471671746>
- [21] Stepanov LV, Koltsov AS, Parinov AV, Dubrovin AS. Mathematical modeling method based on genetic algorithm and its applications. InJournal of Physics: Conference Series 2019 Apr 1 (Vol. 1203, No. 1, p. 012082). IOP Publishing. <https://doi.org/10.1088/1742-6596/1203/1/012082>
- [22] Yu X, Gen M. Introduction to evolutionary algorithms. Springer Science & Business Media; 2010 Jun 10. <https://doi.org/10.1007/978-1-84996-129-5>
- [23] Ahmed ZH. Genetic algorithm for the traveling salesman problem using sequential constructive crossover operator. International Journal of Biometrics & Bioinformatics (IJBB). 2010Jan; 3(6):96. <https://doi.org/10.14569/IJACSA.2020.0110275>
- [24] Hassanat A, Almohammadi K, Alkafaween EA, Abunawas E, Hammouri A, Prasath VS. Choosing mutation and crossover ratios for genetic algorithms—a review with a new dynamic approach. Information. 2019Dec10; 10(12):390. <https://doi.org/10.14569/IJACSA.2020.0110275>.
- [25] Meng X, Song B. Fast genetic algorithms used for PID parameter optimization. In2007 IEEE International Conference on Automation and Logistics 2007 Aug 18 (pp. 2144-2148). IEEE. <https://doi.org/10.1109/ICAL.2007.4338930>

# The Impact of Digital Accounting Technologies in Achieving the Quality of Accounting Information

Ammar Ghazi Ibrahim

Faculty of Management and Economics - Accounting  
Department  
Mosul University  
Mosul, Iraq  
ammar.ezzi82@gmail.com

Kubra Mohammed Tahir

Faculty of Management and Economics - Accounting  
Department  
Mosul University  
Mosul, Iraq  
mohamed\_kubraa@uomosul.edu.iq

**Abstract— Introduction.** Digital technologies are more and more present in the financial accounting activities of large companies, as well as companies specialising in the provision of accounting and auditing services, because of the large amount of data required for processing and because of the reduction of costs, processing time. **Aim.** It is of interest to evaluate the world experience in the application of different digital technologies in accounting, to assess their advantages and disadvantages. **Material and methods.** This study used analysis and synthesis as methods, namely analysing various digital technologies such as Robotic Process Automation, Artificial Intelligence, machine learning, blockchain and others and synthesising them in accounting and evaluating their advantages and disadvantages. **Results and Conclusion.** The accounting profession can play an even more important role in the future by adapting and mastering the demands of new emerging technologies. Future generations of successful professional accountants will need, in addition to traditional knowledge, additional knowledge in areas called IT, AI and RPA, but also develop social skills. The results of the study showed that applying artificial intelligence in conjunction with Robotic Process Automation in the accounting department can automate its work, reducing routine tasks by up to 95%. Despite the greater speed of processing information with cloud-based systems, accounting departments are less likely to utilise them due to data security. Blockchain digital technology has a high speed of information exchange and a high level of reliability, which is maximally acceptable for application in accounting. The difficulties of integrating these systems are related to the high cost of technology and the need for advanced training of specialists.

**Keywords—** Accounting, Data Science, Blockchain, Robotic Process Automation, Artificial Intelligence, Cloud Computing

## I. INTRODUCTION

In today's rapidly evolving digital landscape, organizations around the world are using digital transformation as a catalyst for change and innovation. The emergence of the digital industry is affecting all forms of human activity, including accounting processes. With the help of digital technology, data entry can be much faster, and data can be made accessible to interested users. Based on this, it can be concluded that the formation of accounting analysis is much more effective mainly due to self-improvement [1–3].

At the same time, the digitization of accounting has an impact on many constituent parts of the parts:

- Methods of collecting, storing and exchanging information with interested users;
- Digital technologies in accounting;
- Ways of reconciling information, i.e. the nature of accounting or accounting method [4, 5].

Digital transformation is a fundamental shift in the way organizations conduct administrative and financial activities. It involves the integration and activation of various digital technologies, such as automated information systems, Internet communications, artificial intelligence, blockchain and big data processing, to achieve accuracy, efficiency and speed in organizational operations (Fig. 1).

Financial reporting plays a key role in every organization by providing a comprehensive view of its financial position and performance [6]. Financial performance is a critical indicator of an organization's ability to optimize resources and achieve desired outcomes.

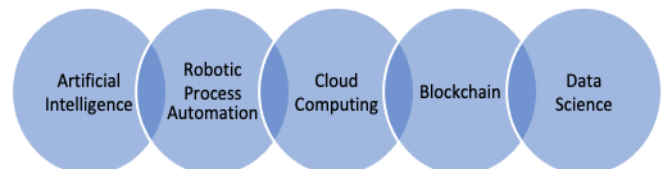


Fig. 1. New trends in accounting technology

This paper analyzes the application of various digital technologies in accounting, evaluating their advantages and disadvantages.

## II. RESEARCH METHODS

In this study, analysis and synthesis were used as methods, namely analysing various digital technologies such as Robotic Process Automation, Artificial Intelligence, Machine Learning, Blockchain and others and synthesising them in accounting by evaluating their merits and demerits.

## III. RESEARCH RESULTS

Robotic Process Automation (RPA) is a technology based on the use of multiple digital robots (bots) that solve routine business tasks and mimic some of the user's actions. RPA is a simple programming tool that does not require a large amount of computer memory and automates manual rule-based tasks faster and cheaper than other technological tools [7, 8]. Typical tasks of the finance department can be summarized into two groups:

1) Receipt and processing of primary documents and preparation of periodic reports (month, quarter, year).

2) Preparation of operational reports, thanks to which top managers can make timely decisions.

These tasks are handled by two related departments: accounting and management. Therefore, the same information may be duplicated. With the help of robotization of accounting processes, unnecessary work can be avoided. Tasks that can be delegated to robotic automation of RPA processes (Fig. 2):

- compiling and separating the accounting statements of multiple legal entities;
- presenting monthly variation and unusual deviations;
- filling out a dashboard, sending an email with access links;
- automatic completion of letters to business partners;
- tracking of deviations from expected norms during inventory counts;
- reflection of transactions within the company in consolidated reporting.

Entering primary documents into the system today is the subject of manual labor. To automate this process requires the introduction of OCR tools (ABBYY, Yandex, Dbrain, Soika) [9].

The difficulties of RPA integration are related to the fact that creating your own solution is long, expensive, will require significant employee resources and support, and using a third-party solution also requires modifications to the existing system.

Robots are a great solution for increasing the efficiency of the accountant's work and saving considerable time. The production of operational reports for management accounting tasks can also be automated. RPA transforms accounting functions from traditional to advisory accounting, deals with email collection of financial data from spreadsheets, deals with invoice routing via PDF files, distributes internal financial reports to business units, etc. However, interoperability with other internal systems is far from guaranteed, not to mention the flexibility of the solutions provided [7, 8].

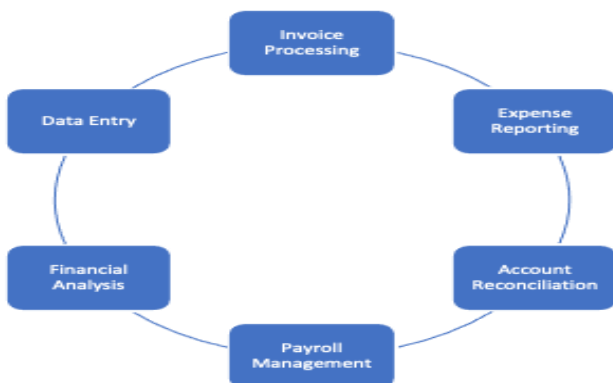


Fig. 2. Examples of RPA in accounting

In general, the results of RPA implementation are as follows:

- Increased speed of work of the accounting department.
- Improvement of quality and accuracy of work, reduction of the "human factor".
- Increased security by protecting access with privacy settings.
- Improved quality of customer service.
- Reduction of routine tasks up to 95%.
- Increase in labor productivity by 50%.
- Reduction of economic costs up to 45 %.

As research [10–12] shows RPA technology allows to automate routine and quite standardized tasks; the technology is in an acceptable state and scalable, quite reliable and stable.

Robotic Process Automation (RPA) and Artificial Intelligence (AI) are two related concepts that have a great impact on accounting and auditing practices. RPA is process-oriented and AI manages data [13]. The peculiarity of robots is that their standardized processes are technically limited. They are unable to process unstructured data. AI and RPA can add value to each other. AI encompasses a number of interrelated technologies including data mining, machine learning, speech recognition, image recognition, and semantic analysis [13].

AI and RPA together mean that large amounts of data can be turned into actionable insights, allowing entire processes to be fully automated. Investing in intelligent automation is very expensive, so it is important to consider which combinations of RPA and AI are most suitable for a company.

The benefits of AI in accounting are summarized in Fig. 3.



Fig. 3. Advantages of AI in accounting

Intelligent data analysis uses a combination of statistical, machine learning and artificial intelligence techniques. The need to use this combination is due to the ever-increasing volume of data, 90% of which is unstructured data.

The central part of artificial intelligence is machine learning (ML). ML systems learn knowledge from data and

decisions, thus learning to classify and make simple decisions [14].

Part of artificial intelligence are NLP (natural language processing), NLU (natural language understanding) and NLG (natural language generation). NLP allows machines to understand and process human language. It analyzes large amounts of text and speech data, identifies patterns and generates intelligent responses. The NLP principle is used in applications such as voice-activated GPS, speech-to-text dictation, Siri, digital assistants, and translators. It performs real-time summarization of big data and text.

NLU is fixed on interpreting the meaning of natural language to better understand its context using syntactic and semantic analysis. It aims at recognizing intentions, objects, sentiment analysis. NLU is applied in automated customer service systems, intelligent virtual assistants, search engines, business chatbots.

NLG focuses on creating natural language from structured data. Its main goal is to create answers in human language and convert the data into speech format.

Currently, only 5% of companies consider AI and 15% in - RPA as an option for them.

By the Internet of Things (IoT) is meant a multilevel system including sensors, microchips, controllers, etc., installed in the nodes and units of an industrial object, means of data transmission, as well as analytical tools for interpreting the received information to make management decisions of current and strategic nature.

Some IoT users cite as its advantage the ability to quickly provide information and respond to process failures, automate decision-making, and reduce operational costs.

Other users argue that IoT technologies have the potential to revolutionize accounting information systems through the emergence of new data sources. Thus, through more data (physical measurements, location, monetary values, etc.), IoT will contribute significantly to improving the completeness, timeliness, and quality of accounting information [15, 16].

To summarize, the Internet of Things (IoT) is a technology based on bridging the gap between the physical and digital worlds through the flow of information to enhance the integration of an organization's activities and operations by providing identification, sensing, and information processing capabilities to communicate with other devices and services through the Internet of Things (IoT).

Big data is a relatively recent phenomenon resulting from the development of a technological environment in which almost everything can be digitally recorded and measured, and a huge number of events can be monitored simultaneously.

This requires special tools and software for storing, managing and analyzing data [17]. Big Data is a set of approaches, tools and methods for processing structured and unstructured data of huge volumes and significant variety to produce human-perceivable results.

The advent of Big Data with its intelligent tools such as smartphones, space sensors, GPS and others will have an impact on the changing role of accountants and auditors. For

management accountants, the amount of data available is increasing significantly, with a fairly large percentage of data residing in cloud storage.

Cloud accounting is a set of accounting programs that reside on a remote server and are designed to operate over the Internet. Data is stored and processed in the so-called cloud, which is, from the client's point of view, one large virtual server.

Studies have shown [18] that accounting and management accounting systems are least likely to favor cloud systems, which is probably due to data security and the decision to store part of the data internally. However, from a technical point of view, cloud service providers can usually offer a higher level of security than most small businesses can certainly afford.

Cloud technology has many advantages. You don't have to purchase licensed accounting and business management software, you can simply rent it. There is no need to install and configure, as well as administer and update the programs, because the configuration and maintenance are handled by special firms. There is also no need to purchase powerful PCs and servers, payment is made only for storage space, there are no restrictions on the number of users. This technology can be used while being in different parts of the world [19].

Cloud technology also has its own disadvantages. The security and storage of company's data depends on the provider. The material liability of the service provider for unauthorized access and distribution of data is insignificant. Insurance of such risks is not provided for today. The speed and reliability of such programs depends on the quality of the Internet and its provider.

Through the use of big data tools, unwanted activities such as accounting fraud and fraud (2017, Zabiholla & Wang) can be detected very quickly and can provide the results of big data analytics in accounting when dealing with changing and sometimes conflicting data and embedded in the financial reporting process, as well as the ability to use big data to predict and further improve the quality of financial statements.

Thus, the challenge for qualified accounting professionals will be to move beyond today's role - beyond the traditional budgeting, cost formation, and control functions. It will be a movement from management accounting to strategic management accounting. Less emphasis on the past, on numbers, on accounting standards and accounting "best practices"; more emphasis on gathering relevant information for future decisions, on using a "control panel" through which various financial and non-financial data can be selected in such a way as to generate a coherent set of information components suitable for decision-making on their basis.

In an increasingly competitive market, blockchain is a good tool for the development of the sector. Blockchain is a distributed database that contains information about all transactions conducted by participants in the system. The information is stored as a chain of blocks. Each block contains a certain number of transactions. Blockchain is an unbreakable digital record of actions. The reliability of this system allows it to be used to improve the efficiency of

monetary transactions and information exchange among individuals, corporations, and even the public sector.

Initially, blockchain was used for cryptocurrency transactions. But later in the process of development on the basis of blockchain and its modifications, other, not related to crypto-tools forms of effective use of this technology appeared. The introduction of blockchain increases the speed of exchange, improves the quality, reliability and availability of services. The scope of application of blockchain has expanded significantly. It can also be applied in the banking sector, financial services, payment services, etc. [20-22].

When it comes to the use of blockchain technology in accounting and auditing, blockchain is a distributed (decentralized) and open digital ledger that allows the transfer of ownership (not copies) from one party to another in real time without the need for an intermediary (third party). It achieves a high degree of security and encryption of the transfer process to counter fraud and manipulation attempts. Those recorded on it cannot be changed retroactively.

Thanks to blockchain technology, the process of creating records and capturing the time of their creation will affect accounting in such a way that all events will forever remain preserved and unchanged. Documents cannot be changed during their life cycles. Business processes spanning multiple departments or even companies will be recorded and fully

traceable. In addition, smart contracts can result in invoices being paid automatically upon confirmation of receipt of goods. Of course, the active implementation of blockchain technology in accounting processes raises the issue of professional development of accountants.

On the one hand, the accountant's functions at the level of primary documentation are reduced due to active automation of accounting, and accountants will spend much less time entering data into the information base and registering facts of economic life. But, on the other hand, accountants will be able to spend more time preparing professional judgments, interpreting the economic content of transactions and correctly reflecting transactions in accounting and reporting. For this purpose, it is necessary to improve their qualifications and restructure their way of thinking. An accountant should become not just an accountant, but also a professional analyst, who understands the economic processes of his/her organization [21, 22].

Technological changes may also affect the software. Many modern accounting systems use "cloud storages" for accumulation of financial information, which are centralized in some data centers.

Analyzing the use of digital technologies in accounting allows us to compare them according to the main indicators (Table 1).

TABLE I. COMPARATIVE ANALYSIS OF DIGITAL TECHNOLOGIES IN ACCOUNTING

Digital technology	Speed	Financial / energy costs	Security	Level of development/use of technology
RPA	High speed of information processing and transmission	High payback and low implementation cost	Dependent on equipment	Dependent on equipment Insufficient technology for accounting - update failures occur
AI	Quick decision making	Low costs	No human error	Little use in accounting
Blockchain	Ability to work in real time or close to it	High material and energy costs	High degree of safety, reliability	Wide application
BigData	Real-time mode	High cost of maintenance	There are privacy issues	Widely used in accounting
Cloud technologies	High speed	Only rent a cloud accounting service	High level	Advanced technology
Internet of Things	Depends on internet speed and quality of equipment	Significant investment for network installation, maintenance and expansion	Insufficient security	Underutilized technology in accounting - many "weaknesses"

Research [23, 24] on the use of blockchain technology and Big Data in the financial sector in Iraq has shown a positive impact on reducing the cost of data processing, the security of information flow in the company's working environment, the quality of financial and economic data, and the quality of financial information.

#### IV. CONCLUSION

Nowadays, organisations that have begun to adopt digital technologies are benefiting from the automation of specific activities and processes in accounting and auditing. The results of the study showed that applying artificial intelligence in conjunction with Robotic Process Automation in the accounting department can automate its work, reducing routine activities by up to 95 %. Despite the greater speed of information processing with cloud-based systems, accounting departments are less likely to utilise them due to data security.

Blockchain digital technology has a high speed of information exchange and a high level of reliability, which is

most suitable for accounting applications. Some types of digital technologies such as big data, blockchain are quite active in the financial sector in Iraq. Some technologies are less so, as this requires significant investment in their development. The implementation of these technologies also requires upgrading the skills of professionals.

The results of the benchmarking of digital methods for accounting showed the prioritized technologies - blockchain, big data and cloud technologies - and revealed the "weaknesses" for accounting of technologies such as Internet of Things, RPA and AI.

By applying digital tools in their work, accountants will be able to spend more time preparing professional judgements, interpreting the economic content of transactions and correctly reflecting transactions in accounting and reporting. To achieve this, it is necessary to upgrade their qualifications and restructure their way of thinking. An accountant should become not just an

accountant, but a professional analyst who understands the economic processes of their organisation.

#### REFERENCES

- [1] S. Kraus, P. Jones, N. Kailer, A. Weinmann, N. Chaparro-Banegas, and N. Roig-Tierno, "Digital transformation: an overview of the current state of the art of research," *Sage Open*, vol. 11(3), September 2021, doi.org/10.1177/21582440211047576.
- [2] R. Martínez-Peláez, A. Ochoa-Brust, S. Rivera, V.G. Félix, R. Ostos, H. Brito, R.A. Félix, L.J. Mena, "Role of Digital Transformation for Achieving Sustainability: Mediated Role of Stakeholders", *Key Capabilities, and Technology. Sustainability*, vol. 15(14), 2023, doi.org/10.3390/su151411221.
- [3] A. Hund, Heinz-Theo Wagner, D. Beimbom, T. Weitzel, "Digital innovation: Review and novel perspective", *The Journal of Strategic Information Systems*, vol. 30, is. 4, 2021, 101695, doi.org/10.1016/j.jsis.2021.101695.
- [4] O. Rozhnova, "Harmonization of Accounting, Auditing and Analysis in a Digital Economy", *Accounting. Analysis. Auditing*, vol. 5. pp.16-23, 2018, doi: 10.26794/2408-9303-2018-5-3-16-23.
- [5] D.M. Coman, C.A. Ionescu, A. Duică, M.D. Coman, M.C. Uzlau, S.G. Stanescu, V. State, "Digitization of Accounting: The Premise of the Paradigm Shift of Role of the Professional Accountant", *Applied Sciences*, vol. 12(7), pp. 3359, 2022, doi:10.3390/app12073359.
- [6] Ahannaya Chinedu, "Financial Reporting and Organization Performance in Nigeria", *Innovation*, vol. 3, pp. 26-32, 2022, doi: 10.11648/j.innov.20220301.15.
- [7] McKay Dwayne, "Robotic process automation", *IJRAR*, vol. 10, is. 2, pp. 338-343, June 2023.
- [8] Ch. Flechsig, F. Anslinger, R. Lasch, "Robotic Process Automation in purchasing and supply management: A multiple case study on potentials, barriers, and implementation", *Journal of Purchasing and Supply Management*, vol. 28, is. 1, 2022, pp. 100718, doi: 10.1016/j.pursup.2021.100718.
- [9] How we tackled document recognition issues for autonomus and automatic payments using OCR and NER. *Teleofis*. <https://habr.com/ru/articles/673142/>
- [10] M. Hafner, P. P. Raimondi, B. Bonometti, "The Energy Sector and Energy Geopolitics in the MENA Region at a Crossroad", Springer, p. 216, 2023.
- [11] Ch. Zhang, H. Issa, A. Rozario and J.S. Søgaard, "Robotic Process Automation (RPA) Implementation Case Studies in Accounting: A Beginning to End Perspective", *Accounting Horizons*, vol. 37, 2022, doi: 10.2308/HORIZONS-2021-084.
- [12] Ramona Lacurezeanu, Tiron Tudor, Adriana, Bresfelean Vasile, "Robotic Process Automation in Audit and Accounting", *Audit Financiar*, vol. 18, 2020, doi: 10.20869/AUDITF/2020/160/024.
- [13] M. Gotthardt, D.Koivulaakso, O. Paksoy, C. Saramo, M. Martikainen, O. Lehner, "Current State and Challenges in the Implementation of Smart Robotic Process Automation in Accounting and Auditing", *ACRN Journal of Finance and Risk Perspectives*, vol. 9, pp. 90-102, 2020.
- [14] N. Tyagi, B. Bhushan, "Demystifying the Role of Natural Language Processing (NLP) in Smart City Applications: Background, Motivation, Recent Advances, and Future Research Directions", *Wireless personal communications*, vol. 130(2), pp. 857–908, 2023, doi: 10.1007/s11277-023-10312-8.
- [15] Pallavi Sethi, Smruti R. Sarangi, "Internet of Things: Architectures, Protocols, and Applications", *Journal of Electrical and Computer Engineering*, vol. 2017, Article ID 9324035, 25 p., 2017, doi: 10.1155/2017/9324035.
- [16] M. Mansour, A. Gamal, A.I. Ahmed, L.A. Said, A. Elbaz, N. Herencsar, A. Soltan, "Internet of Things: A Comprehensive Overview on Protocols, Architectures, Technologies, Simulation Tools, and Future Directions", *Energies*, vol. 16(8), pp. 3465, 2023, doi: 10.3390/en16083465.
- [17] OECD (2021), *Artificial Intelligence, Machine Learning and Big Data in Finance: Opportunities, Challenges, and Implications for Policy Makers*, <https://www.oecd.org/finance/artificial-intelligence-machine-learning-big-data-in-finance.htm>.
- [18] M. Quinn, E. Strauss, G. Kristandl, "The effects of cloud technology on management accounting", *Management and Financial Accounting Report*, vol. 10(6), 2015.
- [19] Sh. Qi, "Fintech and the Digital Transformation of Financial", *BCP Business & Management*, vol. 41, pp. 289-294, 2023, doi: 10.54691/bcpbm.v41i.4445
- [20] L. Albshaiyer, S. Almarri, M.M. Hafizur Rahman, "A Review of Blockchain's Role in E-Commerce Transactions: Open Challenges, and Future Research Directions", *Computers*, vol. 13(1):27, 2024, doi: 10.3390/computers13010027.
- [21] V.S. Elagin, V.I. Fedorovskikh, A.V. Spirikina, *Development and analysis of a blockchain system based on javascript*, T-Comm, No. 2, 2021.
- [22] S. Al-Farsi, M.M. Rathore, S. Bakiras, "Security of Blockchain-Based Supply Chain Management Systems: Challenges and Opportunities", *Applied Sciences*, vol. 11(12):5585, 2021, doi: 10.3390/app11125585.
- [23] B.K.A. Alkafaji, M.L. Dashtbayaz, M. Salehi, "The Impact of Blockchain on the Quality of Accounting Information: An Iraqi Case Study", *Risks*, vol. 11(3):58, 2023, doi: 10.3390/risks11030058.
- [24] Ahmed Waheed Hadi, Ph Dr. Amal Abdel Hussein Kahit, Prof. Dr. Fadel Abdul Abbas Al-Abedi, "The role of big data analytics in measuring and analyzing the quality costs of economic units", *Journal of Accounting & Financial Studies*, vol. 18, No. 63, 2023.

# Модель компьютерного зрения для анализа загруженности регистратуры КДЦ «НМИЦ им. В.А. Алмазова»

Г. В. Орлов, А. Н. Калиниченко

Санкт-Петербургский государственный электротехнический университет  
«ЛЭТИ» им. В.И. Ульянова (Ленина)

sifon8998@mail.ru

**Аннотация.** В данной работе проведен сравнительный анализ эффективности работы дообученной модели YOLO с её изначальной версией для подсчета людей в регистратуре «клинико-диагностического НМИЦ им. В.А. Алмазова». Результаты показали, что обучение модели на собственном наборе данных значительно повысило точность определения количества людей, что важно для оценки загруженности регистратуры. Полученные результаты могут быть использованы для оптимизации работы клинико-диагностического центра и улучшения обслуживания пациентов.

**Ключевые слова:** компьютерное зрение; нейронные сети; машинное обучение; Python; YOLO; здравоохранение

## I. ВВЕДЕНИЕ

Регистратура является важной и неотъемлемой частью медицинского учреждения, обеспечивающей организацию приема пациентов. Однако, несмотря на непрерывный прогресс в области медицины, многие регистратуры сталкиваются с проблемами неоптимальной работы, такими как долгое время ожидания, неэффективное распределение нагрузки и недостаточная удовлетворенность пациентов.

Поставленная заказчиком, клинико-диагностическим центром «НМИЦ им. В.А. Алмазова», задача звучит следующим образом: разработать систему, способную проводить анализ загруженности регистратуры медицинского центра с целью детального изучения проблематики перегрузки. Таким образом, целью настоящего проекта является создание системы для клинико-диагностического центра «НМИЦ им. В.А. Алмазова», направленной на повышение эффективности работы регистратуры. Система предоставит инструменты для прогнозирования загруженности регистратуры и оценки эффективности процессов, что позитивно отразится на обслуживании пациентов, сократит время ожидания и повысит общую производительность регистратуры. Также в рамках проведенных исследований была проведена дополнительная настройка модели YOLOv8, это алгоритм машинного обучения для обнаружения объектов на изображениях и видео. Он является развитием более ранних версий алгоритма YOLO и представляет собой сеть глубокого обучения, способную определять и классифицировать объекты в реальном времени на собственном наборе данных. Результаты этой работы показали значительное улучшение точности

модели в сравнении с исходной версией. Данный подход позволил значительно повысить качество распознавания объектов и повысить общую производительность модели в предсказании загруженности регистратуры. Полученные результаты говорят об эффективной адаптации модели под конкретные условия и специфику исследуемых данных, что открывает новые перспективы для применения YOLOv8 в решении задач прогнозирования и оптимизации работы регистратур в медицинских учреждениях.

Важность данного исследования обусловлена несколькими факторами. Во-первых, оптимизация работы регистратуры напрямую влияет на уровень удовлетворенности пациентов и на посещаемость ими медицинского учреждения. Недостаточная эффективность регистратуры может вызвать проблемы, такие как долгие очереди и нерациональное использование ресурсов, что в конечном итоге может повлечь за собой снижение качества обслуживания и потерю доверия со стороны пациентов.

Во-вторых, разработка информационной системы для оптимизации работы регистратуры является перспективным направлением, которое поможет современным медицинским учреждениям увеличить эффективность деятельности и внедрить инновационные подходы к управлению потоком пациентов.

Таким образом, данное исследование обладает практической значимостью и предлагает новые подходы для улучшения работы регистратуры в медицинских учреждениях.

## II. МОДЕЛЬ КОМПЬЮТЕРНОГО ЗРЕНИЯ YOLO

Модели компьютерного зрения – это мощные инструменты и алгоритмы, разработанные для обработки изображений и видео с целью распознавания и классификации объектов, а также обнаружения различных структур и паттернов. Они играют ключевую роль в сфере искусственного интеллекта и компьютерного зрения, обеспечивая автоматизацию процессов анализа и интерпретации визуальных данных. [1]

В этом контексте модель YOLO занимает особое место, представляя собой эффективный подход к обнаружению объектов в реальном времени. Её последняя версия, YOLOv8, объединяет в себе высокую



скорость обработки с достаточно высокой точностью обнаружения, что делает эту модель популярным выбором для различных задач компьютерного зрения.

Погружаясь в исследования в области компьютерного зрения, важно понимать особенности и потенциал моделей, таких как YOLOv8, чтобы использовать их на практике с максимальной эффективностью.

YOLO (You Only Look Once) – это один из популярных алгоритмов для решения задачи обнаружения объектов в реальном времени. YOLOv8, как последняя версия из семейства YOLO, является результатом эволюции предыдущих моделей, например, YOLOv4 и YOLOv5. Эта модель сочетает в себе высокую скорость обработки видео и изображений с достаточно высокой точностью обнаружения объектов. YOLO определяет объекты на изображении путем предсказания ограничивающих рамок (далее bounding boxes), классов объектов и оценки достоверности этих предсказаний. [2] Процесс определения объектов представлен на рис. 1 и включает в себя несколько шагов:

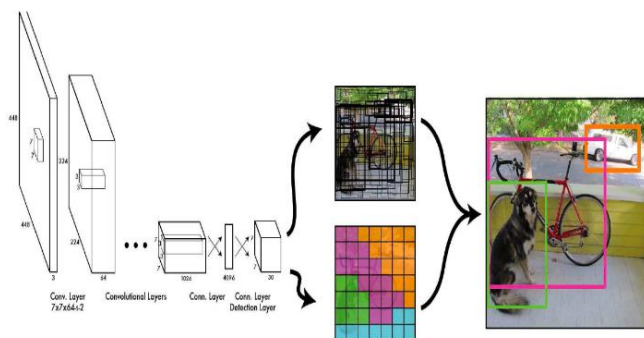


Рис. 1. Алгоритм работы модели типа YOLO [3]

1. Разделение изображения на сетку: Исходное изображение разделяется на сетку ячеек, обычно размером, например, 13x13 или 19x19. Каждая ячейка будет отвечать за обнаружение объектов в своей области.

2. Предсказание bounding boxes и классов: для каждой ячейки сетки модель YOLO предсказывает несколько bounding boxes со своими координатами, размерами, классами объектов и оценками достоверности предсказаний.

3. Фильтрация предсказаний: для улучшения точности и устранения лишних или неправильных предсказаний применяется фильтрация на основе порога уверенности (confidence threshold) и метода Non-Maximum Suppression (NMS).

4. Объединение результатов: после фильтрации и применения NMS каждая оставшаяся bounding box соответствует одному объекту, а несколько bounding boxes могут соответствовать одному объекту.

5. Интерпретация результатов: для каждого определенного объекта модель возвращает информацию о его классе, координатах на изображении и оценки достоверности классификации.

При всех своих преимуществах, модель YOLO обладает некоторыми недостатками. Несмотря на свою высокую скорость работы, модель YOLOv8 сталкивается с проблемами точности при обнаружении объектов на сложных датасетах с большим количеством классов, что может снизить общую надежность предсказаний, а также может потребовать значительных вычислительных ресурсов в зависимости от конфигурации и параметров обучения, что подчеркивает значимость оптимизации ресурсов и параметров для эффективного обучения. В следующем разделе будет рассмотрено обучение предобученной модели на собственном наборе данных для минимизации недостатков модели YOLO.

### III. ОБУЧЕНИЕ И МЕТРИКИ

Для начала стоит рассмотреть и определить, какую модель YOLOv8 мы хотим дообучить. В нашем распоряжении представлены следующие модели: YOLOv8n, YOLOv8s, YOLOv8-m, YOLOv8-l, и YOLOv8-x. Каждая из этих моделей имеет свои особенности и применение. Модель YOLOv8n обладает небольшим размером и высокой скоростью обнаружения, что делает её подходящей для ресурсоемких приложений. YOLOv8s представляет собой стандартную модель, характеризующуюся сбалансированными характеристиками скорости и точности. YOLOv8-m, YOLOv8-l и YOLOv8-x отличаются увеличенным размером и сложностью, что позволяет им обрабатывать более сложные задачи с высокой точностью, но с более низкой скоростью. [3]

Для дообучения была выбрана версия nano. YOLOv8n применяется в рассматриваемом контексте ввиду своих очевидных преимуществ, определяющих выбор именно этой модели. Применение нормализации на каждом слое обеспечивает стабильность процесса обучения и улучшенную сходимость модели, что сказывается на ее обобщающей способности. В результате усовершенствованных процессов обучения возможно добиться повышения общей точности модели YOLOv8n, что важно в контексте задач обнаружения объектов на изображениях. Дополнительно, способность модели успешно обрабатывать разнообразные типы данных говорит в пользу ее универсальности и адаптивности к различным задачам. Усовершенствованная работа с аномальными данными, обеспечиваемая нормализацией, является критически важным аспектом, который может повысить эффективность модели в реальных условиях применения. Благодаря оптимизациям и эффективной реализации, YOLOv8n также отличается хорошим уровнем производительности в обработке изображений. Модель способна достигать высоких значений скорости (кадров в секунду, FPS) на поддерживаемых устройствах. Это обеспечивает быструю и эффективную обработку больших объемов данных, что особенно важно в реальном времени или при работе с большими наборами изображений. В совокупности данные особенности делают YOLOv8n привлекательным выбором для задач обнаружения объектов, а ее преимущества в области стабильности обучения, точности и гибкости придадут исследованию дополнительную значимость и методологическую

обоснованность. При тестировании модели был получен следующий результат, представленный в табл. 1.

ТАБЛИЦА 1.

Class	'person'
Images	250
Instance	700
Box (P	0.58
R	0.37
mAP50	0.403
mAP50-95)	0.192

Полученные результаты на проверочном наборе представлены следующим образом:

Для всех классов обнаружены 700 экземпляров объектов на 250 изображениях, что соответствует показателям точности  $P=0.58$ , это означает, что около 58% всех выявленных моделью объектов действительно являются искомыми, и полноты  $R=0.37$ , следовательно, модели удастся обнаружить около 37% всех существующих объектов данного класса на изображениях. Средние значения метрик mAP50 и mAP50-95 составляют 0.403 и 0.192 соответственно. [2]

Максимальное значение средней точности обнаружения объектов (далее mAP) при пороге перекрытия (далее IoU) 0.50 составляет 40.3, эта метрика показывает насколько хорошо модель обнаруживает объекты с достаточным перекрытием с истинной разметкой, а mAP50 при диапазоне равном 0.50–0.95 IoU = 19.2, это позволяет оценить производительность модели при различных уровнях требуемого перекрытия с объектами. [2]

Хотя эти значения могут показаться невысокими, учитывая использование модели Nano, они демонстрируют приемлемые результаты.

Далее необходимо было подготовить собственный набор данных на основе записей с видеокamer регистратуры клинично-диагностического центра «НМИЦ им. В.А. Алмазова». Пример изображения представлен на рисунке 2.

Для создания собственного датасета, предназначенного для обучения модели YOLO, необходимо провести процесс разметки данных с соблюдением определенных стандартов. Разметка данных является важным этапом обучения моделей компьютерного зрения и представляет собой процесс присвоения аннотаций объектам на изображениях, указывая их координаты и классы.

Первым этапом является подготовка изображений, которые будут использоваться для создания датасета. У заказчика были получены порядка сотни записей с камер видеонаблюдения (пример на рис.2), которых насчитывается порядка сотни. Затем необходимо определить классы объектов, которые будут размечаться на изображениях, и составить файлы аннотаций, указывающие для каждого изображения информацию о прямоугольных областях, где располагаются объекты, и их классах. Значительно упрощает работу специализированные программы, такие как Supervisely или MakeSense.



Рис. 2. Запись с камер наблюдения в КДЦ «НМИЦ им. В.А. Алмазова»

Для модели YOLO в файле аннотаций каждая строка должна содержать следующую информацию: номер класса объекта, нормированные координаты центра объекта по осям X и Y, нормированные размеры объекта по ширине и высоте. Формат аннотации может варьироваться, но должен быть совместим с требованиями YOLO. [4]

После создания файлов аннотаций необходимо связать их с соответствующими изображениями для каждого объекта в датасете. Важно также уделить внимание качеству разметки, точности указания границ объектов и соответствию классов, чтобы обучение модели проходило эффективно.

Таким образом, процесс разметки данных для собственного набора данных для обучения модели YOLO включает в себя подготовку изображений, определение классов объектов, создание файлов аннотаций с необходимой информацией о расположении объектов, а также обеспечение качественной разметки данных для последующего обучения модели.

Также стоит отметить, что процесс разметки данных для создания собственного датасета, предназначенного для обучения модели YOLO, является длительным и малоприятным. Так как требуется вручную размечать каждый объект на изображениях, указывая его класс и координаты, этот этап работы может занимать много времени и требовать значительных усилий. Однако точность и качество разметки играют ключевую роль в обучении модели, поэтому необходимо уделить этому этапу должное внимание и терпение, несмотря на его длительность и малоприятные аспекты.

Пример размеченного изображения для обучения модели представлен на рис. 3.

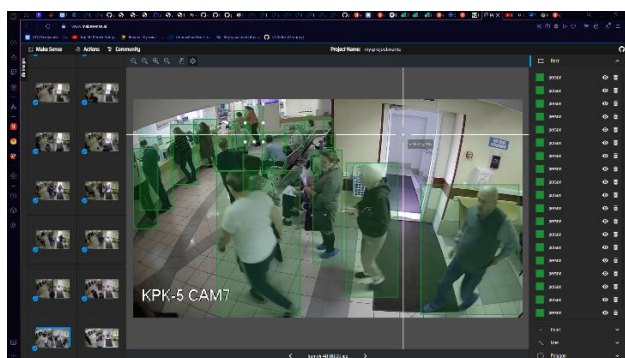


Рис. 3. Пример размеченного изображения для обучения модели

Теперь переходим к этапу самого обучения модели. Вы можете запустить следующую команду через терминал, чтобы инициировать процесс обучения с использованием командной строки YOLO:

```
yolo task=detect mode=train model=yolov8n.pt
imgsz=1280 data=custom.yaml epochs=50 batch=8
name=yolov8custom
```

Нижеприведенный код демонстрирует ту же конфигурацию обучения, но с применением Python API [3–5]:

```
from ultralytics import YOLO
# Загрузка модели.
model = YOLO('yolov8n.pt')
# Обучение.
results = model.train(
    data='custom.yaml',
    imgsz=1280,
    epochs=50,
    batch=8,
    name='yolov8custom'
)
```

В зависимости от характеристик оборудования, процесс обучения может занимать несколько часов. В данном исследовании на обучение уходило порядка 2–3 часов. На выходе получается результат, отображенный в виде карты и графиков потерь, представленных на рис. 4.

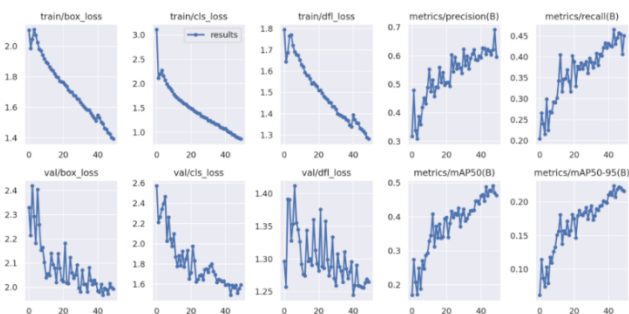


Рис. 4. Карта и графики потерь

При использовании новой модели YOLOv8 удалось достичь значения средней площади под кривой точности и полноты (mAP) равного почти 50 при пороге пересечения/объединения (IoU) 0.50. Для проверки фактического значения предлагается выполнить команду оценки:

```
yolo task=detect mode=val model=runs/detect/
yolov8custom/weights/best.pt name=yolov8custom_eval
data=custom.yaml imgsz=1280
```

Результаты оценки представлены в табл. 2:

ТАБЛИЦА II.

Class	'person'
Images	250
Instance	700
Box (P)	0.705
R	0.568
mAP50	0.589
mAP50-95)	0.324

По представленным метрикам результаты оценки модели выглядят следующим образом:

- точность (Precision) для обнаружения объектов составляет примерно 70.5 %;
- полнота (Recall) равна примерно 56.8 %;
- средняя точность обнаружения объектов при IoU 0.50 (mAP50) составляет примерно 58.9%, что является показателем чуть выше среднего качества модели в выявлении объектов на изображениях;
- средняя точность обнаружения объектов с учетом интервала IoU от 0.50 до 0.95 (mAP50-95) равна примерно 22.4 %, что указывает на уровень качества модели при более широком диапазоне оценки обнаруженных объектов.

Исходя из этих результатов, можно сделать вывод о том, что модель превышает средний уровень качества в обнаружении объектов на изображениях. Для улучшения ее производительности можно проанализировать и внести коррективы в процесс обучения, модифицировать архитектуру модели или провести дополнительную настройку гиперпараметров. [6]

#### IV. ЗАКЛЮЧЕНИЕ

В результате проведенного исследования, посвященного обучению модели на собственном наборе данных и дообучению YOLOv8n, удалось достичь уровня метрик, который не только приблизился к показателям более крупной версии YOLOv8s, но в некоторых случаях оказался даже лучше. Это свидетельствует о высоком потенциале и эффективности дообучения модели на специализированных данных.

Особенно значимо, что при этом удалось сохранить хороший уровень скорости работы модели. Этот факт подчеркивает возможность успешной реализации системы анализа загруженности регистратуры клинико-диагностического центра "НМИЦ им. В.А.Алмазова" с использованием полученной модели.

Таким образом, результаты исследования говорят о перспективности использования такого подхода в области компьютерного зрения и машинного обучения для решения конкретных задач, связанных с анализом и управлением в клинической практике.

#### СПИСОК ЛИТЕРАТУРЫ

- [1] Goodfellow I., Bengio Y., Courville A. Deep Learning. Moscow: DMK Press, 2019.
- [2] O'Rourke B., Irish T., Kaneelingham K. Computer Vision Technologies: Principles and Applications. Saint Petersburg: Piter, 2018.
- [3] Redmon J., Divvala S., Girshick R., Farhadi A. "You Only Look Once: Unified, Real-Time Object Detection." Magazine: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016.
- [4] Kohlin S.V., Vinogradova P.Yu., Kazakova G.S. Application of Machine Learning in Medicine: Research Overview. Informatics and Control Systems, 2020, No. 2 (62), pp. 179-189.
- [5] Bochkovskiy A., Wang C-Y., Liao H-Y.M. "Yolov4: Optimal Speed and Accuracy of Object Detection." Magazine: IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020.
- [6] Li N., Zhang H., Li Z. Application of Computer Vision and Machine Learning in Medical Research. Information Technology, 2019, Vol. 25, No. 2, pp. 93-103.

# Разработка нейросетевого модуля автозаполнения протокола осмотра для модульной медицинской информационной системы

И. М. Киряков<sup>1</sup>, С. А. Молодяков<sup>2</sup>

Санкт-Петербургский политехнический университет Петра Великого  
Высшая школа программной инженерии

<sup>1</sup> vania.ru.ru@mail.ru, <sup>2</sup> molodyakov\_sa@spbstu.ru

**Аннотация.** Рассматриваются вопросы разработки медицинской информационной системы, особенностью которой является ее модульное построение. Представлен модуль автозаполнения протокола осмотра пациента. В основе модуля лежит использование нейросетевых алгоритмов распознавания. Для перевода речи из аудио в текст используется нейросеть SpeechKit, а для извлечения из текста ответов на вопросы YandexGPT. Программное обеспечение разработано с использованием технологии микросервисной архитектуры. Проведен анализ механизмов обмена данными между модулями, выделена технология удаленного вызова процедур gRPC. Приводятся результаты применения разработанного модуля.

**Ключевые слова:** информационная система, микросервисная архитектура, обмен данными, автозаполнение протокола, API

## I. ВВЕДЕНИЕ

Современная медицина отличается тем, что в ней широко применяются информационные системы. Такими системами являются системы определения заболеваний, поиска медицинской информации, анализа большого массива данных из истории болезней и другие. Часто в медицинских информационных системах (МИС) применяются методы искусственного интеллекта (ИИ), в частности методы, связанные с использованием нейронных сетей [1]. Особенностью современных информационных технологий является непрерывное совершенствование методов и алгоритмов, используемых в МИС. Так за последние два года сделан существенный шаг в развитии средств и методов определения заболеваний легких по рентгеновским снимкам. В результате возникает необходимость добавления и/или замены в МИС отдельных алгоритмов на другие. Такая замена возможна, если каждый алгоритм упаковать в отдельный модуль, а МИС будет представлять собой модульную систему.

В настоящее время наиболее современный подход для построения модульной системы связан с использованием микросервисной архитектуры для построения программного обеспечения. В этом случае приложение представляет собой совокупность слабосвязанных сервисов. Разработка каждого сервиса

может вестись независимо от других, возможна замена отдельных сервисов [2, 3].

Настоящая работа является первым этапом создания модульной МИС по автоматизации рабочих процессов, совершаемых врачами в медицинских организациях. Программное обеспечение (ПО) разрабатывается с использованием микросервисной архитектуры. Разработаны первые модули, связанные с задачей автозаполнения протокола осмотра при помощи инструментов на основе ИИ. Рассмотрены технологии обмена данными, которые используются в МИС.

## II. КОНЦЕПЦИЯ РАЗРАБАТЫВАЕМОЙ СИСТЕМЫ

В мире разработки ПО преимущественно используются две популярные архитектуры разработки приложений: монолитная и микросервисная. Выбор конкретного архитектурного решения зависит от многих факторов, в частности, назначения разрабатываемой системы. В рамках современных проектов, нуждающихся в легкой масштабируемости, в последнее время предпочтение отдается микросервисам [2]. Поэтому основной разработкой МИС является микросервисная архитектура. Каждый сервис работает в своем процессе и взаимодействует с другими сервисами при помощи API-шлюза т. е. советующей технологии обмена данными. На рис. 1 представлена схема системы.

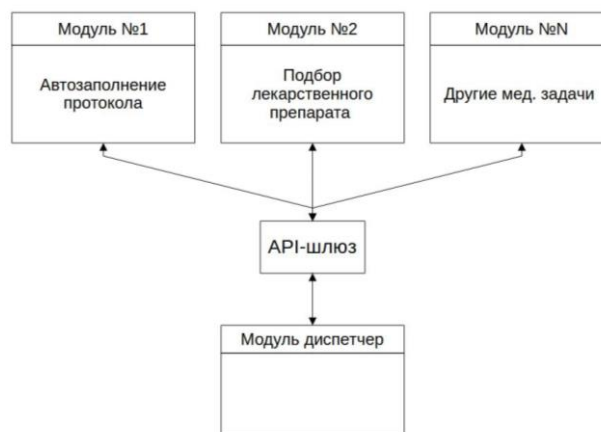


Рис. 1. Архитектура системы

Это схема медицинского модульного комплекса, которая, предназначена для автоматизации процесса медицинского обслуживания пациента.

Данный модульный комплекс будет иметь возможность динамического приседания и отсоединения нужных модулей, которые могут быть использованы как последовательно, так и на прямую по отдельности. Каждый модуль представляет самостоятельную единицу, которая в полной мере выполняет свою задачу и имеет возможность передавать нужные данные по API «следующему модулю». В качестве «следующего модуля» может выступать модуль Диспетчер, который будет поддерживать интерфейс пользователя, организовывать вызов других модулей. В данной работе излагается разработка конкретного модуля «Автозаполнение протокола осмотра», таким образом был разработан модуль «Подбор лекарственного препарата» [9], также может быть преобразована в модуль разработанная нами программа обнаружения COVID-19 [4].

### III. АНАЛИЗ ТЕХНОЛОГИЙ ОБМЕНА ДАННЫМИ ПРИ МИКРОСЕРВИСНОЙ АРХИТЕКТУРЕ

При проектировании приложения на микросервисной архитектуре одним из основных вопросов является взаимодействие между ее модулями. Рассмотрим известные решения и определим наилучшее решение для обмена данными между приложением-клиентом и приложением-сервером. Обмен данными между приложениями относится к элементу программного интерфейса API (Application Programming Interface), который может быть вызван или выполнен на различных уровнях абстракции в системе.

Для рассмотрения были выбраны следующие используемые технологии [5, 6]:

- протокол простого доступа к объектам (SOAP);
- передача репрезентативного состояния (REST);
- удаленный вызов процедур (gRPC).

Акцентируем внимание на особенности применения технологии обмена данными. Весь процесс обмена данными сводится к принципу изображенный на рис. 2.



Рис. 2. Принцип обмена данными

В настоящее время разработка приложений введётся на объектно-ориентированном языке программирования, например, Java, C# и др. Соответственно данные, над которыми осуществляется манипуляция в алгоритмах приложения, находятся в состоянии объекта, представляющий класс на языке программирования ООП. Для передачи объекта осуществляется его преобразование в соответствующий формат данных, т. е. сериализация, а при получении осуществляется его

обратный процесс, т. е. десериализация. Самыми популярными форматами сериализации данных на сегодняшний день являются JSON (JavaScript Object Notation), XML (eXtensible Markup Language) и Protocol Buffers (Protobuf). Для передачи данных в SOAP используется XML, для REST нет фиксированного формата передачи сообщений, а gRPC используется Protocol Buffers.

Реализация механизма обмена данными сопровождается некоторыми трудностями, которые необходимо преодолеть разработчикам и программистам, чтобы обеспечить согласованность между участниками обмена данными.

Для оценки и сравнительного анализа быстродействия для каждой технологии обмена данными были реализованы программы, как клиентская, так и серверная часть. При реализации тестовых программ был соблюден принцип обмена данными, описанный выше (рис. 2). По окончании обмена данными учитывалось время сериализации и десериализации данных. В качестве объекта, используемого для обмена, выступала структура данных, состоящая из 3-х полей и заполняемая строковыми данными различной длины в диапазоне от 1000 до 1000000 символов. Тестовым стендом служит ПК с характеристиками: ЦП – Intel Core i5-9600KF (6 ядер по 4.6 ГГц), ОЗУ –DDR4 16 Гб.

На рис. 3 представлен график зависимости времени обмена от количества передаваемых данных. Можно сделать вывод, что REST является легковесной и быстрой технологией, но проигрывает по времени при передаче больших данных по причине того, что у него отсутствует встроенной системы сериализации и десериализации данных. SOAP и gRPC обладают встроенной системой сериализации и десериализации данных, но SOAP уступает gRPC по причине того, что gRPC использует более быстрый механизм конвертации данных на основе Protobuf.

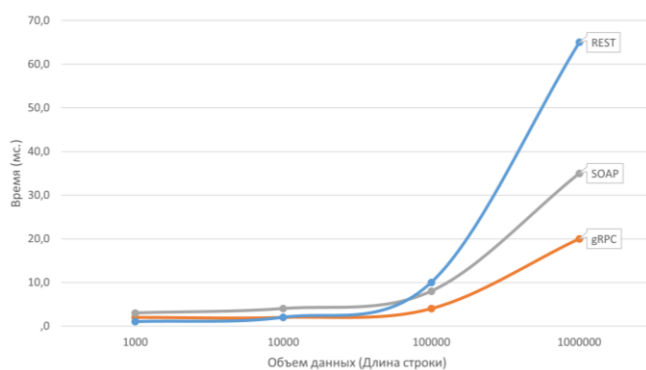


Рис. 3. График времени обмена данными

Рассмотрев представленные технологии обмена, можно сделать вывод, что наилучшим вариантом при реализации приложения является gRPC. Он будет лучшим вариантом, особенно когда реализация алгоритма обмена данными в приложении является основной задачей, а для реализации её требуется надёжное и нетрудоёмкое в применении технология.

#### IV. ОПИСАНИЕ РАЗРАБОТАННОГО МОДУЛЯ

Рассмотрим разработанный модуль. Он предназначен для автоматического заполнения протокола (АЗП) с использованием методов искусственного интеллекта (ИИ). В основе модуля лежит использование нейросетевых алгоритмов распознавания сведений.

Прием к врачу начинается с получения анамнеза, т. е. врач осуществляет процесс получения от пациента совокупности сведений, путем расспроса и медицинского обследования. Полученную информацию от пациента врачу необходимо вносить в электронную медицинскую карту (ЭМК).

Данный модуль предоставит возможность автоматизировать данный процесс, требуя от врача минимальные усилия.

При реализации данного модуля были изучены существующие методы решения аналогичной задачи. Был рассмотрен подход, реализованный в системе «АрхиМед» [10]. «АрхиМед» предлагает метод заполнения протекла с использованием системы распознавания речи. Голосовое заполнение осуществляется по отдельным полям протокола, требуя от врача голосовой команды переключения по полям.

В реализуемом модуле также применяется система распознавания речи, но с применением вспомогательного ИИ обеспечивающего единого заполнения протокола, не требующая от врача отдельного заполнения каждого поля. Система сама распознаёт из записанной речи, в какие поля протокола занести соответствующий текст.

На рис. 4 представлена архитектура разрабатываемой системы. Показана концепция подключения модуля к МИС и использования его для расширения функциональных возможностей системы.

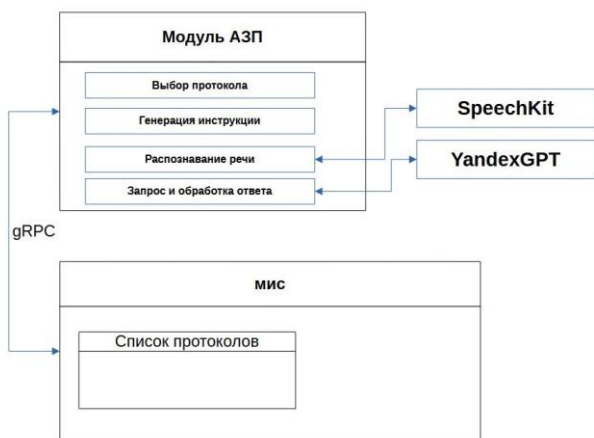


Рис. 4. Архитектура системы с использованием модуля АЗП

Модуль представляет из себя микросервис, реализованный по стандартам, позволяющий подключать его к МИС. Архитектуру системы на рис. 4 можно представить как шаблон проектирования, необходимый для расширения функционала МИС, в основе которого лежат следующие составляющие:

- МИС (с диспетчером);

- Модуль АЗП;
- API-шлюз.

Модуль АЗП, представляет из себя микросервис, выполняющий конкретную медицинскую задачу с применением ИИ для МИС. В данной работе микросервис реализуется на платформе ASP.NET Core. Работа модуля АЗП устроена следующим образом:

Шаг 1. Врач осуществляет выбора протокола, который необходимо автоматически заполнить. Пример приведён на рис. 5.

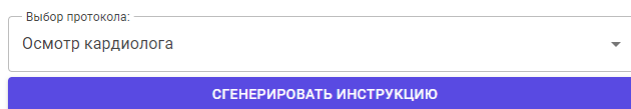


Рис. 5. Выбор протокола

Шаг 2. На основании данного протокола осуществляется генерация инструкций по нажатии кнопки «Сгенерировать инструкцию» для нейронной сети, которые она должна применить к тексту запроса. Пример приведён на рис. 6.

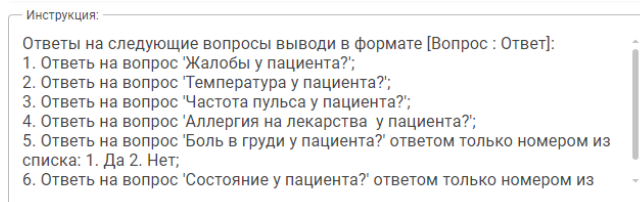


Рис. 6. Генерация инструкции по протоколу

Шаг 3. Врач нажимает кнопку «Начать запись». Осуществляется аудио запись осмотра и перевод речи из аудио в текст с помощью нейросети SpeechKit [7]. Пример приведён на рис. 7.

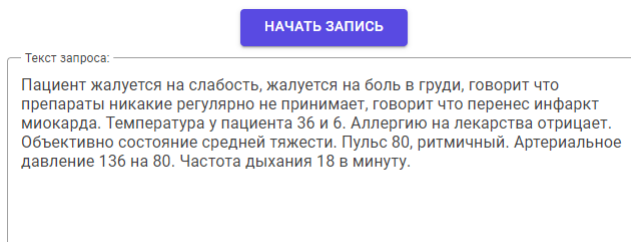


Рис. 7. Формирование записи

Шаг 4. К полученному тексту применяется инструкция с использованием нейросети YandexGPT для извлечения из текста ответов на вопросы, требуемые для заполнения протокола [8]. Пример обработки приведён на рис. 8.

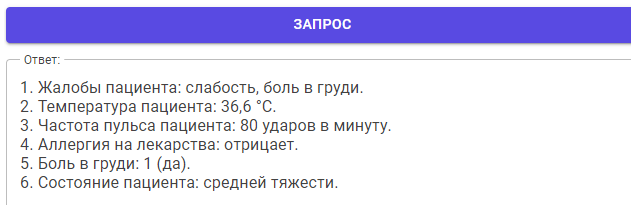


Рис. 8. Результат обработки

API-шлюз, в данной системе это средство универсального подключения модуля к МИС. Он обеспечивает коммуникацию между клиентом (МИС) и сервисом (Модуль), в основе которого лежит фреймворк gRPC, который использует протокол RPC (Remote Procedure Call) для обмена сообщениями между клиентом и сервером.

gRPC использует подход contract-first для построения API. Это некоторый контракт, которому сервер и клиент должны следовать. Этот контракт описывается с помощью файлов с расширением .proto.

Proto-файлы выполняют две основные задачи:

- описание сервиса gRPC;
- описание формата сообщений, которыми обмениваются клиент и сервер.

Для описания сервиса gRPC и сообщений proto-файле используется специальный синтаксис proto.

Подобных proto-файлов МИС может подключить несколько. Таким образом, МИС может внедрить в свою систему несколько необходимых модулей, расширить функционал и всего за пару шагов обновить МИС.

## V. ЗАКЛЮЧЕНИЕ

В данной работе была рассмотрена модульная МИС, связанная с использованием микросервисной архитектуры. Модульность построения предполагает постепенное наращивание возможностей информационных систем путем добавления новых модулей и замены старых модулей новыми более совершенными. Причем в качестве модулей можно использовать программы, в которых используются методы искусственного интеллекта.

Проведен анализ технологий обмена данными SOAP, REST, gRPC. После проведенного тестирования, выделена технология удаленного вызова процедур gRPC, которая используется в системе. Представлена архитектурная схема системы с модулем автоматического заполнения протокола. Рассмотрен механизм работы с модулем. Определен шаблон и механизм взаимодействия с модулями.

Будущие направления разработок и исследований связаны с разработкой интерфейса-диспетчера для работы с врачами, расширением возможностей модуля автоматического заполнения протокола, созданием других модулей.

## СПИСОК ЛИТЕРАТУРЫ

- [1] Мазеев В.П., Рязанова С.В. Основные направления развития искусственного интеллекта в медицине. // Научное обозрение. Медицинские науки. 2020. №5. С.33-40. doi: 10.17513/srms.1141.
- [2] Порфильева А.Э., Шайхутдинов Р.Ф., Нуриева Г.А. Эффективная разработка приложений при микросервисной архитектуре // Электронные библиотеки. 2018. Т. 21. № 3-4. С. 357-368. EDN XVASHB.
- [3] Chris Richardson (2017). Pattern: Microservice Architecture. URL: <http://microservices.io/patterns/microservices.html>
- [4] Думаев Р.И., Киряков И.М., Молодяков С.А. Особенности предобработки и сегментации изображений в задаче обнаружения COVID-19 по рентгеновским снимкам // Современная наука: актуальные проблемы теории и практики. Серия: Естественные и Технические Науки. 2022. №09. С. 88-95. doi: 10.37882/2223-2966.2022.09.08
- [5] Различия REST и SOAP [Электронный ресурс]. – URL: <https://habr.com/ru/post/483204/> (19.01.2023)
- [6] Авельцов Д.О. Применение протокола сериализации структурированных данных Protobuf в микросервисной архитектуре // Проблемы автоматизации и управления. 2022. № 3(45). С. 185-196. EDN SJWCHJ  
Обзор технологий Yandex SpeechKit. URL: <https://cloud.yandex.ru/ru/docs/speechkit/overview>  
YandexGPT API. URL: <https://cloud.yandex.ru/ru/services/yandexgpt>
- [7] Киряков И.М., Молодяков С.А. Метод построения модульной медицинской информационной системы на примере разработки программы подбора лекарственного препарата. Современная наука: актуальные проблемы теории и практики. Серия: Естественные и технические науки. 2023. № 4. С. 71-77.
- [8] «АрхиМед». Экосистема ИТ-решений и сервисов для медицины будущего. URL: [https://archimedical.ru/wp-content/uploads/2022/01/archimed\\_brochure\\_10.2022.pdf](https://archimedical.ru/wp-content/uploads/2022/01/archimed_brochure_10.2022.pdf)

# Выявление рентгенологической стадии остеоартроза коленного сустава с применением сверточных нейронных сетей

Н. А. Цыганюк, Е. Ю. Белова

Санкт-Петербургский государственный электротехнический университет  
«ЛЭТИ» им. В.И. Ульянова (Ленина)

natsygaryuk@stud.etu.ru, eyshukeylo@gmail.com

**Аннотация.** По данным Всемирной организации здравоохранения, остеоартроз диагностируется у 7,6 % населения во всем мире. При данном заболевании чаще всего поражается коленный сустав в связи с его анатомическими и функциональными особенностями. Выявление остеоартроза путём изучения рентгенологических снимков связано с определенными временными затратами, а качество диагностики напрямую зависит от квалификации врача. В настоящее время большое количество исследований посвящено теме анализа предварительно обработанных рентгенологических изображений с использованием методов глубокого обучения. Такой подход позволяет сократить сроки постановки диагноза, снизить риск врачебных ошибок, повысить эффективность лечения пациентов. В данном исследовании в результате обучения сверточных нейронных сетей разных архитектур, в частности MobileNet, на основе открытого набора данных The Osteoarthritis Initiative (OAI) точность выявления стадий остеоартроза составляет 77,4 %.

**Ключевые слова:** остеоартроз; рентгенограмма; OAI; нейронные сети; MobileNet; ResNet; глубокое обучение

## I. ВВЕДЕНИЕ

Остеоартроз коленного сустава является одной из наиболее часто встречающихся форм артроза, которая характеризуется дегенерацией суставного хряща и появлением костных наростов. На сегодняшний день выявлены несколько основных факторов риска появления остеоартроза: высокая физическая активность, пожилой возраст, ожирение. Также риску развития больше подвержены женщины, нежели мужчины [1, 2].

Показатель заболеваемости остеоартрозом по всему миру растет с каждым годом. Так, например, при анализе открытой базы данных «Глобальное бремя болезней» выяснилось, что общее число заболеваний на 2019 год составляло 364.5 миллиона, в то время как по состоянию, например, на 2010 год было на 80 миллионов меньше – 283. Около половины всех выявленных случаев только за 2019 год приходится на людей старше 55 лет, среди которых около 65 % женщины [3].

В большинстве стран мира в качестве шкалы, которая определяет стадию остеоартроза коленного сустава, принята классификация Kellgren-Lawrence (KL). Согласно ей, заболевание имеет 5 стадий [4]:

- 0 – отсутствие радиографических признаков остеоартроза;

- 1 – возможное сужение суставного пространства и образование остеофитов;
- 2 – точное образование остеофитов с возможным сужением суставного пространства;
- 3 – множественные остеофиты, определенное сужение суставного пространства, склероз и возможная костная деформация;
- 4 – большие остеофиты, выраженное сужение суставного пространства, выраженный склероз и определенная костная деформация.

Для диагностики остеоартроза коленного сустава используются различные методы, самым распространенным из которых является рентгенография. Она позволяет достаточно быстро и без больших финансовых затрат получить снимок коленного сустава, по которому врач впоследствии определяет наличие и стадию остеоартроза. По сравнению с другими методами выявления остеоартроза рентгенография обладает несомненными достоинствами. Однако полученные в ходе исследования снимки коленного сустава часто имеют явные помехи (шумы), затрудняющие процесс диагностики заболевания [5]. Безусловно, уровень квалификации врача также оказывает влияние на корректность определения стадии остеоартроза. От правильной постановки диагноза зависит последующая тактика лечения пациента.

С целью снижения риска появления ошибки в ходе ручного анализа рентгенографических снимков, применяются сверточные нейронные сети. Их использование имеет сразу несколько преимуществ: сокращение финансовых затрат, уменьшение времени на выявление заболевания и определение его стадии, улучшение качества диагностики.

Помимо сверточных нейронных сетей, прослеживается тенденция на применение предварительной цифровой обработки рентгенограмм, которая позволяет уменьшить шумы на изображении. Данный подход упрощает процесс обучения моделей и, в конечном счете, улучшает точность выявления стадии остеоартроза коленного сустава. Особенность текущего исследования заключается в применении цифрового контрастирования к изображению для лучшего выделения значимых участков.



Целью данного исследования является выявление рентгенологической стадии остеоартроза коленного сустава с применением сверточных нейронных сетей.

В процессе исследования решаются следующие задачи:

- предварительная обработка набора данных;
- выбор архитектур сверточных нейронных сетей;
- обучение выбранных моделей и оценка качества прогнозирования.

## II. МАТЕРИАЛЫ И МЕТОДЫ

В качестве набора данных в проведенном исследовании используется коллекция рентгеновских снимков коленного сустава, собранная в рамках программы The Osteoarthritis Initiative (OAI) [6]. Целью данной программы является выявление признаков возникновения и прогрессирования остеоартроза, а также разработка новых методов его лечения, направленных на замедление развития данного заболевания. Участниками программы OAI являются 4796 пациентов в возрасте от 45 до 79 лет. Набор данных содержит 8260 рентгеновских снимков коленного сустава, которые разделены в соответствии со шкалой KL. Выбор данного набора обусловлен следующими фактами. Во-первых, он находится в открытом доступе. Во-вторых, все рентгенограммы предварительно проверяются группой врачей, что обеспечивает точность определения стадии заболевания.

Язык программирования Python и фреймворк PyTorch, который содержит набор функций и архитектур сетей для упрощения обучения моделей, используются для обучения сверточных нейронных сетей. Данные разделены на тренировочную и тестовую выборки в соотношении 80:20 соответственно. С целью интерпретации процесса принятия решения моделью используется метод Grad-CAM, который генерирует тепловую карту на изображении. Гиперпараметры для настройки проведения обучения представлены в табл. I.

ТАБЛИЦА I. ГИПЕРПАРАМЕТРЫ ОБУЧЕНИЯ

Количество эпох	40
Функция потерь	Кросс-энтропия
Оптимизатор	Adam

В данном исследовании применяется библиотека OpenCV. В качестве методов предварительной обработки рентгеновских снимков выбраны изменение контраста изображения и локально-адаптивная эквализация гистограммы. Также проведено предварительное уменьшение размера изображений до 224×224 пикселей для уменьшения объема занимаемой оперативной памяти при обучении.

В качестве метрики для оценки результатов классификации используется точность.

## III. ОСНОВНЫЕ РЕЗУЛЬТАТЫ

При изменении контраста и локально-адаптивной эквализации гистограммы получены более четкие изображения и устранены артефакты, которые могут негативно повлиять на классификацию. Примеры

результатов обработки изображений представлены на рис. 1 и 2.



Рис. 1. Результаты обработки снимка для класса 3 KL



Рис. 2. Результаты обработки снимка для класса 0 KL

В данном исследовании использовано трансферное обучение, в котором в качестве базовых архитектур сверточных нейронных сетей выбраны MobileNetV3, ResNet50 и DenseNet121, предобученные на наборе данных ImageNet. Затем проведена тонкая настройка каждой полученной модели. Результаты прогнозирования рентгенологической стадии остеоартроза коленного сустава для тестовой выборки представлены в табл. II.

ТАБЛИЦА II. РЕЗУЛЬТАТЫ ОБУЧЕНИЯ

Архитектура СНС	Точность
MobileNetV3	77.4%
ResNet50	65%
DenseNet121	71.2%

Как можно заметить, самое высокое значение точности получено для модели, основанной на архитектуре MobileNet, и составляет 77.4%. Стоит также отметить, что у данной архитектуры самый низкий уровень требований к вычислительным ресурсам, и процесс ее обучения занимает меньшее количество времени по сравнению с другими выбранными архитектурами.

На основе результатов обучения модели MobileNet построены карты активации с применением метода Grad-CAM. Пример полученного изображения приведен на рис. 3.



Рис. 3. Карта активации для снимка класса 0KL

#### IV. ДИСКУССИЯ

С каждым годом количество исследований, посвященных классификации стадий остеоартроза, растет, что указывает на актуальность проблемы диагностики данного заболевания. Одной из лучших работ в плане достижения высокого показателя точности определения рентгенологической стадии остеоартроза является исследование Тюльпина А. и Саараккалы С. [7]. В нем использована обработка изображений фильтром Гаусса, а обучение проведено с помощью ансамбля сетей SE-Resnet-50 и SE-Resnet-50-32x4d. Точность составила в среднем 82 %. Это одно из самых высоких значений, полученных с помощью сверточных нейронных сетей. Однако на стадиях 0–2 классификации KL наблюдается уменьшение точности до 67 %. Также при обучении ансамбля нейронных сетей предъявляются высокие требования к вычислительным ресурсам.

В работе Суреша С. и др. используется перспективный подход с предварительной обработкой изображения. Он заключается в обрезке снимка до области околосуставной щели, что позволяет удалить лишнюю информацию с изображения и снизить требования к объему памяти вычислительной машины при обучении [8]. При классификации стадий 2–4 шкалы KL точность составляет 88.2 %. Однако, среднее значение точности падает до 67.2 % для всех классов.

Основываясь на ранее рассмотренных исследованиях, можно сделать вывод, что использование предварительной обработки рентгеновских снимков позволяет добиваться лучших показателей точности при классификации стадий остеоартроза.

#### V. ЗАКЛЮЧЕНИЕ

В результате проведенного исследования получены модели сверточных нейронных сетей, которые позволяют определить рентгенологическую стадию остеоартроза коленного сустава на основе рентгеновских снимков коленного сустава, собранных в рамках программы ОАИ. Их точность по сравнению с точностью моделей, полученных в приведенных исследованиях, несколько ниже для некоторых классов KL но в то же время имеет хорошие показатели при совокупном выявлении всех классов KL градации.

Дальнейшие улучшения результатов проведенной работы возможны за счёт внедрения более глубокой системы обработки рентгеновских снимков и изменения параметров обучения. Увеличение значения показателя точности возможно путём добавления автоматического выделения участка изображения с суставной щелью и применения различных гиперпараметров обучения, например, ordinalloss в качестве функции потерь.

#### СПИСОК ЛИТЕРАТУРЫ

- [1] Кульчицкая Д.Б., Юрова О.В., Севрюгина О.А. Применение магнитно-резонансной томографии в оценке эффективности медицинской реабилитации пациентов с остеоартрозом коленного сустава: обзорная статья. // Вестник восстановительной медицины. 2023. № 22(2). С. 96-101.
- [2] Бадочкин В.В. Остеоартроз коленного сустава: клиника, диагностика, лечение. // Современная ревматология. 2013;(3):70–5.
- [3] Global Burden of Disease. URL: <https://vizhub.healthdata.org/gbd-results>. (дата обращения: 12.03.2024)
- [4] Kellgren JH., Lawrence J.S. Radiological assessment of osteoarthritis. // Ann Rheum Dis. 1957. Т. 16. №. 4. С. 494-502.
- [5] Сигал З.М., Сурнина О.В., Брындин В.В. Экспресс-диагностика ревматоидного артрита и остеоартроза // Вестник Мордовского университета. 2017. Т. 27, № 3. С. 330-344.
- [6] The Osteoarthritis Initiative (OAI). URL: <https://nda.nih.gov/oai>
- [7] Tiulpin A., Saarakkala S. Automatic Grading of Individual Knee Osteoarthritis Features in Plain Radiographs Using Deep Convolutional Neural Networks. // Diagnostics. 2020. Т. 10. №. 11. С. 932.
- [8] Suresha, S., Kidziński, L., Halilaj, E., Gold, G. E., & Delp, S. L. Automated staging of knee osteoarthritis severity using deep neural networks. // Osteoarthritis and Cartilage. 2018. Т. 26. С. S441.

# Эффективность выявления лейкемии: сравнительное исследование EfficientNetB3 и EfficientNetB5

Асил Альшурайхи<sup>1</sup>, Хусам Хасан Бу Исса<sup>2</sup>,  
Анагим Ибрагим<sup>3</sup>, Agbonrhienrhien Osazee Osa<sup>4</sup>

Санкт-Петербургский государственный электротехнический университет  
«ЛЭТИ» им. В.И. Ульянова (Ленина)

<sup>1</sup>Russiaaal\_shuraikhi@stud.etu.ru, <sup>2</sup>houssam.bouissa@gmail.com,  
<sup>3</sup>Russiaaibrakhim@etu.ru, <sup>4</sup>OsazeeOsa@gmail.com

**Аннотация.** Лейкемия, сложное гематологическое злокачественное новообразование, требует точного и своевременного выявления для эффективного лечения. В этом исследовании исследуется производительность двух моделей глубокого обучения, EfficientNetB3 и EfficientNetB5, в контексте выявления лейкемии. Используя набор данных, включающий различные изображения лейкозных клеток, мы провели всесторонний сравнительный анализ для оценки эффективности этих моделей. Наше исследование углубляется в сложность чувствительности, специфичности и общей точности обнаружения путем тщательного рассмотрения экспериментов и оценки показателей производительности. Полученные результаты подчеркивают различия в возможностях обнаружения между EfficientNetB3 и EfficientNetB5, проливая свет на их сильные и слабые стороны. Выводы, собранные в ходе этого исследования, вносят вклад в развитие области выявления лейкемии, а также предлагают ценные рекомендации для практикующих врачей и исследователей, стремящихся использовать методы глубокого обучения для улучшения диагностики и лечения заболеваний.

**Ключевые слова:** лейкемия, машинное обучение, EfficientNetB3, EfficientNetB5, точность

## I. ВВЕДЕНИЕ

Лейкоз — это сложный рак крови, который возникает в костном мозге и приводит к чрезмерной выработке аномальных лейкоцитов [1]. Эти клетки вытесняют здоровые клетки крови, что приводит к таким симптомам, как анемия, инфекция и нарушения свертываемости крови [2]. Заболевание классифицируется на различные типы, при этом острый лимфобластный лейкоз (ОЛЛ) и острый миелоидный лейкоз (ОМЛ) являются наиболее распространенными у детей и взрослых соответственно [3].

Диагностика лейкемии традиционно основывалась на микроскопическом исследовании мазков крови опытными гематологами. Однако этот метод может быть субъективным и подверженным человеческим ошибкам [4]. С появлением методов глубокого обучения произошло значительное повышение точности и эффективности диагностики лейкозов [5]. Особенно эффективными в этом отношении оказались такие

модели, как EfficientNetB3 и EfficientNetB5, входящие в семейство сверточных нейронных сетей (СНС) EfficientNet [2].

EfficientNetB3 продемонстрировал высокую точность классификации ALL за счет использования метода комплексного масштабирования, который равномерно масштабирует глубину, ширину и разрешение сети. С другой стороны, EfficientNetB5 при своем большем размере и сложности показал еще лучшую производительность за счет увеличения емкости и использования трансферного обучения для тонкой настройки модели на конкретных наборах данных [6].

Эти модели были обучены на обширных наборах данных изображений мазков крови, что позволило им изучить сложные закономерности, связанные с различными подтипами лейкемии [7]. В результате они могут оказать быструю и надежную диагностическую помощь, что имеет решающее значение для начала своевременного и адекватного лечения [10].

## II. ОБЗОР ЛИТЕРАТУРЫ

Хандекар и его коллеги представили систему автоматизации на основе искусственного интеллекта, направленную на улучшение обнаружения взрывных камер. Их подход включал в себя интеграцию методов обнаружения объектов в систему, которая анализировала микроскопические изображения мазков крови для прогнозирования лейкозных клеток. Используя четвертую версию алгоритма You Only Look Once (YOLO), они классифицировали клетки как здоровые (НЕМ) или бластные клетки (ALL), рассматривая это как задачу бинарной классификации. Система прошла тщательное обучение и тестирование с использованием изображений из наборов данных ALL\_IDB1 и C\_NMC. Результаты оказались многообещающими, алгоритм продемонстрировал высокую точность в идентификации лейкозных клеток. Предлагаемая система может служить ценным инструментом для предварительного скрининга, помогая в ранней идентификации лейкемии по микроскопическим изображениям мазка крови [8].

В исследовании, проведенном Касани и его коллегами, была предложена агрегированная модель глубокого обучения (DL) для классификации лейкозных

В-лимфоцитов. Чтобы решить проблему небольшого набора данных, авторы использовали методы аугментации данных для увеличения количества обучающих выборок. Кроме того, было реализовано трансферное обучение для ускорения обучения и повышения производительности предлагаемой сети, обеспечивая ее надежность и точность. Метод достиг заметных успехов в диагностике лейкозных В-лимфоцитов, превзойдя отдельные сети за счет объединения признаков из различных моделей глубокого обучения [9].

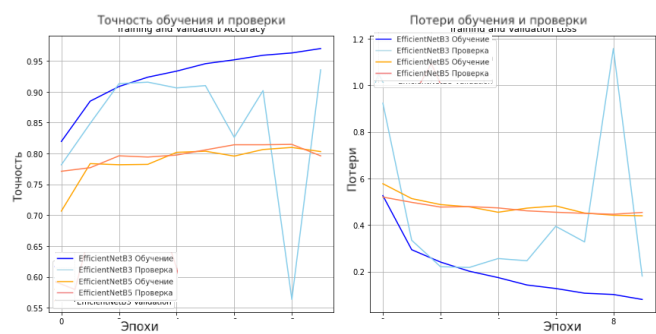


Рис. 1. График для алгоритмов

Альмадор и его коллеги использовали наивные алгоритмы Байеса (NB), К-ближайшего соседа (KNN), случайного леса (RF) и метода опорных векторов (SVM) для разработки ансамблевой автоматизированной стратегии прогнозирования. В их исследовании использовался набор данных о лейкемии C-NMC, полученный от Kaggle, где клетки были разделены на здоровые и раковые группы. При анализе результаты показали, что SVM продемонстрировал более высокую производительность по сравнению с другими алгоритмами, достигнув точности 90 % [10].

### III. ЭКСПЕРИМЕНТЫ И РЕЗУЛЬТАТЫ

Наши модели, показанные на рис. 2, были протестированы на наборе данных из 10000+ изображений 118 пациентов с лейкемией из The Cancer Imaging Archive (TCIA) [7], достигнув точности 93,56 % с моделью EfficientB3 и 79,62 % с моделью EfficientB5. Изображение содержит двухлинейные графики, каждый из которых отображает сравнение производительности двух архитектур нейронных сетей, EfficientNetB3 и EfficientNetB5, в течение ряда эпох во время обучения.

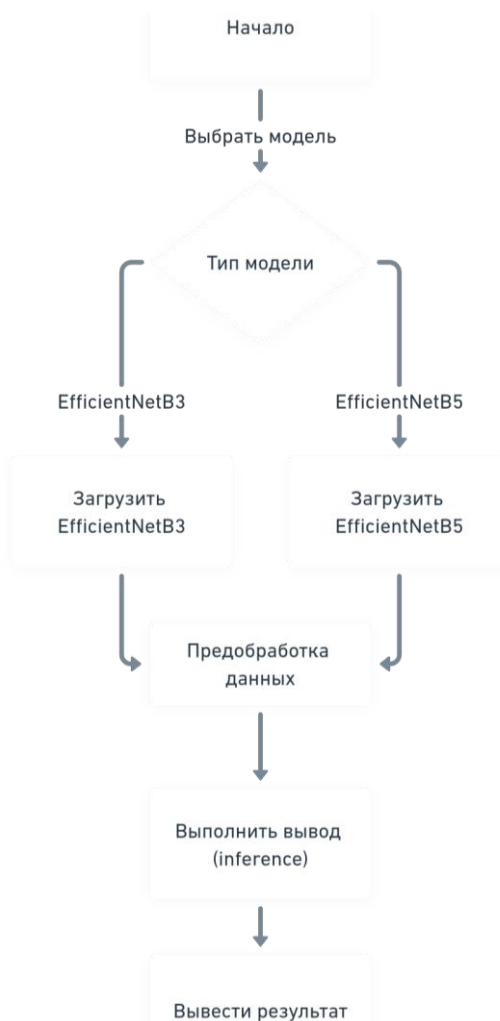


Рис. 2.

На левом графике показаны четыре линии, представляющие точность обучения и проверки как для EfficientNetB3, так и для EfficientNetB5. Ось X представляет количество эпох, а ось Y – точность в диапазоне от 0 до 1. Линии точности обучения EfficientNetB3 и EfficientNetB5 относительно гладкие, в то время как линии точности валидации показывают некоторые флуктуации, в частности, значительное падение точности валидации EfficientNetB3 около эпохи 7. Кроме того, на графике справа показаны четыре линии, представляющие потери в обучении и проверке как для EfficientNetB3, так и для EfficientNetB5. Ось X – это снова количество эпох, а ось Y – убыток в диапазоне от 0 до чуть более 1. Потери в линиях для EfficientNetB3 и EfficientNetB5 со временем уменьшаются, что ожидается в процессе обучения. Тем не менее, линии потерь при валидации показывают волатильность, при этом потери при валидации EfficientNetB3 резко возрастают в эпоху 7, отражая падение точности валидации, наблюдаемое на левом графике. Оба графика полезны для диагностики поведения моделей во время обучения, таких как переобучение, недостаточное обучение или другие проблемы, которые могут потребовать корректировки процесса обучения. Кроме того, волатильность, наблюдаемая в потерях валидации как для EfficientNetB3, так и для EfficientNetB5,

особенно значительный всплеск, отмеченный в EfficientNetB3, может быть связана с различными факторами. Во-первых, основной проблемой является переобучение, возникающее, когда модель чрезмерно изучает обучающие данные, включая их шум и выбросы, что приводит к плохим результатам работы с новыми, невидимыми данными. Заметный всплеск потерь валидации в эпоху 7 для EfficientNetB3 указывает на потенциальное переобучение на этом этапе. Кроме того, неподходящая скорость обучения может вызвать волатильность потерь валидации; слишком высокая скорость может привести к тому, что модель выйдет за пределы оптимальной точки во время оптимизации, что приведет к колебаниям потерь, в то время как слишком низкая скорость может привести к тому, что модель попадет в ловушку локальных минимумов, что также повлияет на потери при проверке. Кроме того, выбор размера партии во время обучения может повлиять на стабильность потерь; меньшие пакеты могут вносить больше шума в обновления градиента, вызывая флуктуации, в то время как большие пакеты могут обеспечить более стабильные оценки, но потенциально сглаживают полезные сигналы. Тщательное применение аугментации данных имеет решающее значение для того, чтобы избежать появления шума в процессе обучения, особенно если дополненные данные значительно отличаются от проверочного набора. Кроме того, отсутствие или неправильное использование методов регуляризации может привести к переобучению и последующим колебаниям потерь валидации, что подчеркивает важность регуляризации для предотвращения чрезмерной фиксации модели на обучающих данных. Сложность модели по отношению к задаче также может способствовать волатильности потерь при валидации, при этом слишком сложные модели склонны к появлению шума при обучении на основе обучающих данных, что приводит к плохому обобщению. Наконец, различия в качестве данных и их распределении между наборами данных для обучения и проверки могут привести к неожиданным скачкам потерь при валидации, что подчеркивает необходимость репрезентативных наборов данных для обеспечения стабильности модели. Учет этих факторов влечет за собой точную настройку скорости обучения, корректировку размера пакета, применение соответствующих методов регуляризации, обеспечение качества данных и, возможно, упрощение архитектуры модели для лучшего соответствия сложности задачи.

#### IV. МЕТОДОЛОГИЯ

Методология обучения и оценки моделей EfficientNetB3 и EfficientNetB5 включает в себя несколько ключевых этапов. Первоначально собирается и предварительно обрабатывается подходящий набор данных, включая такие задачи, как изменение размера изображений, нормализация значений пикселей и дополнение данных для повышения разнообразия. После этого выбраны архитектуры EfficientNetB3 и EfficientNetB5 настраиваются для конкретной задачи, при необходимости корректируя такие параметры, как размер входных данных и частота выпадения. Затем модели обучаются с помощью соответствующего оптимизатора и функции потерь на разделенном наборе

данных, состоящем из обучающего, проверочного и тестового наборов. На протяжении всего обучения производительность валидационного набора контролируется для предотвращения переобучения, при этом при необходимости вносятся тонкие корректировки. Затем обученные модели оцениваются на тестовом наборе, чтобы оценить их производительность на основе невидимых данных, а метрики оценки рассчитываются для количественной оценки их эффективности. Сравнительный анализ моделей EfficientNetB3 и EfficientNetB5 проводится для определения того, какая из них лучше справляется с поставленной задачей, с последующей интерпретацией результатов для выявления сильных и слабых сторон, а также областей для улучшения. Если модели соответствуют желаемым критериям, они могут быть рассмотрены для развертывания в реальных приложениях с постоянным мониторингом и обновлениями для поддержания эффективности с течением времени.

#### V. ВЫВОД

Сравнение точности обучения и валидации между EfficientNetB3 и EfficientNetB5 дает ценную информацию об их поведении при обучении и возможностях обобщения. Для EfficientNetB3 точность обучения неуклонно возрастает с течением эпох, что свидетельствует об эффективном обучении на основе обучающих данных. Тем не менее, точность валидации демонстрирует больше флуктуаций, при этом заметное падение наблюдалось около эпохи 7, что совпало со всплеском потерь при валидации. Это несоответствие указывает на потенциальное переобучение, когда модель может запоминать тонкости обучающих данных, а не учиться эффективно обобщать новые, неизвестные данные.

С другой стороны, EfficientNetB5 следует аналогичной тенденции с устойчивым повышением точности обучения, что указывает на надежное обучение на основе обучающих данных. Тем не менее, его точность валидации, по-видимому, более стабильна по сравнению с EfficientNetB3, с меньшим количеством колебаний, наблюдаемых с течением времени. Эта стабильность позволяет предположить, что EfficientNetB5 может лучше обобщать данные проверки, что указывает на его потенциал для повышения производительности на невидимых данных.

Подводя итог, можно сказать, что, несмотря на то, что обе модели демонстрируют повышение точности обучения в течение разных эпох, EfficientNetB5 демонстрирует более стабильную и потенциально более высокую производительность обобщения с точки зрения точности проверки. Это различие особенно примечательно, учитывая значительное снижение точности валидации EfficientNetB3 в эпоху 7, что сигнализирует о потенциальных ограничениях в его способности к эффективному обобщению. Эти результаты подчеркивают важность не только мониторинга точности обучения, но и оценки точности валидации для оценки истинных возможностей обобщения модели.

СПИСОК ЛИТЕРАТУРЫ

- [1] Райна Р., Гондхи Н.К., Чаахат и др. Систематический обзор по выявлению острого лейкоза с использованием методов глубокого обучения. Методы архитектурных вычислений Eng 30, 251–270 (2023). <https://doi.org/10.1007/s11831-022-09796-7>
- [2] Чжоу М., Ву К., Юй Л., Сюй М., Ян Дж., Шен К., Лю Б., Ши Л., Ву С., Донг Б., Ван Х., Юань Дж., Шэнь С. и Чжао Л. (2021) Разработка и оценка системы диагностики лейкемии с использованием глубокого обучения в реальных клинических сценариях. Фронт. Педиатрия. 9:693676. DOI: 10.3389/fped.2021.693676
- [3] Вахидул Хасан Абир, доктор медицины Фахим Уддин, Фариа Рахман Ханам, Тахия Тазин, Мохаммад Монируджаман Хан, Мехеди Масуд, Султан Альджахдали, «[Отозванный] объяснимый ИИ в диагностике и прогнозировании лейкемии с использованием метода трансферного обучения», Вычислительный интеллект и нейронаука, том 2022 г., идентификатор статьи 5140148, 14 страниц, 2022 г. <https://doi.org/10.1155/2022/5140148>
- [4] Тауфик Шоули, Ахмед А. Альшейхи, «Биомедицинская диагностика лейкемии с использованием классификатора глубокого обучения», Вычислительный интеллект и нейронаука, том 2022 г., идентификатор статьи 1568375, 9 страниц, 2022 г. <https://doi.org/10.1155/2022/1568375>
- [5] Сампатхила, Н., Чадага, К., Госвами, Н., Чадага, Р., Пандья, М., Прабху, С., Байри, М. Г., Катта, С. С., Бхат, Д., и Упадья, С. (2022). Настраиваемый классификатор глубокого обучения для выявления острого лимфобластного лейкоза с использованием изображений мазка крови. *Здравоохранение (Базель, Швейцария)*, 10(10), 1812. <https://doi.org/10.3390/healthcare10101812>
- [6] Абд эль-Гани, Самех, Мохаммед Эльмоги и А. А. Абд эль-Азиз. 2023. «Система компьютерной диагностики заболеваний крови с использованием EfficientNet-V3 на основе алгоритма динамического обучения» *Диагностика* 13, No 3: 404. <https://doi.org/10.3390/diagnostics13030404>
- [7] Mourya, S., Kant, S., Kumar, P., Gupta, A., & Gupta, R. (2019). Набор данных ALL Challenge ISBI 2019 (C-NMC 2019) (версия 1) [набор данных]. Архив визуализации рака. <https://doi.org/10.7937/tcia.2019.dc64i46r>
- [8] Кхандекар Р., Шастрий, Джайшанкар С., Фауст О. Сампатхила Н. Автоматизированное обнаружение бластных клеток для диагностики острого лимфобластного лейкоза. *Биомед. Сигнальный процесс. Контроль*. 2021, 68, 102690.
- [9] Касани Х., Вон Парк С., Вон Чан, Дж. Агрегированный метод глубокого обучения для классификации лейкозных В-лимфоцитов. *Диагностика* 2020, 10, 1064.
- [10] Альмадор А., Саттар У., Аль-Хеджайли А., Мохаммад У.Г., Тарик У., Бен Чиха Х. Эффективный подход на основе компьютерного зрения для прогнозирования острого лимфобластного лейкоза. *Фронт. Вычислительная техника. Невроз*. 2022, 16, 1083649.

# Моделирование теплопроводности черного фосфорена методом классической молекулярной динамики с использованием потенциала, полученного на основе глубокого обучения сверточной нейронной сети

Д. В. Завьялов<sup>1</sup>, Д. Н. Жариков<sup>1</sup>, В. И. Конченков<sup>1,2</sup>, Д. В. Шеин<sup>1</sup>

<sup>1</sup>Волгоградский государственный технический университет

<sup>2</sup>Волгоградский государственный социально-педагогический университет

kontchenkov@yandex.ru

**Аннотация.** Представлены результаты моделирования теплопроводности черного фосфорена методом классической молекулярной динамики в пакете LAMMPS с использованием потенциала межатомного взаимодействия, полученного на основе глубокого обучения сверточной нейронной сети, реализованной в пакете SchNetPack. Проведено сравнение результатов, полученных при помощи пакета SchNetPack, с результатами, полученными с помощью пакета DeePMD, основанном на использовании полносвязной нейронной сети.

**Ключевые слова:** черный фосфорен, сверточные нейронные сети, потенциал межатомного взаимодействия, классическая молекулярная динамика, теплопроводность

## I. ИСПОЛЬЗОВАНИЕ НЕЙРОННЫХ СЕТЕЙ РАЗЛИЧНЫХ ТИПОВ ДЛЯ ПРЕДСКАЗАНИЯ ПОТЕНЦИАЛА МЕЖАТОМНОГО ВЗАИМОДЕЙСТВИЯ

Моделирование методом классической молекулярной динамики находит применение в самых разных исследованиях кристаллических и аморфных материалов. В последнее время развивается подход, основанный на использовании нейронных сетей для вычисления потенциала межатомного взаимодействия (силового поля), который затем используется при расчете траекторий частиц в рамках молекулярной динамики. Поскольку обучающая выборка формируется на основе моделирования изучаемых молекул методами квантовой молекулярной динамики (ab initio molecular dynamics, AIMD) на коротких временных интервалах (порядка фемтосекунд), такой подход в принципе позволяет предсказывать силовое поле с точностью, доступной методам теории функционала плотности (density functional theory, DFT). При этом использование

классической молекулярной динамики с подобранным таким образом силовым полем позволяет рассматривать образцы, содержащие тысячи и десятки тысяч атомов, и времена порядка наносекунд.

В настоящее время широкое развитие получило использование полносвязных [1], графовых [2–5], сверточных [6, 7], генеративных [9–10] нейронных сетей в подобных задачах. Столь широкий охват используемых методов глубокого обучения обусловлен, во-первых, испытанием различных типов описания материала, архитектур и методов обучения нейронных сетей, наиболее точно представляющих силовое поле. С другой стороны, использование широкой номенклатуры сетей связано с особенностями строения исследуемых веществ, приводящих к тому, что сети определенного типа удовлетворительно описывают силовое поле одних материалов, но плохо справляются с представлением потенциала взаимодействия в других материалах. Так, в [11] проведено исследование применимости нейронных сетей с прямой связью, реализованных в пакете DeePMD [1], для вычисления потенциала межатомного взаимодействия черного фосфорена, представляющего собой двумерный кристалл, и полифениленсульфида, являющегося полимером. Было установлено, что используемая в [11] нейронная сеть плохо учитывает особенности силового поля полифениленсульфида, что связано со значительно большим количеством конфигураций полимера (по сравнению с кристаллическим веществом), которые должны быть представлены в обучающей выборке.

Собственно, сравнение полученных силовых полей является только первым этапом оценки применимости нейронных сетей для моделирования рассматриваемого материала. Необходимо вычислить какую-то физическую величину, чтобы сравнить ее значение со значениями, получаемыми на эксперименте. В работе [12] выполнен расчет плотности и теплопроводности черного фосфорена при помощи силового поля, построенного нейронной сетью пакета DeePMD.

Работа поддержана грантом РФФИ 23-22-00461 «Исследование тепловых свойств, упорядоченных и неупорядоченных низкоразмерных материалов методом молекулярного моделирования с потенциалами, полученными при помощи глубокого машинного обучения» (конкурс 2022 года «Проведение фундаментальных научных исследований и поисковых научных исследований малыми отдельными научными группами»).

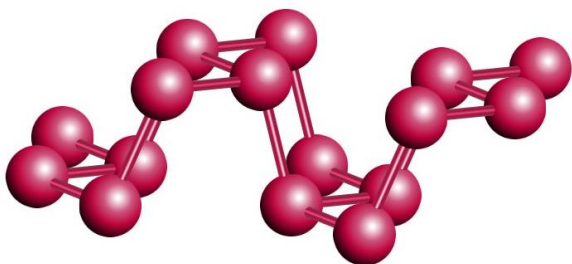


Рис. 1. Элементарная ячейка черного фосфора, используемая при моделировании силового поля

В настоящей работе эта же задача решается средствами пакета SchNetPack [8], основанного на так называемой сверточной сети с непрерывной фильтрацией (continuous-filter convolutional neural network).

В ряде работ эту сеть классифицируют также как графовую [5, 13]. В данном случае исследование теплопроводности фосфора выступает как тестовая задача, позволяющая провести сравнение с результатами работы [12]. В перспективе планируется испытать сеть SchNet для моделирования полимерных материалов. Ожидается, что, подобно тому, как сверточные нейронные сети выполняют агрегацию большого числа пикселей при обработке изображений, так и используемая в настоящей работе сеть SchNet позволит учесть информацию об особенностях конфигурации полимеров, недоступную для сети с прямой связью.

Следует отметить, что большинство работ, посвященных сети SchNet, основаны на использовании датасетов QM9 и MD17 (см. ссылки в [6]), содержащих данные для сравнительно небольших органических молекул. Исследование кристаллических веществ при помощи этой сети, насколько нам известно, не проводилось. В работе [5] проводится сравнение точности вычисления силовых полей нейронными сетями различных типов, в том числе сети SchNet. Авторы [5] выполнили собственную реализацию этой сети [14] в соответствии с описанием авторов [8], что дало возможность более тонкой конфигурации сети.

## II. МОДЕЛИРОВАНИЕ СИЛОВОГО ПОЛЯ ЧЕРНОГО ФОСФОРА СЕТЬЮ SCHNET

В настоящей работе рассматривается черный фосфор [15], фрагмент кристаллической решетки которого представлен на рис. 1. Данные для обучения получены в результате моделирования ячейки, содержащей 16 атомов, при помощи AIMD Кара-Парринелло, реализованной в пакете Quantum ESPRESSO. Для обучения использовались два датасета, содержащие 2000 записей и 100000 записей. Запись содержит сведения о 16 атомах: координаты каждого атома в трехмерном пространстве, три проекции силы, действующих на каждый из атомов, и энергию всей конфигурации атомов.

Архитектура сети SchNet представлена на рис. 2. На каждом из слоев этой сети молекула представляется поатомно. Вектор признаков  $X_l = (x_1^l, x_2^l, \dots, x_n^l)$  формируется из комбинации координат всех атомов  $R = (\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_n)$  и их зарядовых чисел  $Z = (Z_1, Z_2, \dots, Z_n)$ ,  $n$  — количество атомов в исследуемой молекуле,  $l$  — номер слоя нейронной сети. Величина  $x_i^l$  представляет собой, в свою очередь, вектор длиной  $F$ , где  $F$  определяется параметром n\_atom\_basis класса representation.SchNet.

На начальном этапе ( $l = 0$ ) векторы признаков  $x_i^0$  формируются на основе так называемого погружения (англ. embedding), выполняемого над списком зарядовых чисел. Далее в слоях взаимодействия (англ. interaction) к  $x_i^l$  добавляется информация о координатах атомов. Собственно, в этих слоях и выполняется операция свертки. Условно можно сказать, что набор  $x_i^l$  из  $F$  чисел хранит информацию о расположении  $i$ -го атома относительно остальных атомов.

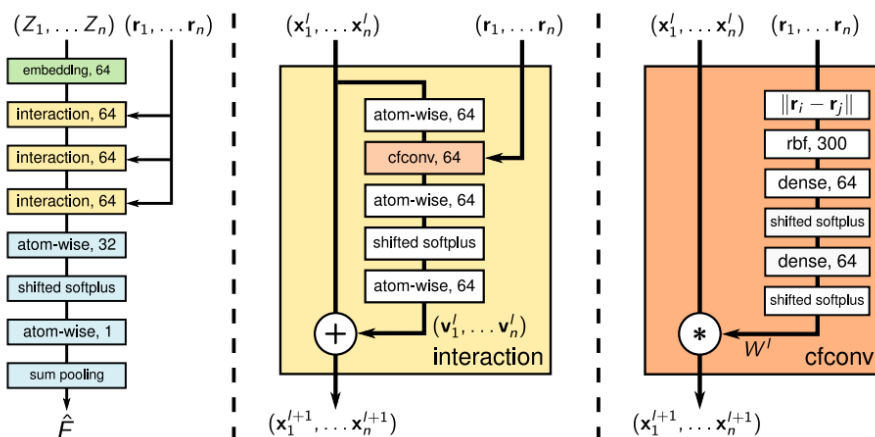


Рис. 2. Архитектура сети SchNet [6]



Используя SchNetPack, можно варьировать следующие параметры:

- $n\_atom\_basis$  ( $F$ ) – количество элементов карты объектов (англ. number of feature map);
- количество «слоев взаимодействия» ( $n\_interactions$ ) – фактически, количество сверточных слоев;
- тип используемых радиально-базисных функций и их количество  $n\_rbf$  (рис. 2, слой rbf, блок cfcov);
- радиус «обрезки»  $cutoff$ , определяющий количество учитываемых соседних атомов;
- алгоритм построения списка соседей.

На основе проведенных вычислительных экспериментов и анализа кривых обучения выбраны  $n\_atom\_basis = 256$ ,  $n\_interactions = 9$ ,  $n\_rbf = 300$ ,  $cutoff = 5.0 \text{ \AA}$ , алгоритм построения списка соседей ASENeighborList, позволяющий учитывать периодические граничные условия. Значение параметра  $cutoff$  выбрано несколько больше постоянной решетки черного фосфора ( $a_x = 4.376 \text{ \AA}$ ,  $a_y = 3.314 \text{ \AA}$  [15], рис. 3). Использование указанного типа формирования списка соседей и значения  $cutoff$  накладывает ограничения на размер батча (параметр  $batch\_size$ ), который при расчетах на видеокарте NVIDIA GeForce RTX 3060 полагали равным 100–200 (при больших значениях  $batch\_size$  возникало переполнение памяти видеокарты). В качестве входных параметров обученная нейронная сеть получает координаты атомов, на выходе формируются значения векторов сил, действующих на каждый атом, а также значение полной энергии системы. Поскольку рассматривается периодическая структура и предполагается в дальнейшем проводить моделирование методом классической молекулярной динамики, весовые коэффициенты при вычислении функции потерь на валидационной выборке выбраны равными 0.99 и 0.01 для сил и энергий, соответственно. В качестве функции потерь используется среднеквадратическая ошибка.

На рис. 4 приведены графики функции потерь на валидационной выборке для двух вариантов обучения сети: 1 – датасет DS1, содержит 2000 записей, размеры обучающей  $num\_train$ , валидационной  $num\_val$  и тестовой  $num\_test$  выборок, соответственно, 1200, 600 и 200, периодические граничные условия не вводятся; 2 – датасет DS2, содержит 100000 записей,  $num\_train=80000$ ,  $num\_val=15000$ ,  $num\_tes = 5000$ , учитываются периодические граничные условия на границах области кристалла, показанной на рис. 1.

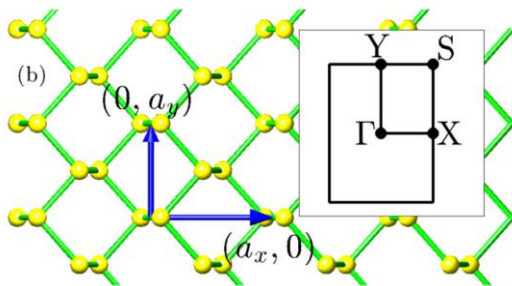


Рис. 3. Кристаллическая решетка черного фосфора [15]

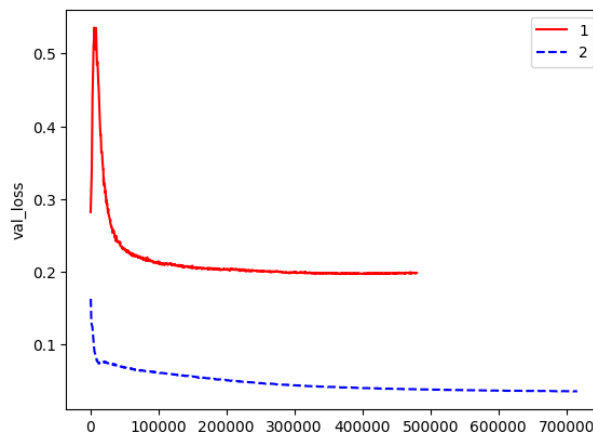


Рис. 4. Функция потерь на валидационной выборке

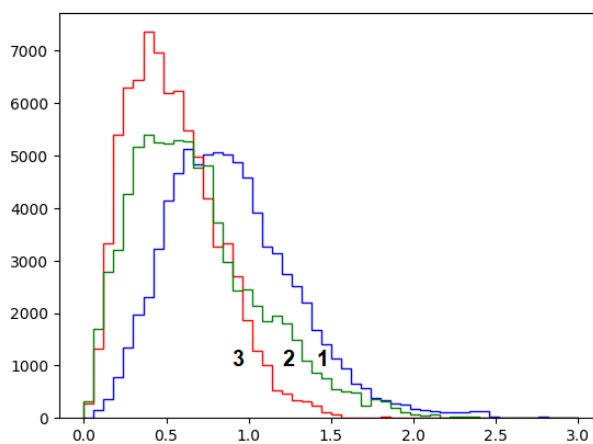


Рис. 5. Гистограмма абсолютных величин сил, действующих на атом: 1 – данные AIMD, 2 – модель, обученная на датасете DS2, 3 – модель, обученная на датасете DS1

Из рис. 4 видно, что в обоих случаях функция потерь выходит на плато, причем в случае большей по объему выборки и использования периодических граничных условий установившееся значение  $val\_loss$  ниже.

Для сравнения качества предсказания силовых полей на рис. 5 показана гистограмма распределения абсолютных величин силы, действующей на атом фосфора. По горизонтальной оси отложены значения модуля силы в единицах  $\text{эВ}/\text{Å}$ , по вертикальной оси – количество значений модуля силы, попадающих в заданный промежуток  $(F, F + \Delta F)$ . Видно, что обе модели несколько занижают модуль силы. Распределение значений абсолютных величин силы, действующей на атом, даваемое моделью, обученной на датасете DS2, ближе к исходным данным, полученным на основе AIMD. Следует отметить, что при приведении к термодинамическому равновесию образца фосфора в ходе молекулярной динамики лучшие результаты показывает, наоборот, модель, обученная на меньшем датасете DS1. В литературе (например, [13]) встречаются сообщения о том, что сеть SchNet дает более точные результаты при использовании обучающих выборок в несколько тысяч записей по сравнению с обучающими выборками, содержащими сотни тысяч записей.

### III. МЕТОД ВЫЧИСЛЕНИЯ ТЕПЛОПРОВОДНОСТИ

В работе используется методика вычисления теплопроводности, описанная в руководстве к пакету LAMMPS (команда `fix heat`). Рассматривается образец, полученный повторением области кристалла, изображенной на рис. 1, 23 раза вдоль оси X (вдоль границы типа `armchair`) и 16 раз вдоль оси Y (вдоль границы типа `zigzag`). Общее число атомов фосфора составило 5888. Длина образца вдоль оси X составляет  $l_x = 202.5 \text{ \AA}$  (рис. 6). Ширина области источника и приемника тепла  $\Delta l = 10 \text{ \AA}$ . Поскольку мы используем периодические граничные условия, располагаем источник тепла на краю образца, а приемник тепла – в середине образца.

Коэффициент теплопроводности  $\kappa$  определяется как коэффициент пропорциональности между тепловым потоком  $J$  и градиентом температуры  $\nabla T$ :

$$\mathbf{J} = -\kappa \nabla T. \quad (1)$$

В представленной на рис. 6 конфигурации рассматривается поток тепловой энергии вдоль оси X, поэтому в (1) переходим к одномерному случаю:

$$J = -\kappa (\nabla T)_x = -\kappa \frac{dT}{dx} \approx -\kappa \cdot \frac{\Delta T}{\left(\frac{l_x}{2}\right)}, \quad (2)$$

где  $\Delta T$  – разность температур между источником и приемником тепла,  $l_x/2$  – расстояние между ними.

Тепловой поток определяется как количество теплоты, прошедшее через поперечное сечение образца в единицу времени:

$$J = \frac{\Delta Q}{\Delta t \cdot S}. \quad (3)$$

Комбинируя (2) и (3), получаем:

$$\kappa = \frac{\Delta Q}{\Delta t \cdot S} \frac{l_x}{2 \Delta T}. \quad (4)$$

Значения теплопроводности вычисляются в среде LAMMPS на основе усреднения по времени разности температур между источником и приемником тепла.

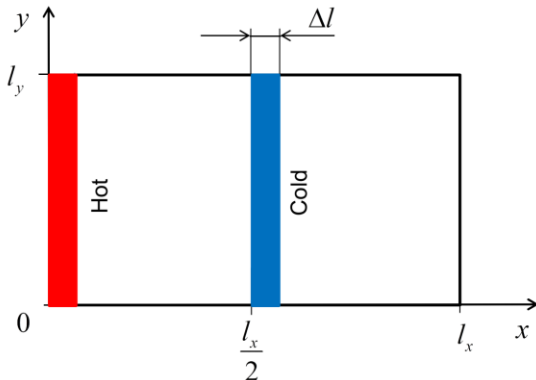


Рис. 6. Расположение областей источника (Hot) и приемника (Cold) тепла

Поскольку мы имеем дело с двумерным образцом, в качестве размера вдоль оси Z возьмем расстояние между

слоями фосфорена  $d \approx 5.455 \text{ \AA}$  [16] в составе объемного образца, так что  $S = d \cdot l_y$ .

Моделирование теплопроводности в среде LAMMPS проводится в два этапа. На первом этапе система приводится к равновесию при температуре 273 K: 10000 шагов при «нагревании» от 0 до 20 K с использованием ансамбля NVE и термостата Берендсена с корректировкой скоростей на каждом временном шаге, 20000 шагов – при нагревании от 20 K до 273 K с использованием ансамбля NVE и термостата Берендсена с корректировкой скоростей каждые 100 шагов, 10000 шагов при поддержании температуры 273 K с использованием ансамбля NVT с корректировкой скоростей каждые 100 шагов. Один шаг по времени соответствует 1 фс. На втором этапе задаем положение «горячего» и «холодного» блоков (источника и приемника тепла), используем ансамбль NVT с корректировкой скоростей каждые 100 шагов (всего 10000 шагов) и вычисляем среднее значение разности температур между блоками. Усреднение проводится через каждые 200 временных шагов, исходные данные обновляются каждые 20 шагов, при усреднении используются 6 предыдущих значений, соответствующих моментам обновления исходных данных.

Моделирование дает значения теплопроводности при распространении тепла вдоль направления границы `armchair`  $\kappa \approx 3 \text{ Вт/(м}^2 \cdot \text{К)}$ . Следует отметить, что в [12] получено значение, примерно в 2 раза меньшее. Причиной расхождения, на наш взгляд, является малое время моделирования методом молекулярной динамики. Кроме того, рассматриваемая методика использует довольно грубое приближение для градиента температуры.

### IV. ЗАКЛЮЧЕНИЕ

В рамках работы были выполнены конфигурация и обучение сети SchNet для получения потенциала межатомного взаимодействия черного фосфорена, проведено сравнение гистограмм модулей сил, действующих на атом, для разных наборов данных и способов обучения. При помощи молекулярной динамики, реализованной в пакете LAMMPS, выполнено приведение системы к равновесию и вычислена теплопроводность образца вдоль направления границы `armchair`. По сравнению с данными, полученными при помощи потенциала, вычисленного пакетом DeePMD, используемый в настоящей работе подход приводит к завышенному значению теплопроводности. В дальнейшем планируется рассчитать также теплопроводность вдоль направления границы `zigzag`, используя другие методики, реализованные в пакете LAMMPS.

### СПИСОК ЛИТЕРАТУРЫ

- [1] Zeng J. et al. DeePMD-kit v2: A software package for deep potential models // Journal of Chemical Physics. 2023. V. 159. P. 054801. DOI: 10.1063/5.0155600
- [2] Batatia I. et al. A foundation model for atomistic materials chemistry // arXiv, Mar 1 2024. DOI: 10.48550/arXiv.2401.00096
- [3] Deng B., Zhong P., Jun K., Riebesell J., Han K., Bartel Ch. J., Ceder G. CHGNet as a pretrained universal neural network potential for

- charge-informed atomistic modelling // *Nature Machine Intelligence*. 2023. V. 5. Pp. 1031-1041. DOI: 10.1038/s42256-023-00716-3
- [4] Chen Chi, Ong Sh. P. A universal graph deep learning interatomic potential for the periodic table // *Nature Computational Science*. 2022. V. 2. Pp. 718-728. DOI: 10.1038/s43588-022-00349-3
- [5] Wollschläger T., Gao N., Charpentier B., Ketata M.A., Günnemann S. Uncertainty estimation for molecules: desiderata and methods // *arXiv*, Jun 20 2023. DOI: 10.48550/arXiv.2306.14916
- [6] Schütt K.T., Saucedo H.E., Kindermans P.-J., Tkatchenko A., Müller K.-R. SchNet – A deep learning architecture for molecules and materials // *Journal of Chemical Physics*. 2018. V. 148. P. 241722. DOI: 10.1063/1.5019779
- [7] K.T. Schütt, P. Kessel, M. Gastegger, K.A. Nicoli, A. Tkatchenko, K.-R. Müller. SchNetPack: A Deep Learning Toolbox for Atomistic Systems // *Journal of Chemical Theory and Computation*. 2019. V. 15. Is. 1. Pp. 448–455. DOI: 10.1021/acs.jctc.8b00908
- [8] Schütt K.T., Hessmann S.S.P., Gebauer N.W.A., Lederer J., Gastegger M. SchNetPack 2.0: A neural network toolbox for atomistic machine learning // *Journal of Chemical Physics*. 2023. V. 158. Is. 14. P. 144801. DOI: 10.1063/5.0138367
- [9] Gebauer N.W.A., Gastegger M., Schütt K.T. Symmetry-adapted generation of 3d point sets for the targeted discovery of molecules // *arXiv*, Jan 9 2020. DOI: 10.48550/arXiv.1906.00957
- [10] Gebauer N.W.A., Gastegger M., Hessmann S.S.P., Müller K.-R., Schütt K.T. Inverse design of 3d molecular structures with conditional generative neural networks // *Nature Communications*. 2022. V. 13. Art. Num. 973. DOI: 10.1038/s41467-022-28526-y
- [11] Шейн Д.В., Завьялов Д.В., Конченков В.И. Исследование применимости нейронных сетей с прямой связью для компьютерного моделирования полимеров // *Журнал технической физики*. 2023. Т. 93. Вып. 12. С. 1732–1735. DOI: 10.61011/JTF.2023.12.56806.f242-23
- [12] Шейн Д.В., Завьялов Д.В., Жариков Д.Н. Моделирование фосфорена методом классической молекулярной динамики с использованием глубокого обучения // *Физика. Технологии. Инновации. ФТИ-2022 : тез. докл. IX Междунар. молодеж. науч. конф., посвящ. 100-летию со дня рожд. проф. С.П. Распопина*. Екатеринбург, 2022. С. 330-331.
- [13] Khrabrov K., Shenbin I., Ryabov A., Tsypin A., Telepov A., Alekseev A., Grishin A., Strashnov P., Zhilyaev P., Nikolenko S., Kadurin A. nablDFT: large-scale conformational energy and hamiltonian prediction benchmark and dataset // *Physical Chemistry Chemical Physics*. 2022. V. 24. P. 25853. DOI: 10.1039/d2cp03966d
- [14] [https://github.com/wollschl/uncertainty\\_for\\_molecules/tree/main/ue4mol](https://github.com/wollschl/uncertainty_for_molecules/tree/main/ue4mol)
- [15] Li P., Appelbaum I. Electrons and holes in phosphorene // *Physical Review B*. 2014. V. 90. P. 115439. DOI: 10.1103/PhysRevB.90.115439
- [16] Luo Z., Maassen J., Deng Y., Du Y., Garrelts R.P., Lundstrom M.S., Ye P.D., Xu X. Anisotropic in-plane thermal conductivity observed in few-layer black phosphorus // *Nature Communications*. 2015. V. 6. P. 8572.

# Использование искусственного интеллекта и технологии перовскитных материалов для мониторинга светового загрязнения

Д. Ю. Терехов, А. Р. Попова, Е. В. Патока,  
В. Д. Бурлака, М. С. Козлов

Санкт-Петербургский государственный электротехнический университет  
«ЛЭТИ» им. В.И. Ульянова (Ленина)

ditereh@gmail.com, renoridoru@gmail.com, yegor-patoka@mail.ru,  
byrlaka60@gmail.com, mex.koz01@gmail.com

**Аннотация.** В данной статье исследуется использование искусственного интеллекта и технологии перовскитных материалов для мониторинга светового загрязнения. Рассматривается методика, включающая установку цифровой камеры и прибора для измерения качества ночного неба на беспилотный летательный аппарат (БПЛА) с последующим анализом данных с помощью алгоритмов машинного обучения и глубокого обучения. Кроме того, обсуждается перспективность применения перовскитных материалов для повышения точности сбора данных о световом загрязнении.

**Ключевые слова:** искусственный интеллект, перовскитные материалы, световое загрязнение, мониторинг, беспилотные летательные аппараты, машинное обучение, глубокое обучение, оптические приборы, ночное небо

## I. ВВЕДЕНИЕ

Ухудшение качества окружающей среды и здоровья человека – насущные проблемы, над которыми бьются исследователи и ученые.

Одной из таких проблем является световое загрязнение – постоянно увеличивающееся количество искусственного света (уличное освещение, автомобильные фары, рекламные вывески и щиты), оказывающее негативное влияние на окружающую среду и здоровье человека. С увеличением количества света нарушаются биологические ритмы растений, животных и, конечно, человека, разрушаются экосистемы и климат [1].

Мониторинг светового загрязнения позволит оценить влияние света на здоровье человека, качество окружающей среды и разработать стратегии по снижению его негативного воздействия.

Данная статья направлена на изучение методов и перспектив использования новых технологий мониторинга светового загрязнения для улучшения качества жизни.

В статье рассматривается использование беспилотных летательных аппаратов (БПЛА) с различными приложениями и инструментами для измерения светового загрязнения. Собранные данные будут проанализированы с помощью алгоритмов

машинного обучения и глубокого обучения для прогнозирования светового загрязнения, мониторинга эффектов искусственного света и снижения его негативного воздействия. Кроме того, мы обсудим перовскитные материалы как перспективное направление в области фотоприемников [2].

В статье будет представлена методика использования БПЛА с перовскитным напылением для сбора данных, а также алгоритмы искусственного интеллекта для их анализа.

Исследование будет полезно для специалистов в области экологии, астрономии, медицины, а также для тех, кто интересуется развитием технологий в области экологического мониторинга.

## II. ПРЕДПОСЫЛКИ МЕТОДА

В этом разделе рассматривается методика использования беспилотных летательных аппаратов (БПЛА) с фотодетектором для мониторинга светового загрязнения и анализа полученных данных с помощью искусственного интеллекта. Методика состоит из нескольких ключевых этапов.

На первом этапе необходимо выбрать БПЛА и установить на его корпус фотодетекторы. Фотодетекторы представляют собой электронные элементы (фотодиод, фототранзистор и т.п.) с напылением перовскита. Этот материал в последние годы привлекает внимание научного сообщества, так как, в отличие от фотодетекторов на основе сульфида кадмия и свинца, он обладает высокой способностью поглощать электромагнитное излучение в видимом диапазоне [3], фотолуминесценцией с возможностью управления длиной волны поглощаемого излучения. Эффективность рассматриваемых фотодетекторов была продемонстрирована как в широком, так и в узком оптических диапазонах [4]. Еще одним преимуществом является их чувствительность к рентгеновскому и гамма-излучению [5–8]. Это дает возможность расширить спектр получаемых данных. Датчики используются для измерения таких параметров, как интенсивность света, спектральный состав и другие характеристики, которые могут быть важны для оценки уровня светового загрязнения.

Второй этап – составление карты маршрута полета БПЛА над выбранными районами города. Благодаря автоматизации полета обеспечивается высокая точность и большая площадь покрытия, что позволяет расширить спектр собираемых данных. Особенностью сбора необходимых данных с помощью БПЛА можно отметить низкую высоту полета, что существенно влияет на повышение точности получаемой информации. Это является большим преимуществом перед традиционным методом мониторинга – спутниковым анализом.

На третьем этапе собранные данные анализируются с помощью алгоритмов искусственного интеллекта (ИИ). ИИ может использоваться для обработки больших объемов данных, выявления сложных закономерностей и составления прогнозов на основе обучаемых моделей. В контексте мониторинга светового загрязнения ИИ может определять уровень светового загрязнения, идентифицировать источники света, визуализировать и прогнозировать тенденции негативного воздействия. Для анализа результатов необходимо запрограммировать модель ИИ на основе обучающего набора данных. Этот набор данных может включать в себя измерения, проведенные в различных условиях и в разное время суток, а также информацию об известных источниках светового загрязнения.

Модель ИИ обучается распознавать и классифицировать различные типы загрязнения. После обучения ИИ можно использовать для анализа новых данных, собранных БПЛА. Это позволит быстрее и точнее оценивать уровень светового загрязнения в режиме реального времени. Результаты анализа данных ИИ могут быть визуализированы и интерпретированы для получения полезной информации. Это может включать создание карт загрязнения, анализ негативных тенденций и выводы о влиянии светового загрязнения на здоровье человека и окружающую среду.

### III. ПОЛУЧЕНИЕ И АНАЛИЗ ДАННЫХ О СВЕТОВОМ ЗАГРЯЗНЕНИИ

В этом разделе мы подробно рассмотрим, как БПЛА будут летать и передавать данные для анализа.

Перед началом полета БПЛА программируется либо блок управления летательным аппаратом как отдельное устройство, либо сам дрон, после чего определяется предварительный маршрут и время работы. Примером такой работы с БПЛА, обученными ИИ, является использование беспилотников для разведки в различных погодных условиях [9].

Маршрут полета планируется таким образом, чтобы максимально охватить выбранную территорию. Алгоритм полета предполагает отсутствие человека в качестве пилота. Использование БПЛА с искусственным интеллектом в качестве пилота было исследовано в работе «Искусственный интеллект в беспилотных авиационных системах» компании «Аэросвит» [9]. Изначально существуют только стартовые и финишные точки и ограничения по покрытию, чтобы обеспечить безопасность в зоне полета. Это позволит БПЛА летать над различными типами местности при разных условиях освещения.

Фотодетекторы измеряют интенсивность света, спектральный состав и другие параметры. Собранные БПЛА данные передаются на землю в режиме реального времени по различным каналам связи, включая GSM, Wi-Fi, LTE. После передачи данные обрабатываются и анализируются с помощью алгоритмов искусственного интеллекта. Метод работы описывается следующим образом.

Сначала аналоговый сигнал, полученный БПЛА, декодируется в цифровые данные, которые собираются в массив.

Затем обработанные данные могут быть предоставлены искусственному интеллекту либо для обучения, либо для обнаружения закономерностей.

Алгоритм обучения может быть основан на методе конволюционных нейронных схем. Их идея заключается в чередовании конволюционных, субдискретных и полностью связанных слоев на выходе.

Изображения классифицируются с помощью фильтра, который представляет собой массив коэффициентов, задаваемых исследователем вручную. Эталонный массив с помощью математического метода, называемого сверткой, сравнивается с полученными измерениями. Операция заключается в том, что каждая ячейка имеющейся матрицы суммируется и записывается в аналогичную позицию выходного изображения. Особенность этого фильтра в том, что значение его выхода становится тем больше, чем больше изображение похоже на сам фильтр. Таким образом, свернутое изображение даст нам еще одно изображение, каждый пиксель которого будет указывать на степень сходства фрагмента изображения с фильтром. Результатом такой операции является карта признаков. Каждый фрагмент изображения умножается на небольшую матрицу весов, затем результат суммируется, в данном случае сумма – это пиксель выходного изображения, то есть карта признаков [10].

Постоянное чередование слоев позволит создавать карты признаков из карт признаков, что поможет распознавать самые сложные классификации.

На основе этих классификаций можно составить такие классы изображений:

**Загрязненные территории:** к этому классу относятся территории, где световое загрязнение превышает определенный порог. К таким областям можно отнести те, где интенсивность света превышает нормативные значения, что может негативно сказаться на здоровье человека и окружающей среде.

**Незагрязненные территории:** к этому классу относятся территории с уровнем светового загрязнения, соответствующим нормативным значениям. Эти зоны можно определить как те, где интенсивность света находится в пределах допустимых значений, что соответствует здоровым условиям жизни и работы.

**Источники света:** в некоторых случаях классификация может включать определение конкретных источников света, таких как уличные фонари, рекламные вывески, автомобильные фары и т. д. Это позволяет не только оценить общий уровень

загрязнения, но и определить, какие источники света подвержены наибольшему воздействию.

**Временные периоды:** в зависимости от целей исследования, занятия могут включать различные временные периоды, например, ночное время, утренние и вечерние часы, а также дни недели. Это позволит вам проанализировать, как меняется уровень светового загрязнения в течение дня и в разные дни недели.

**Времена года:** в некоторых случаях классы можно разделить на сезонные периоды, такие как зима, весна, лето и осень. Это может помочь в анализе того, как сезонные изменения влияют на уровень светового загрязнения.

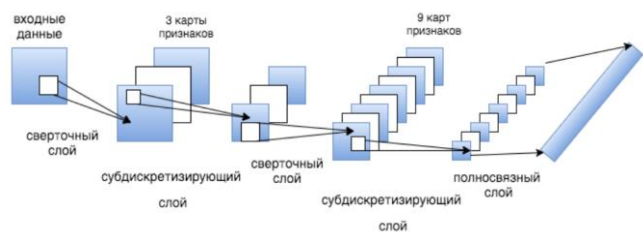


Рис. 1. Алгоритм метода свертки, интегрированного в ИИ

#### IV. ОЦЕНКА ВИЗУАЛЬНЫХ ДАННЫХ, ПОЛУЧЕННЫХ В ХОДЕ ИЗМЕРЕНИЙ

В этой части статьи рассматривается визуальная часть анализа, проводимого искусственным интеллектом на основе данных с БПЛА в режиме реального времени.

Цветовая палитра на картах светового загрязнения может быть использована для визуализации уровня загрязнения. Например, темные области могут указывать на низкий уровень загрязнения, а светлые – на высокий. Анализ цветовой палитры может помочь определить места, на которые стоит обратить внимание.

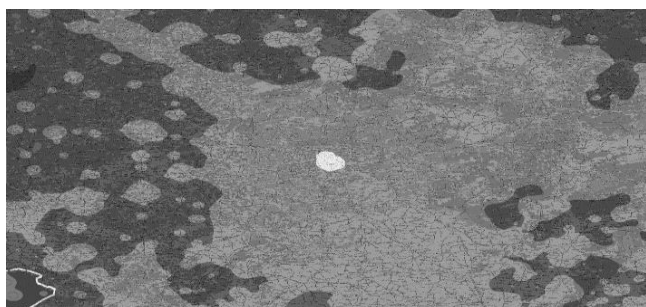


Рис. 2. Пример полученной карты светового загрязнения Москвы

- Критический уровень, необходимы срочные меры по снижению светового загрязнения
- Высокий уровень, необходимы решительные меры по снижению светового загрязнения
- Повышенный уровень, необходимы серьезные меры по снижению светового загрязнения
- Средний уровень, необходимы меры по снижению светового загрязнения
- Умеренный уровень, необходимы упреждающие меры по снижению светового загрязнения
- Низкий уровень, оптимальный уровень светового загрязнения
- Незначительный уровень, необходимых мер не требуется

Рис. 3. Оттенки цветов, воспринимаемые нейронной сетью

Сравнение визуальных данных с исходными или предыдущими измерениями может помочь оценить изменения в уровне загрязнения. Например, можно сравнить карту светового загрязнения текущего года с картой, созданной в прошлом году, чтобы увидеть, как изменилась ситуация.

Алгоритмы искусственного интеллекта могут быть обучены распознавать и классифицировать различные типы загрязнения на основе визуальных данных. Модель обучается на изображениях, которые ранее были помечены как загрязненные или не загрязненные, что позволяет использовать модель для анализа новых данных.

В некоторых случаях может потребоваться профессиональная оценка визуальных данных (рис. 4). Эксперты могут использовать свой опыт для оценки визуальных данных, выявляя области, которые могут быть недооценены или неверно интерпретированы автоматизированными методами.

Предположим, существует карта светового загрязнения в Москве. Чтобы оценить эту карту, необходимо начать с анализа цветовой палитры, чтобы определить, какие районы наиболее загрязнены. Затем, используя инструменты анализа изображений, извлекаются цветовые и яркостные характеристики этих районов. Сравнив карту с исходными данными (рис. 5), можно оценить, как изменилось текущее состояние. Наконец, для автоматического анализа изображений можно использовать алгоритмы машинного обучения. Экспертная оценка может быть использована для подтверждения или корректировки результатов, полученных искусственным интеллектом.

Перечисленные выше методы позволят нам оценить визуализированные данные о световом загрязнении, предоставив ценную информацию для планирования и принятия мер по решению описанной проблемы.

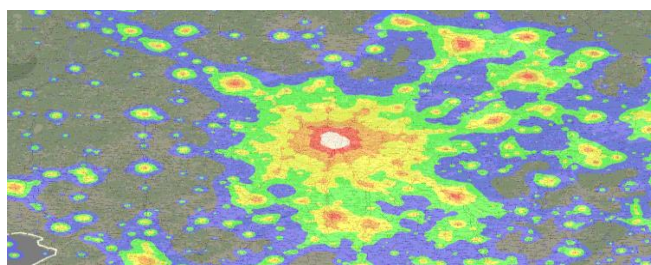


Рис. 4. Визуальная карта светового загрязнения в Москве

- Критический уровень, необходимы срочные меры по снижению светового загрязнения
- Высокий уровень, необходимы решительные меры по снижению светового загрязнения
- Повышенный уровень, необходимы серьезные меры по снижению светового загрязнения
- Средний уровень, необходимы меры по снижению светового загрязнения
- Умеренный уровень, необходимы упреждающие меры по снижению светового загрязнения
- Низкий уровень, оптимальный уровень светового загрязнения
- Незначительный уровень, необходимых мер не требуется

Рис. 5. Цветовая палитра загрязнений, удобная для визуального восприятия человеком

## V. ЗАКЛЮЧЕНИЕ

В заключение следует отметить, что предложенная методика оценки светового загрязнения, основанная на использовании БПЛА, перовскитных датчиков и искусственного интеллекта, открывает перспективные возможности для анализа и предотвращения негативного влияния светового загрязнения на здоровье человека и окружающую среду. В исследовании используется методологический подход для отслеживания, сравнения и анализа полученных данных.

Предварительные результаты показывают, что эта методология может стать ценным инструментом в руках инженеров, светотехников и астрономов, которая поможет предотвратить пагубные последствия светового загрязнения. Опираясь на существующие данные и установленные закономерности с помощью искусственного интеллекта, исследование призвано предоставить информацию, полезную для экологических и инженерных приложений.

Будущие исследования в этой области будут направлены на совершенствование алгоритмов обучения ИИ, использование самых передовых технологий перовскитных датчиков и модернизацию БПЛА, что в конечном итоге повысит точность, скорость и реакцию на полученные данные.

## СПИСОК ЛИТЕРАТУРЫ

- [1] Longcore Travis, and Rich Catherine. Ecological light pollution. // *Frontiers in Ecology and the Environment* 2.4 (2004): 191-198.
- [2] Zhao Y., Zhu K. Organic–inorganic hybrid lead halide perovskites for optoelectronic and electronic applications // *Chem. Soc. Rev. Royal Society of Chemistry*, 2016. Vol. 45, №3. P.655–689.
- [3] Fujiwara H. et al. Optical Characteristics and Operational Principles of Hybrid Perovskite Solar Cells // *Phys. Status Solidi Appl. Mater. Sci.* 2018. Vol. 215, № 12. P. 1–10.
- [4] Saidaminov M.I. et al. Perovskite Photodetectors Operating in Both Narrowband and Broadband Regimes // *Adv. Mater.* 2016. Vol. 28, № 37. P. 8144–8149. 170
- [5] Yakunin S. et al. Detection of X-ray photons by solution-processed lead halide perovskites // *Nat. Photonics. Nature Publishing Group*, 2015. Vol. 9, № 7. P. 444–449.
- [6] Birowosuto M.D. et al. X-ray scintillation in lead halide perovskite crystals // *Sci. Rep. Nature Publishing Group*, 2016. Vol. 6, № November. P. 1–10.
- [7] Xu Q. et al. Bulk Organic-Inorganic Methylammonium Lead Halide Perovskite Single Crystals for Indirect Gamma Ray Detection // *ACS Appl. Mater. Interfaces.* 2019. Vol. 11, № 50. P. 47485–47490.
- [8] He Y. et al. Resolving the Energy of  $\gamma$ -Ray Photons with MAPbI<sub>3</sub> Single Crystals // *ACS Photonics.* 2018. Vol. 5, № 10. P. 4132–4138.
- [9] Matyukha S. et. al. Artificial intelligence in unmanned aircraft systems // Ph.D., Deputy General Director of Aerosvet, LLC. DOI 10.52375/20728689\_2022\_1\_8
- [10] Mustaev A.F. The use of neural networks in image recognition // UDC004.932

# 3D-сегментация опухоли головного мозга с интерполяцией с использованием глубоких нейронных сетей для последующей 3D-визуализации

А. С. Сырых

Национальный исследовательский  
университет ИТМО

alyoshca.syryh@mail.ru

Г. О. Бондаренко

Санкт-Петербургский государственный  
электротехнический университет  
«ЛЭТИ» им. В.И. Ульянова (Ленина)

Olakola9@gmail.com

**Аннотация.** В этой статье представлен метод 3D-сегментации опухоли головного мозга на основе медицинских 3D-сканирований с использованием глубокого обучения. Этот метод позволяет уменьшить размер шага при выполнении 2D-срезов, чтобы получить более информативный анализ и создать 3D-объект, который более точно соответствует форме опухоли, чем существующие аналоги. Представлена подготовка данных для обучения нейронных сетей, практические примеры исследований, демонстрирующие потенциал этой передовой методологии в области медицинской диагностики, хирургических процедур и обучения.

**Ключевые слова:** 3D сегментация головного мозга, 3D визуализация, МРТ головного мозга

## I. ВВЕДЕНИЕ

В современной медицинской практике инновационные технологии и методы анализа данных играют центральную роль в повышении точности диагностики и терапии. Среди них нейронные сети выделяются своей способностью анализировать и интерпретировать сложные медицинские изображения, включая трехмерное (3D) медицинское сканирование, с беспрецедентным уровнем точности и детализации. В этой статье мы исследуем применение нейронных сетей для 3D-сегментации опухолей на основе 3D-медицинских снимков опухолей головного мозга.

Опухоли головного мозга представляют собой серьезную медицинскую проблему, и их диагностика и лечение требуют высокой степени точности и детализации. Традиционные методы анализа медицинских изображений, хотя и полезны, часто имеют ограничения в способности обнаруживать и характеризовать опухоли головного мозга, особенно в случаях сложной формы и структуры.

С развитием глубокого обучения и нейронных сетей стало возможным применять современные методы машинного обучения для автоматической сегментации опухолей головного мозга на 3D-медицинских снимках. Это открывает новые перспективы для раннего выявления и лечения опухолей головного мозга, а также для мониторинга их динамики.

Актуальность разработки нейронных сетей для 3D-сегментации и визуализации медицинских 3D-изображений трудно переоценить, поскольку эти технологии представляют собой значительный прорыв в области медицины. Нейронные сети позволяют обнаруживать и сегментировать опухоли и другие патологические образования в органах с высокой точностью и детализацией, способствуя раннему выявлению и точной диагностике заболеваний [1].

Точная сегментация опухоли на медицинских 3D-изображениях позволяет разрабатывать индивидуальные планы лечения. Это важно для определения объема хирургического вмешательства и выбора оптимального метода лечения для каждого пациента.

3D-сегментация позволяет хирургам планировать и выполнять хирургические процедуры с большей точностью. Это особенно важно при проведении сложных операций, таких как удаление опухоли головного мозга, где точная навигация играет решающую роль.

Сегментация и визуализация изменений на 3D-медицинских изображениях также позволяют отслеживать динамику заболевания у пациентов. Это полезно для оценки эффективности лечения и принятия решения о необходимости корректировки терапии.

Трехмерная сегментация также имеет большое значение для медицинских исследований, позволяя анализировать большие массивы данных и выявлять новые закономерности. Она также может быть использована для обучения будущих медицинских работников [2].

Пациенты также могут воспользоваться 3D-визуализацией сегментации для лучшего понимания своего состояния и плана лечения. Это может помочь повысить осведомленность пациентов и вовлеченность в процесс лечения.

## II. ЛИТЕРАТУРНЫЙ ОБЗОР

Многие современные исследования в области сегментации медицинских изображений прибегают к глубокому обучению, в частности, к сверточным



нейронным сетям (CNN), для сегментации опухолей. В данной сфере лидируют такие архитектуры, как FCN и U-Net, которые построены полностью на сверточных слоях формата энкодер-декодер нейросети без полносвязных слоев. Данная архитектура в отличие от других нейросетей для сегментации, как Mesh-RCNN или YOLO более точно справляется с задачей получения четких контуров интересующих областей, что является критическим и основополагающим фактором для сегментации головного мозга и медицинских изображений в целом.

Последующие исследования были направлены на улучшение существующих архитектур и решения проблем переобучения, скорости обучения нейросети и повышения точности сегментации. Некоторые исследования исследуют подходы, комбинирующие классические методы обработки изображений с методами глубокого обучения [3]. Например, в работе [4] авторы предложили гибридную модель, которая комбинирует в себе несколько архитектур сверточных нейронных сетей, таких как U-Net и ResNet, с целью повышения точности сегментации опухолей головного мозга.

Результаты экспериментов показали, что гибридная модель демонстрирует улучшение производительности по сравнению с другими существующими моделями и методами сегментации опухолей головного мозга, а также проявляет большую адаптивность к различным типам опухолей головного мозга и способна более эффективно сегментировать как крупные, так и мелкие опухоли, чем не гибридные варианты, что указывает на потенциал дальнейшего развития гибридных архитектур сверточных нейронных сетей для улучшения сегментации опухолей головного мозга и их применения в клинической практике.

Некоторые исследования также рассматривают использование ансамблевых методов или комбинацию данных с разных модальностей МРТ (например, T1, T2, FLAIR и т. д.) для более точной сегментации. В работе [5] предлагается использовать ансамбль, состоящий из нескольких 2D сверточных нейронных сетей (CNN), каждая из которых обучается независимо на различных срезах МРТ-изображений. Для каждого слоя ансамбля используется различная архитектура CNN, такая как U-Net, ResNet и другие, что позволяет учесть разнообразные аспекты и структуры опухолей головного мозга. После обучения каждой нейронной сети, их предсказания объединяются в единое решение путем агрегации или голосования, что позволяет улучшить общую точность сегментации. Это исследование подчеркивает эффективность и применимость ансамблей из двумерных сверточных нейронных сетей для задачи сегментации опухолей на МРТ-изображениях головного мозга, предоставляя важные результаты для улучшения диагностики и лечения пациентов с опухолями головного мозга.

На данный момент U-Net является одной из наиболее широко используемых архитектур для сегментации медицинских изображений. Модификации U-Net, такие как 3D U-Net для трехмерных данных и различные

вариации U-Net с адаптацией под конкретные задачи и типы данных, демонстрируют высокую эффективность.

Применение различных методов аугментации данных, таких как повороты, отражения, масштабирование и т. д., в сочетании с глубокими сверточными сетями позволяет улучшить обобщающую способность моделей и снизить риск переобучения. Хотя архитектуры DeepLab и Mask R-CNN не предназначены изначально для сегментации медицинских изображений, они широко применяются в медицинских изображениях после некоторых адаптаций, которые позволяют достичь качественного результата за более короткое время.

Для улучшения результатов сегментации также применяют использование информации из нескольких модальностей (например, совмещение МРТ и ПЭТ изображений), что позволяет улучшить точность сегментации, а также совместное обучение сетей для выполнения нескольких связанных задач (например, сегментации и классификации), применение CRF для улучшения границ сегментации и снижения ошибок путем учета контекстной информации и использование методов, таких как активное обучение и самообучение, для эффективной аннотации медицинских изображений и улучшения обучения моделей.

Для решения второго этапа задачи – регрессии для получения промежуточных срезов МРТ-снимков опухолей головного мозга – могут применяться различные методы машинного обучения и глубокого обучения такие, как линейная регрессия для аппроксимации зависимости между существующими срезами и их позициями в пространстве, методы интерполяции для генерации промежуточных срезов на основе существующих данных, CNN для выполнения регрессии с использованием пары входных и выходных снимков, чтобы предсказать промежуточные срезы, а также GAN, которая генерирует реалистичные промежуточные снимки на основе существующих данных, без явного определения зависимости между входом и выходом.

### III. МЕТОДЫ

В этом разделе мы опишем подготовку набора данных к формату 3D-медицинских изображений для обучения нейронной сети задаче 3D-сегментации, представим алгоритм и обоснуем выбор архитектуры для получения масок сегментации опухолей головного мозга.

Каждое трехмерное медицинское изображение состоит из 23 двумерных срезов головного мозга на разных уровнях, полученных в результате одного сканирования. В свою очередь, метки для сегментации опухоли головного мозга содержат маски из 2D-срезов головного мозга, указывающие область опухоли на каждом уровне, начиная с соответствующего уровня, на котором расположены конкретный срез головного мозга и опухоль. На рис. 1 показан 2D-срез 3D-медицинского изображения и маска опухоли головного мозга.

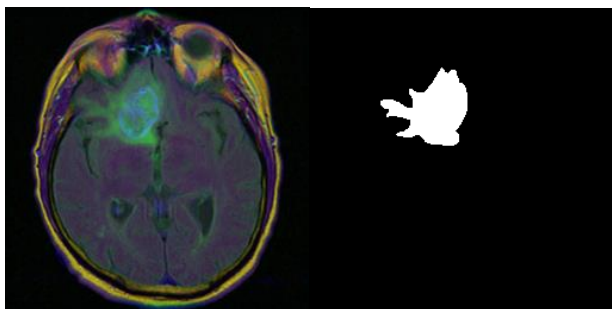


Рис. 1. 2D-снимки головного мозга и метки опухоли

Нейронная сеть должна выводить маски, соответствующие уровням 2D-срезов, и между каждой маской 2D-среза нейронная сеть интерполирует выбранное количество масок для уровней, где на медицинском изображении не было соответствующих 2D-срезов. В результате нейронная сеть должна выдавать более точный и плавный результат по сравнению со 23 срезами вместо 70.

Для получения интересующего результата необходимо:

- Аугментация данных, масштабирование и увеличение контрастности.
- Выбрать архитектуру нейронной сети для 3d-сегментации головного мозга и подбор гиперпараметров для получения наилучшей точности.
- Полученные маски после сегментации передаются в следующую нейросеть для получения большего количества масок и более плавного перехода между слоями.
- Визуализация результата алгоритма.

Для 3d-сегментации была выбрана модификация архитектуры 3D-UNET, потому что данная модификация учитывает объемную информацию, что важно для медицинских изображений, где анатомические структуры имеют трехмерную природу, как в случае с сегментацией головного мозга по МРТ-снимкам. Это позволяет получать более полное представление о форме и структуре объектов. Благодаря использованию трехмерных данных 3D-UNET может обеспечивать более точную сегментацию, особенно для объемных структур или структур, которые распространяются в трех измерениях, а также может уменьшить потерю информации, поскольку он учитывает контекст изображения в трех плоскостях, а не в двух.

Для задачи регрессии был выбран метод GAN, т. к. GAN может помочь улучшить качество сегментации путем генерации дополнительных промежуточных снимков, которые могут устранить артефакты, шумы или недостаточное разрешение на исходных снимках, генерировать маски опухолей, которые соответствуют анатомическим особенностям и характеристикам масок на реальных МРТ-снимках, что предоставит более реалистичную 3d-визуализацию и параметры опухоли головного мозга. В качестве генератора будет выступать архитектура UNET, где на вход будут подаваться маски двумерных срезов опухоли головного мозга, а на выходе – промежуточные двумерные срезы между

входными срезами. В качестве дискриминатора будет выступать классификатор с архитектурой ResNet.

#### IV. РЕЗУЛЬТАТЫ

Для оценки эффективности метода мы используем две ключевые метрики: Intersection over Union (IoU) и индекс структурного сходства (SSIM). IoU – это отношение общей площади пересечения между сегментированным объектом и эталонным образцом к общей площади обоих объектов. Эта метрика часто используется для анализа результатов сегментации, чем ближе результат к 1, тем выше вероятность правильно сегментировать область. Результат анализа IoU показан на рис. 2.

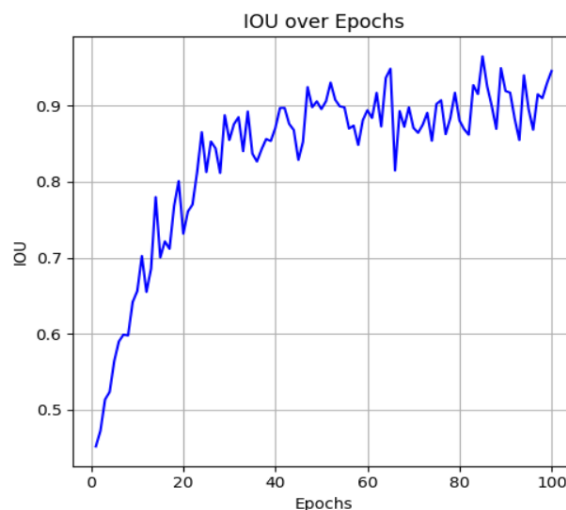


Рис. 2. Оценка IoU на каждую эпоху

SSIM как метрика для оценки качества изображения, которая в данном случае используется для оценки улучшения разрешения трехмерных объектов.

Когда речь идет о сегментации трехмерных медицинских изображений, улучшение разрешения может означать повышение детализации и четкости границ таких структур, как опухоли или анатомические элементы. Чем отчетливее и детальнее структуры на изображении, тем выше значение SSIM. Результаты SSIM представлены на рис. 3.

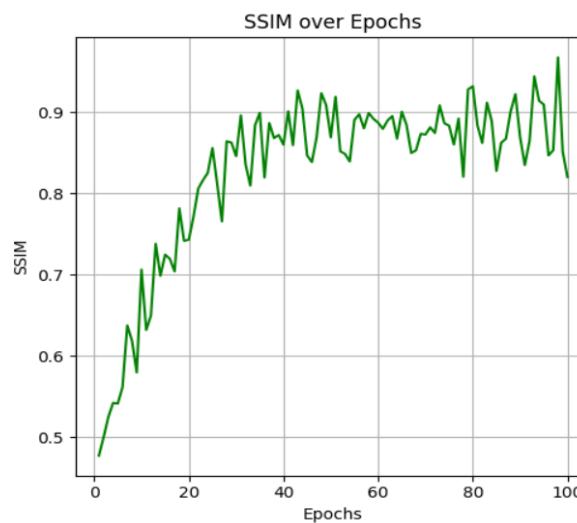


Рис. 3. Оценка SSIM на каждую эпоху

Анализ результатов показал, что значения IOU и SSIM имеют устойчивую тенденцию к увеличению с продолжительностью обучения. Это явление характеризуется стремлением метрик к единице, что говорит о высокой способности предложенного метода достигать высокой степени согласованности с эталонными данными при сегментации медицинских изображений мозга.

Результаты исследования свидетельствуют о высокой эффективности предложенного метода сегментации опухолей на медицинских МРТ-изображениях головного мозга. Модификация архитектуры 3D-UNET для сегментации и применение метода генеративной адверсарной сети (GAN) для улучшения качества аугментации позволили достичь высокой точности и качества сегментации

## V. ЗАКЛЮЧЕНИЕ

Предложенный метод 3D-сегментации с интерполяцией с использованием глубокого обучения для последующей визуализации имеет значительные перспективы в области медицинской диагностики и визуализации. Применение технологий, позволяющих реконструировать детальные 3D-модели анатомических структур и патологий, потенциально может повысить точность и информативность диагностики, а также облегчить взаимодействие между медицинскими работниками и пациентами.

Интерполяция с использованием нейронных сетей может быть использована для создания плавных и непрерывных 3D-моделей на основе ограниченного набора данных. Это особенно ценно при визуализации результатов сегментации, поскольку обеспечивает более наглядное представление о структуре опухоли или другой анатомической области [4].

Проведение интерполяции с глубоким обучением во время сегментации также может способствовать более точному определению размера и объема опухоли.

Однако, как и в случае с любым новым методом, важно помнить, что передовые технологии требуют дополнительных исследований и валидации, прежде чем широко внедряться в медицинскую практику. Это включает в себя проверку результатов на больших группах пациентов и оценку их эффективности и безопасности [5].

Тем не менее, перспективы метода 3D-сегментации с интерполяцией с использованием глубокого обучения для визуализации результатов 3D-сегментации весьма многообещающие. Дальнейшие исследования и работа в этой области могут улучшить способность медицинских работников видеть и понимать анатомические особенности и патологии, что приведет к более точной постановке диагноза и более эффективному лечению.

Будущие исследования в этой области будут включать разработку архитектуры глубокого обучения для решения заявленной задачи, улучшение результатов, изучение максимального порога расширения уровня без ущерба для качества прогнозирования и сравнение с существующими методами, подходами и исследованиями.

## СПИСОК ЛИТЕРАТУРЫ

- [1] Smarta Sangui. "3D MRI Segmentation using U-Net Architecture for the detection of Brain Tumor", 2023.
- [2] Rickmann A, Roy A, Sarasua I, Navab N, Wachinger C. 'Project & Excite' modules for segmentation of volumetric medical scans. Image Video Processing. 2019.
- [3] Chenchu Xu. "Contrast Agent-free Synthesis and Segmentation of Ischemic Heart Disease Images using Progressive Sequential Causal GANs", 2020.
- [4] Dintisrang Daimary. "Brain Tumor Segmentation from MRI Images using Hybrid Convolutional Neural Networks", 2020.
- [5] Mark Lyksborg. "An Ensemble of 2D Convolutional Neural Networks for Tumor Segmentation", 2015.

# Применение модуля внимания в свёрточной нейронной сети в задаче анализа ОКТ-снимков

Е. Н. Волков

Федеральный исследовательский центр «Информатика и управление»  
Российской академии наук  
envolkoff@gmail.com

**Аннотация.** Оптическая когерентная томография является одним из самых перспективных методов диагностики дегенеративных заболеваний и патологических состояний сетчатки глаза (диабетический макулярный отёк, возрастная макулярная дегенерация, друзы и т. д.). Однако анализ таких снимков требует от врача офтальмолога не только высокой квалификации, но и значительного времени на поиск патологий. Использование искусственных нейронных сетей для решения данной задачи сможет ускорить анализ снимков, поможет минимизировать или даже избежать ошибок в процессе диагностики. В работе представлен процесс создания интерпретируемого программного решения для распознавания патологий сетчатки глаза по снимкам оптической когерентной томографии, включая обзор существующих решений, подбор набора обучающих данных, создание алгоритма предподготовки данных, выбор архитектуры искусственной нейронной сети, имплементация метода объяснительного искусственного интеллекта, а также перспектив развёртывания решения для применения в клинической практике.

**Ключевые слова:** искусственный интеллект, объяснительный искусственный интеллект, искусственные нейронные сети, система поддержки принятия врачебных решений, оптическая когерентная томография, офтальмология, медицина

## I. ВВЕДЕНИЕ

Использование систем на основе искусственного интеллекта (ИИ) выводит диагностику на основе анализа медицинских изображений на совершенно новый уровень и является одной из ключевых черт парадигмы Здравоохранение 5.0 [1]. Однако в таких областях как офтальмология, технологии ИИ не получили широкого распространения. Особенно это касается такого метода диагностики заболеваний сетчатки глаза как оптическая когерентная томография (ОКТ). Существуют отдельные обзоры исследований данной области, показывающие преимущества использования свёрточных нейронных сетей (СНС) для классификации патологий сетчатки [2,3]. В то время как появление архитектур визуальных трансформеров значительно изменило подходы к нейросетевому анализу изображений, преимущества СНС для решения данной задачи очевидны (малые выборки данных, простота обучения и т. д.). Исследования возможностей интеграции одного из ключевых аспектов архитектуры трансформера – механизма внимания в СНС может дать прирост в точности работы сети.

Исследование выполнено за счет гранта Российского научного фонда (проект № 24-21-00330), <https://rscf.ru/project/24-21-00330/>.

Корректность работы таких модификаций СНС может быть подтверждена использованием методов объяснительного ИИ в рамках построения пайплайна. Использование методов объяснения при анализе офтальмологических изображений имеет высокую актуальность, что показано в работах [4, 5].

## II. МАТЕРИАЛЫ И МЕТОДЫ

### A. Наборы данных

В отличие от наборов данных, составленных из снимков таких исследований лучевой диагностики как КТ, МРТ, ПЭТ КТ, наборы данных ОКТ-снимков имеют невысокое распространение. Особенно наборы данных, находящиеся в открытом доступе [6].

Анализ литературы показал наличие 8 наборов ОКТ-снимков для задачи классификации, находящихся в открытом доступе [7]. При этом для задач анализа фундус-снимков таких наборов в открытом доступе существует не менее 120 [8]. Эти наборы отличает схожесть категорий наименований классов и небольшие объёмы выборок классов. Все найденные наборы данных являются несбалансированными. Характеристики наборов данных представлены в табл. 1.

ТАБЛИЦА I. НАБОРЫ ДАННЫХ ОКТ-СНИМКОВ

Название	Год	Кол-во изображений	Наименования классов
Duke [9]	2015	110	DME
OPTIMA [10]	2016	30	IFR
UMN [11]	2017	600	Exudative AMD
Kermany [12]	2018	207130	CNV, DME, Drusen, NO
OCTID [13]	2018	500	MH, AMD, CSR, DR, NO
RETOUCH [14]	2019	70	AMD, RVO
HCMS [15]	2019	35	Healthy Controls, MS
OCTDL [7]	2023	2064	AMD, DME, ERM, NO, RAO, RVO, VID

Основными классами изображений, представленными в наборах данных табл. 1, являются: Age-related Macular Degeneration (AMD), Diabetic Macular Edema (DME), Epiretinal Membrane (ERM), Intra-retinal Fluid (IRF), Retinal Artery Occlusion (RAO), Retinal Vein Occlusion (RVO), Vitreomacular Interface Disease (VID), Macular Hole (MH), Normal (NO), Diabetic Retinopathy (DR). Схожесть категорий наименований классов, а по сути, наименований патологий исследуемой области отражает стандартизацию применения метода оптической когерентной томографии в диагностике конкретного пула заболеваний глаза [16].

Исходя из исходных характеристик наборов данных, для использования в данном исследовании был выбран OCTDL [7]. Выбранный набор данных является самым новым из всех опубликованных, был представлен Кулябиным и др. в 2023 году и состоит из 2064 изображений, разделённых на 7 классов изображений (AMD, DME, ERM, NO, RAO, RVO, VID). Изображения в наборе данных представляют собой ОКТ-снимки, полученные на нескольких томографах и имеющие высокое разрешение. Соотношение классов датасета представлено в табл. 2, а пример изображений на рис. 2.

ТАБЛИЦА II. ХАРАКТЕРИСТИКИ НАБОРА ДАННЫХ OCTDL

Название класса	Кол-во примеров, %
AMD	1231 / 59,6
DME	147 / 7,1
ERM	155 / 7,5
NO	332 / 16,1
RAO	22 / 1,1
RVO	101 / 4,9
VID	76 / 3,7
Всего	2064

Исходя из данных табл. 2, можно определить, что набор данных OCTDL является несбалансированным в связи с чем, для более корректного обучения модели и получения приемлемых значений метрик классификации необходимо провести его балансировку. В случае работы с наборами медицинских данных количество методов возможных аугментации ограничено. Применяемые методы не должны исказить изображения.

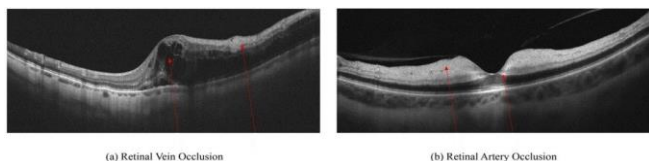


Рис. 1. Изображения из наборов данных OCTDL [7]

Для балансировки набора данных OCTDL применялись как методы аугментации (поворот изображений, изменение яркости снимков), так и дополнение такими же классами из других наборов данных. Количество изображений каждого класса в результате изменений было доведено до 600. При этом классы патологий, в отношении которых не могла быть проведена адекватная балансировка (Vitreomacular Interface Disease, Retinal Artery Occlusion), были исключены из набора. Характеристики получившегося после процедуры балансировки набора данных представлены в табл. 3.

ТАБЛИЦА III. СБАЛАНСИРОВАННЫЙ НАБОР ДАННЫХ

Название класса	Кол-во изображений / Кол-во добавленных изображений
AMD	1231 / -631
DME	147 / +453
ERM	155 / +445
NO	332 / +268
RAO	101 / +499
RVO	1231 / -631
VID	147 / +453
Всего	2400

### В. Архитектура ИНС

Механизм внимания как новая революционная идея в архитектурах ИНС была представлена Vaswani в 2017 году в работе [17]. Основная идея механизма внимания заключается в улучшении восприятия моделью контекстной информации. Для применения вместе с свёрточными нейронными сетями в 2018 году был предложен механизм пространственного внимания (англ. spatial attention module) [18]. В отличие от канального внимания, существовавшего ранее, фокусируется на информации о пространственном положении в расположении входящих данных, то есть пытается извлечь информацию из пространственного расположения пикселей.

Первым шагом в работе модуля является сжатие размерности пространства признаков  $F'$  относительно измерения  $C$ , затем на выходе происходит образование двух матриц  $F_{max}^s$  и  $F_{avg}^s \in \mathbb{R}^{H \times W}$  являющихся картами пространственного внимания, которые впоследствии конкатенируются. К полученному тензору размерности  $\mathbb{R}^{2 \times H \times W}$  применяется свёртка  $f^{m \times n}$ . Поэлементно к результату после прохождения свёртки применяется сигмоидальная функция активации. Происходит расчёт двумерных карт пространственного внимания  $M_s$ :

$$M_s(F) = \sigma(f^{m \times n}([\text{AvgPool}(F); \text{MaxPool}(F)])),$$

$$M_s(F) = \sigma(f^{m \times n}([F_{avg}^s; F_{max}^s])),$$

где  $\sigma$  – сигмоидальная функция. Архитектура модуля пространственного внимания представлена на рис. 2.

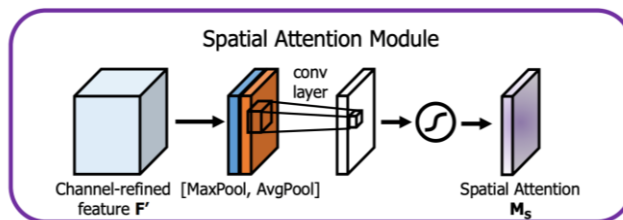


Рис. 2. Модуль пространственного внимания [18]

Модуль пространственного внимания был интегрирован в последовательную свёрточную нейронную сеть, состоящую из 20 слоёв (8 – 2-D Convolutional, 3 – MaxPooling, 4 – BatchNormalization, 1 – Dense, 1 – Softmax). Модуль внимания был интегрирован перед полносвязным слоем (рис. 3).

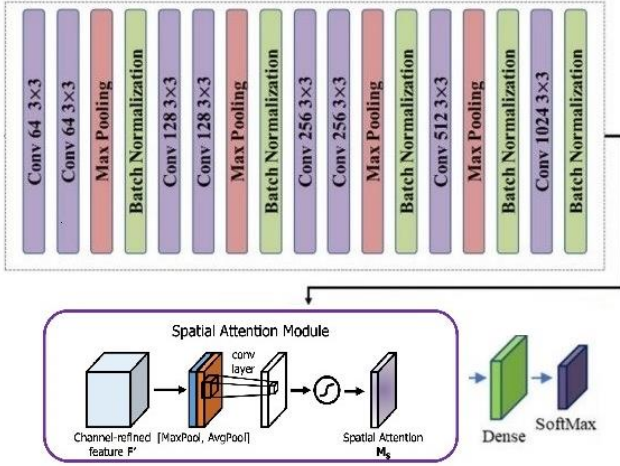


Рис. 3. Предложенная архитектура сети

Сравнение представленной на рис. 3 архитектуры ИНС производилось со свёрточным блоком данной сети. Параметры свёрточных слоёв не изменялись.

### C. Метод объяснительного ИИ

Применительно к поставленной задаче был выбран метод ОИИ Score-CAM, являющийся методом локального объяснения результатов работы непрозрачной модели – ИНС. Score-Weighted Visual Explanations for Convolutional Neural Networks (Score-CAM) был предложен Naofan Wang в 2020 году как обобщённая и усовершенствованная версия появившегося ранее метода CAM (Class Activation Mapping) [19] с модификациями, благодаря которым не требуется внесение изменений в архитектуру сети, что достигается избавлением от линейности в финальной части интерпретируемой модели [20]. Алгоритм получения визуального объяснения с помощью метода Score-CAM представлен на рис. 4.

#### Algorithm 1: Score-CAM algorithm

**Input:** Image  $X_0$ , Baseline Image  $X_b$ , Model  $f(X)$ , class  $c$ , layer  $l$

**Output:**  $L_{Score-CAM}^c$   
initialization;

// get activation of layer  $l$ ;

$M \leftarrow []$ ,  $A_l \leftarrow f_l(X)$

$C \leftarrow$  the number of channels in  $A_l$

**for**  $k$  in  $[0, \dots, C - 1]$  **do**

$M_l^k \leftarrow$  Upsample( $A_l^k$ )

    // normalize the activation map;

$M_l^k \leftarrow s(M_l^k)$

    // Hadamard product;

$M.append(M_l^k \circ X_0)$

**end**

$M \leftarrow$  Batchify( $M$ )

//  $f^c(\cdot)$  as the logit of class  $c$ ;

$S^c \leftarrow f^c(M) - f^c(X_b)$

// ensure  $\sum_k \alpha_k^c = 1$  in the implementation;

$\alpha_k^c \leftarrow \frac{\exp(S_k^c)}{\sum_k \exp(S_k^c)}$

$L_{Score-CAM}^c \leftarrow ReLU(\sum_k \alpha_k^c A_l^k)$

Рис. 4. Алгоритм Score-CAM [20]

### D. Метрики оценки качества

Метрики оценки качества работы ИНС вычисляются на основе значений матрицы ошибок: положительный-предсказано (TP), отрицательный-предсказано (FP), отрицательный-факт (FN), отрицательный-предсказано (FP). Данная матрица ошибок может быть изменена для задачи классификации на более чем два класса, тогда размерность матрицы будет соответствовать количеству классов.

Оценка качества мультиклассовой классификации, учитывая сбалансированность набора данных, проводилось с использованием модификаций метрик Recall, Precision, Accuracy, усреднённых относительно количества классов.

$$Recall_{avg} = \frac{Recall_A + Recall_B + Recall_N}{N},$$

$$Precision_{avg} = \frac{Precision_A + Precision_B + Precision_N}{N}$$

$$Accuracy_{avg} = \frac{\text{correct predictions}}{\text{all predictions}}.$$

## III. РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

Обучение, валидация и тестирование используемых моделей ИНС проводились, исходя из задачи, на разбитых на обучающую, валидационную, тестовую выборки в соотношении 70 %, 20 %, 10 %, собранную из модифицированного набора данных OCTDL. Результаты работы сетей на тестовой выборке, представленные в табл. 4, показывают прирост значений метрик после добавления модуля пространственного внимания к стандартной свёрточной архитектуре (recall +3 %, precision +5.8 %, accuracy +6 %). Значимым прирост является лишь на метриках precision и accuracy. Однако скорость обучения ИНС после добавления модуля внимания значительно снизилась. Так, при размере batch в 64 изображения, скорость обучения, в среднем, составила: 20 сек./эпоха для СНС и 50 сек./эпоха для СНС и модуль внимания. Обучение производилось на графическом процессоре NVIDIA GeForce RTX 4070.

ТАБЛИЦА IV. РЕЗУЛЬТАТЫ МУЛЬТИКЛАССОВОЙ КЛАССИФИКАЦИИ

Модель ИНС	Recall (%)	Precision	Accuracy (%)
Свёрточная	0,890	0,882	0,893
Свёрточная + модуль внимания	<b>0,921</b>	<b>0,940</b>	<b>0,953</b>

На рис. 5 представлено визуальное объяснение классов ОКТ-снимка (слева – нормальная сетчатка, справа – диабетический макулярный отёк), полученное с помощью метода Score-CAM. По интенсивности цвета тепловой карты можно увидеть области с наибольшим количеством релевантных признаков, которые оказали ключевое влияние на процесс принятия решения моделью ИНС. Для снимка без патологий (Normal, слева) – это вся область, кроме центральной «ямки», а для снимка с диабетическим макулярным отёком (DME,

справа) – это область с отёком, находящаяся по центру под «ямкой». Данный пример подтверждает корректность классификации патологии.

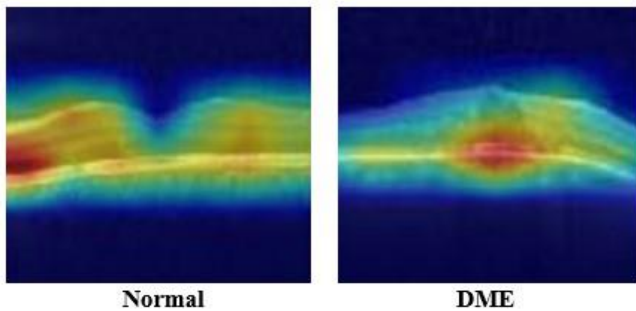


Рис. 5. Визуальное объяснение с помощью Score-CAM (собственные данные)

#### IV. ЗАКЛЮЧЕНИЕ

Таким образом, в данном исследовании представлен вариант интеграции модуля пространственного внимания в структуру свёрточной нейронной сети. Предположение об улучшении качества предсказания ИНС при интеграции в её состав модуля внимания подтверждено, получен статистически значимый прирост на метриках precision accuracy. Результаты работы ИНС: recall – 92 %, precision – 94 %, accuracy – 95 %, что свидетельствует о перспективах использования модифицированного механизма внимания в свёрточные нейронные сети. Однако стоит отметить, что при его интеграции увеличиваются требования к вычислительным мощностям для обучения ИНС. Корректность классификации оценивалась с использованием метода объяснительного ИИ Score-CAM. Использование методов объяснительного ИИ делает пайплайны обучения более прозрачными и, в перспективе, повышает доверие пользователя к конечному программному решению.

#### СПИСОК ЛИТЕРАТУРЫ

- [1] Saraswat D., Bhattacharya P., Verma A. et al. Explainable AI for healthcare 5.0: opportunities and challenges // *IEEE Access*. 2022. Vol. 10. P. 84486-84517. DOI: 10.1109/ACCESS.2022.3197671.
- [2] Ran A., Cheung C. Y. Deep learning-based optical coherence tomography and optical coherence tomography angiography image analysis: an updated summary // *Asia-Pacific Journal of Ophthalmology*. 2021. Vol. 10. no. 3. P. 253-260. DOI: 10.1097/APO.0000000000000405.
- [3] Leandro I. et al. OCT-based deep-learning models for the identification of retinal key signs // *Scientific Reports*. 2023. Vol. 13. no. 1. P. 14628. DOI: 10.1038/s41598-023-41362-4.
- [4] Averkin A.N., Volkov E.N., Yarushev S.A. Explainable Artificial Intelligence in Deep Learning Neural Nets-Based Digital Images Analysis // *Journal of Computer and Systems Sciences International*. 2024. V. 65. № 1. P. 1-29. DOI: 10.1134/S1064230724700138.
- [5] Volkov E.N., Averkin A.N. Possibilities of explainable artificial intelligence for glaucoma detection using the lime method as an example // *2023 XXVI International Conference on Soft Computing and Measurements (SCM)*. IEEE: Saint-Petersburg, 2023. P. 130-133. DOI: 10.1109/SCM58628.2023.10159038.
- [6] Khan S.M., Liu X., Nath S., et al. A global review of publicly available datasets for ophthalmological imaging: barriers to access, usability, and generalisability. *Lancet Digital Health*. 2020. DOI: 10.1016/S2589-7500(20)30240-5.
- [7] Kulyabin M., Zhdanov A., Nikiforova A. et al. Octdl: Optical coherence tomography dataset for image-based deep learning methods // *Scientific Data*. 2024. Vol. 11. no. 1. P. 365.
- [8] Krzywicki T., Brona P., Zbrzezny A.M. et al. A global review of publicly available datasets containing fundus images: characteristics, barriers to access, usability, and generalizability // *Journal of Clinical Medicine*. 2023. Vol. 12. no. 10. P. 3587. DOI: 10.3390/jcm12103587.
- [9] Chiu S.J., Allingham M.J., Mettu P.S. et al. Kernel regression-based segmentation of optical coherence tomography images with diabetic macular edema // *Biomedical optics express*. 2015. Vol. 6. no. 4. P. 1172-1194. DOI: 10.1364/boe.6.001172.
- [10] Wu J., Philip A.M., Podkowinski D. et al. Multivendor spectral-domain optical coherence tomography dataset, observer annotation performance evaluation, and standardized evaluation framework for intraretinal cystoid fluid segmentation // *Journal of Ophthalmology*. 2016. Vol. 2016. no. 1. P. 3898750. DOI: 10.1155/2016/3898750.
- [11] Rashno A., Koozekanani D.D., Drayna P.M. et al. Fully automated segmentation of fluid/cyst regions in optical coherence tomography images with diabetic macular edema using neutrosophic sets and graph algorithms // *IEEE Transactions on Biomedical Engineering*. 2017. Vol. 65. no. 5. P. 989-1001. DOI: 10.1109/tbme.2017.2734058.
- [12] Kermay D.S., Goldbaum M., Cai W. et al. Identifying medical diagnoses and treatable diseases by image-based deep learning // *cell*. 2018. Vol. 172. no. 5. P. 1122-1131. e9. DOI: 10.1016/j.cell.2018.02.010.
- [13] Gholami P., Roy P., Parthasarathy M.K. et al. OCTID: Optical coherence tomography image database // *Computers & Electrical Engineering*. 2020. Vol. 81. P. 106532. DOI: 10.1016/j.compeleceng.2019.106532.
- [14] Bogunovic H., Venhuizen F., Klimscha S. et al. RETOUCH: The retinal OCT fluid detection and segmentation benchmark and challenge // *IEEE transactions on medical imaging*. 2019. Vol. 38. no. 8. P. 1858-1874. DOI: 10.1109/tmi.2019.2901398.
- [15] He Y., Carass A., Solomon S.D. et al. Retinal layer parcellation of optical coherence tomography images: Data resource for multiple sclerosis and healthy controls // *Data in brief*. 2019. Vol. 22. P. 601-604. DOI: 10.1016/j.dib.2018.12.073.
- [16] Viedma I.A., Alonso-Caneiro D., Read S.A. et al. Deep learning in retinal optical coherence tomography (OCT): A comprehensive survey // *Neurocomputing*. 2022. Vol. 507. P. 247-264. DOI: 10.1016/j.neucom.2022.08.021
- [17] Vaswani A., Shazeer N., Parmar N. et al. Attention is all you need // *Advances in neural information processing systems*. 2017. Vol. 30.
- [18] Woo S., Park J., Lee J.-Y. et al. Cbam: Convolutional block attention module // *Proceedings of the European conference on computer vision (ECCV)*. 2018. P. 3-19.
- [19] Zhou B., Khosla A., Lapedriza A. et al. Learning deep features for discriminative localization // *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016. P. 2921-2929.
- [20] Wang H., Wang Z., Du M. et al. Score-CAM: Score-weighted visual explanations for convolutional neural networks // *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*. 2020. P. 24-25.

# Перспективы использования нейронных сетей для диагностики заболеваний сердечно-сосудистой системы

Е. А. Соколова

Лаборатория физиологии биоуправления  
ОЭФ ФГБНУ «Институт экспериментальной медицины»

evgenyaagent@gmail.com

**Аннотация.** В статье рассматриваются современные исследования, направленные на использование искусственных нейронных сетей в задачах диагностики сердечно-сосудистых заболеваний. Рассмотрены основные способы применения нейронных сетей для анализа диагностических данных (электрокардиографических, гемодинамических и других). Описаны перспективы и основные трудности, возникающие при создании диагностических систем на основе нейронных сетей.

**Ключевые слова:** нейронные сети; диагностика; сердечно-сосудистая система; сердечно-сосудистые заболевания

## I. ВВЕДЕНИЕ

Несмотря на стремительное развитие технологий, направленных на повышение качества диагностики заболеваний сердечно-сосудистой системы, они все еще занимают ведущие места по распространенности среди населения. Согласно исследованию 2022 года, одним из осложнений ишемической болезни сердца является инфаркт миокарда, вероятность возникновения которого увеличивается с возрастом [1]. Поэтому вопрос разработки эффективных средств ранней диагностики является наиболее актуальным.

Искусственные нейронные сети (ИНС) являются современным инструментом, который уже сейчас активно применяется в различных отраслях деятельности человека. Не удивительно, что они также начали проникать и в область здравоохранения. Поскольку они способны решать широкий круг задач, таких, как классификация, кластеризация, сегментация, прогнозирование и других. Глубокое обучение демонстрирует высокую эффективность для здравоохранения, например, в анализе медицинских изображений. Однако, его использование требует больших объемов данных, а, следовательно, такие системы требуют больших вычислительных мощностей, что, в свою очередь, может снижать возможности применения нейросетевых моделей на машинах с ограниченной вычислительной мощностью [2].

Тем не менее, ИНС активно применяются в сфере здравоохранения, и пройдет сравнительно немного времени до того момента, когда технологии, основанные на искусственном интеллекте (ИИ), будут внедряться в медицине повсеместно.

## II. НЕЙРОННЫЕ СЕТИ В ДИАГНОСТИКЕ

В частности, искусственные нейронные сети могут помочь в диагностике различных заболеваний сердечно-сосудистой системы. Одним из таких заболеваний является сердечный амилоидоз, который характеризуется внеклеточным отложением амилоидных фибрилл. Как было показано Kamel M.A. с соавторами (2024), сверточные нейронные сети, могли бы выявлять пациентов, подверженных высокому риску амилоидоза, и тем самым способствовать ранней диагностике этого заболевания [3].

Нейронные сети могут определять наличие заболеваний сердечно-сосудистой системы на основании анализа снимков магнитно-резонансной (МРТ) и компьютерной томографии (КТ), результатов эхо кардиографии, ядерной кардиологии и рентгенологических исследований [4]. Например, в исследовании 2018 года Yakun Chang с коллегами использовали сеть YOLO для осуществления сегментации и полносвязную нейронную сеть для классификации кардиопатии на основании МРТ снимков сердца [5]. Результат работы показал, что нейронная сеть может успешно справиться с сегментацией миокарда, левого и правого желудочка, а также достигает высокой точности при классификации заболеваний сердца.

Рассмотрим более подробно применение нейронных сетей в диагностике заболеваний сердечно-сосудистой системы.

### A. Врожденный порок сердца

Нейронные сети можно применять для распознавания врожденных пороков сердца. Так, Md Hassanuzzaman с коллегами использовали сигналы фонокардиограммы (ФКГ) и нейронные сети на базе архитектуры Трансформер (Transformer) для выявления врожденных пороков сердца у детей [6]. Авторами была предложена модель глубокого обучения, которая способна осуществлять классификацию как обработанных, так и не обработанных сигналов ФКГ. Построенная модель показала хорошее качество работы и может использоваться для более быстрого и полного распознавания и анализа сигналов ФКГ.

В исследовании [7] был предложен новый метод для диагностики и обнаружения врожденных пороков сердца у беременных женщин, основанный на



машинном обучении. Prabu Pachiyannan с коллегами используют метод прогнозирования врожденных пороков сердца на основе машинного обучения (Machine Learning-based Congenital Heart Disease Prediction Method (ML-CHDPM)). Указанная модель показала довольно высокую точность и низкий уровень ложноположительных результатов.

#### *В. Инфаркт миокарда*

Искусственные нейронные сети могут также найти применение для автоматического обнаружения и локализации инфаркта миокарда [8], что очень важно, поскольку такое заболевание требует быстрой и точной диагностики. В исследовании [9] нейронные сети использовались для обнаружения и классификации миокарда на кадрах эхокардиограмм. Авторы реализовали усовершенствованный алгоритм сверточной нейронной сети и сети ECV-3D. После проведенных экспериментов предложенный авторами подход достиг высокой точности (97,05%) при обнаружении и классификации инфарктов миокарда.

#### *С. Аритмия сердца*

Еще одной областью применения нейронных сетей, может являться определение наличия аритмии у человека. Расшифровка электрокардиограммы (ЭКГ) с помощью ИНС позволяет дифференцировать наличие синусового ритма и мерцательной аритмии [10]. В исследовании 2024 года, Nelson Lu с соавторами предлагают систему искусственного интеллекта для обнаружения фибрилляции предсердий по результатам анализа эхокардиограмм без проведения ЭКГ [11]. Применение такой модели показало точность диагностики более 90%.

В исследовании 2020 года авторы использовали модель глубокой нейронной сети для обнаружения сердечной аритмии по записям ЭКГ [12]. В результате обучения модели на наборе данных, содержащем более 10 000 записей, ИНС показала высокую точность для двух классов сценариев, включающих в себя семь или четырех типов ритмов, соответственно.

#### *Д. Оценка риска*

Нейронные сети можно использовать для оценки риска возникновения сердечно-сосудистых событий. Так, например, в исследовании 2024 года [13] авторы применяли глубокие сверточные нейронные сети для оценки визуализации перфузии миокарда и для долгосрочного прогнозирования рисков наступления сердечно-сосудистого инцидента. Кроме того, нейронные сети использовались для прогнозирования на основании КТ грудной клетки [14]. При этом автоматизированная оценка сердечно-сосудистых событий на основании кальцификации коронарных артерий показала высокую корреляцию с ручной количественной оценкой. Таким образом, демонстрируется клиническая ценность реализации систем на основе нейронных сетей для прогнозирования сердечно-сосудистых заболеваний.

Поскольку артериальное давление (АД) играет огромную роль в диагностике заболеваний сердечно-сосудистой системы, очень важно регулярно отслеживать этот показатель. Нейронные сети могут использоваться для мониторинга здоровья спортсменов

[15], суть которого заключается в отслеживании изменений АД и пульса во время и после тренировки, последующей оценке воздействия спортивной нагрузки на сердце и предоставлении рекомендаций по тренировочному режиму.

### III. ТРУДНОСТИ ПРИ СОЗДАНИИ ДИАГНОСТИЧЕСКИХ СИСТЕМ С ИСКУССТВЕННЫМ ИНТЕЛЛЕКТОМ

Одной из ключевых проблем при обучении ИНС для задачи диагностики можно назвать подготовку данных, поскольку нейронные сети для своего обучения требуют большой объем информации. Существует большое число болезней сердечно-сосудистой системы, соответственно сбор такого объема данных может вызывать огромные трудности. По этой причине ИНС могут быть сильно ограничены в возможностях своего обучения, что способно негативно сказаться на точности осуществляемой ими диагностике.

Кроме того, для обучения моделей требуются машины с высокой вычислительной мощностью, которые имеют большую стоимость, и их приобретение может быть проблематичным для лабораторий, которые планируют внедрять системы искусственного интеллекта в свою работу.

Еще одной причиной, вызывающей трудности в создании диагностических систем на основе ИНС может являться то, что хоть они и способны работать с высоким уровнем точности при выполнении поставленных перед ними задач, они, тем не менее, могут ошибаться, поэтому возникает закономерный вопрос, кто будет нести ответственность в случае неправильной постановки диагноза, повлекшей за собой отрицательные последствия для пациента. Таким образом, принципиальной задачей является распределение ответственности между разработчиком системы на базе ИИ и врачом [16]. Кроме того, возникает вопрос "предвзятости" таких систем, поскольку ИНС могут быть ограничены в вариативности информации, используемой при их обучении, и следовательно, они могут плохо диагностировать тех пациентов, чьи медицинские параметры являются не стандартными.

### IV. ПЕРСПЕКТИВЫ

Современные тенденции показывают, что технологии искусственного интеллекта уже не являются чем-то фантастическим. Нейронные сети показали свои способности в диагностике множества заболеваний сердечно-сосудистой системы (ишемической болезни сердца, сердечной недостаточности, кардиомиопатии, врожденного порока, мерцательной аритмии). Создаются коммерческие системы, построенные на основе искусственного интеллекта, которые призваны помогать специалистам здравоохранения. Например, компания «СберМедИИ», которая занимается разработкой решений на базе искусственного интеллекта для здравоохранения. Одним из таких решений является диагностический ассистент «AIDA», который призван на основании записей из медицинской карты пациента, помочь диагностировать заболевания [17]. Кроме того, компания предлагает медицинское оборудование со встроенным искусственным интеллектом: MDDC Кардио предлагает мобильное оборудование способное

осуществлять расшифровку записей ЭКГ и проводить функциональную диагностику. Используя такое оборудование можно повысить точность и качество диагностики записей ЭКГ [18].

Кроме того, благодаря системам на основе ИИС, можно будет выявлять людей из группы риска еще в молодом возрасте, тем самым начать профилактику возникновения заболеваний сердечно-сосудистой системы. Также это может способствовать индивидуальной оценке рисков развития сердечно-сосудистых заболеваний у пациентов, поскольку будет учитывать особенности организма отдельно взятого пациента [19]. Благодаря чему, такая система может использоваться для поддержки принятия решений, например помочь врачу выбрать наиболее эффективный метод лечения для конкретного пациента.

#### V. ЗАКЛЮЧЕНИЕ

Подводя итог, следует сказать, развитие искусственного интеллекта позволит значительно повысить качество диагностики сердечно-сосудистых заболеваний, что по-прежнему является актуальной проблемой, и, соответственно, повысить качество жизни людей. В настоящее время результаты применения систем на базе ИИ носят рекомендательный характер. Однако, учитывая темпы развития технологий, можно с уверенностью сказать, что повсеместное внедрение систем искусственного интеллекта в клиническую практику является лишь вопросом времени. В связи с чем требуется параллельная разработка мер по защите пациентов от последствия возможных ошибок в работе ИИ.

#### СПИСОК ЛИТЕРАТУРЫ

- [1] Шальнова С.А., Драпкина О.М., Куценко В.А., Капустина А.В., Муромцева Г.А., Яровая Е.Б., Баланова Ю.А., Евстифеева С.Е., Имаева А.Э., Шляхто Е.В., Бойцов С.А., Астахова З.Т., Барбараш О.Л., Белова О.А., Гринштейн Ю.И., Ефанов А.Ю., Калачикова О.Н., Кулакова Н.В., Недогода С.В., Ротарь О.П., Трубочева И.А., Черных от имени участников исследования ЭССЕ-РФ Т.М. Инфаркт миокарда в популяции некоторых регионов России и его прогностическое значение // Российский кардиологический журнал. 2022. Т. 27. № 6. С:4952.
- [2] Alzubaidi L, Zhang J, Humaidi AJ, et al. Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. *J Big Data*. 2021;8(1):53.
- [3] Kamel MA, Abbas MT, Kanaan CN, Awad KA, Baba Ali N, Scalia IG, Farina JM, Pereyra M, Mahmoud AK, Steidley DE, et al. How Artificial Intelligence Can Enhance the Diagnosis of Cardiac Amyloidosis: A Review of Recent Advances and Challenges. *Journal of Cardiovascular Development and Disease*. 2024; 11(4):118.
- [4] Аббасов И.Б., Дешмух Р.Р. Применение искусственного интеллекта для медицинской визуализации // Международный научно-исследовательский журнал. 2021. Т. 12. № 114, С. 43-49.
- [5] Y. Chang, B. Song, C. Jung and L. Huang, Automatic Segmentation and Cardiopathy Classification in Cardiac Mri Images Based on Deep Neural Networks. 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). 2018, P. 1020-1024.
- [6] Hassanuzzaman M, Hasan NA, Mamun MAA, et al. Recognition of Pediatric Congenital Heart Diseases by Using Phonocardiogram Signals and Transformer-Based Neural Networks. *Annu Int Conf IEEE Eng Med Biol Soc*. 2023;2023:1-4.
- [7] Pachiyannan P, Alsulami M, Alsadie D, Saudagar AKJ, AlKhatami M, Poonia RC. A Novel Machine Learning-Based Prediction Method for Early Detection and Diagnosis of Congenital Heart Disease Using ECG Signal Processing. *Technologies*. 2024; 12(1):4.
- [8] Qiang Y., Dong X., Yang Y. Automatic detection and localisation of myocardial infarction using multi-channel dense attention neural network. *Biomedical Signal Processing and Control*. 2024. Vol. 89:105766.
- [9] Deepika S., Jaisankar N. Detecting and Classifying Myocardial Infarction in Echocardiogram Frames With an Enhanced CNN Algorithm and ECV-3D Network // *IEEE Access*. 2024. Vol. 12:51690–51703.
- [10] Родионов Д. М., Карчков Д.А., Москаленко В. А., Никольский А. В., Осипов Г. В., Золотых Н. Ю. Диагностика синусового ритма и мерцательной аритмии средствами искусственного интеллекта // "Проблемы информатики", 2022, № 1, с.77-88.
- [11] Lu N, Vaseli H, Mahdavi M, et al. Automated Atrial Fibrillation Diagnosis by Echocardiography without ECG: Accuracy and Applications of a New Deep Learning Approach. *Diseases*. 2024;12(2):35.
- [12] Yildirim O, Talo M, Ciaccio EJ, Tan RS, Acharya UR. Accurate deep neural network model to detect cardiac arrhythmia on more than 10,000 individual subject ECG records. *Comput Methods Programs Biomed*. 2020;197:105740.
- [13] Li, YL., Leu, HB., Ting, CH. et al. Predicting long-term time to cardiovascular incidents using myocardial perfusion imaging and deep convolutional neural networks. *Sci Rep*. 2024. 14:3802.
- [14] Zeleznik, R., Foldyna, B., Eslami, P. et al. Deep convolutional neural networks to predict cardiovascular risk from computed tomography. *Nat Commun*. 2021. 12:715.
- [15] Исмаилов О.М., Мирзахалилов С.С., Холдарова Г.Н. Методы применения алгоритмов искусственного интеллекта в мониторинге здоровья спортсменов во время тренировок и соревнований // *Science and innovation*. 2023. 2(3):697–702.
- [16] Третьякова Е.П. Использование искусственного интеллекта в здравоохранении: распределение ответственности и рисков // *Цифровое право*. 2021. Т. 2, № 4. СС 51–60.
- [17] Диагностический ассистент AIDA (AI Diagnostics Assistant), URL: <https://sbermed.ai/our-algorithms/aida> (дата обращения: 28.04.2024)
- [18] MDDC Кардио, URL: <https://sbermed.ai/meditsinskoye-oborudovaniye-s-ii/mddc-cardio> (дата обращения: 28.04.2024).
- [19] Гусев А.В., Гаврилов Д.В., Корсаков И.Н., Кузнецова Т.Ю., Серова Л.М., Новицкий Р.Э. Перспективы использования методов машинного обучения для предсказания сердечно-сосудистых заболеваний // *Врач и информационные технологии*. 2019. № 3. С. 41-47.

# Применение U-Net для выделения рядов виноградника по данным дистанционного зондирования Земли

С. А. Шквиро, М. Д. Поляк

Санкт-Петербургский государственный университет аэрокосмического приборостроения

markpolyak@gmail.com

**Аннотация.** Одним из ключевых шагов в управлении сельскохозяйственными угодьями на сегодняшний день является детекция объектов по данным дистанционного зондирования Земли. Доступность спутниковых снимков с высоким разрешением привела к появлению новых методов классификации и анализа изображений. Целью данного исследования является оценка эффективности алгоритмов машинного обучения для определения урожайности виноградника. Предлагается использовать нейронную сеть с архитектурой U-Net для извлечения информации о рядах виноградника из спутниковых снимков, полученных с картографического сервиса Bing Maps. Обученная нейросетевая модель имеет точность 90 %.

**Ключевые слова:** виноградарство; U-Net; ДЗЗ; классификация

## I. ВВЕДЕНИЕ

Ряды виноградника являются важнейшим элементом виноградного угодья. Обнаружение рядов виноградника на спутниковых снимках либо снимках, полученных с помощью БПЛА, имеет первостепенное значение для оценки урожайности.

В рамках рассматриваемой задачи извлечение информации о рядах виноградника из данных ДЗЗ является основной целью анализа изображений с помощью методов компьютерного зрения. Ряды виноградников имеют выделяющуюся структуру на фоне всего сельскохозяйственного угодья. Таким образом, они могут быть обнаружены на снимках и низкого и высокого разрешения. Из-за быстрых изменений внешнего вида во время роста кустов винограда необходимо своевременно оценивать их характеристики. Например, болезнью Пирса [1] может поразить растения и их цвет на изображении ДЗЗ изменится. В такие моменты, в первую очередь, необходимо приступить к исправлению сложившейся ситуации для сохранения урожая. Разработка подхода для получения информации в реальном времени будет полезна для управления аграрным хозяйством, а также для оптимизации экономической составляющей виноградарства и роста объёма отечественных вин [2]. Это позволит аграрным предприятиям принимать обоснованные решения на основе актуальных данных о состоянии виноградников. Благодаря своевременной информации о росте растений, погодных условиях и других факторах, связанных с

выращиванием винограда, сельскохозяйственные предприятия смогут оптимизировать процессы ведения хозяйства, включая планирование урожая, управление ресурсами и распределение рабочей силы. Кроме того, это также может сыграть важную роль при привлечении инвестиций, взятии кредитов и заключении контрактов на продажу урожая. Кусты виноградника высаживают в ряды через равные промежутки: для сильнорослых сортов – 2,5 м, среднерослых – 2, слаборослых – 1,5 м. [3].

Для решения задачи обнаружения рядов виноградника в данной работе предлагается использовать подход, применяющийся для детекции дорог по спутниковым снимкам [4]. Дороги, как и ряды виноградника, имеют ярко выраженную структуру, хорошо заметную на изображениях ДЗЗ. Таким образом, методы, разработанные для поиска дорог на спутниковых снимках, могут быть адаптированы и применены к обнаружению рядов виноградника.

Цель данного исследования – построить модель для выделения рядов виноградника по спутниковым снимкам.

## II. МОДЕЛЬ U-NET

Архитектура U-Net является мощным инструментом для сегментации изображений. Она успешно применяется в различных областях, включая обработку медицинских изображений, в сельском хозяйстве и др.

Изначально архитектура была разработана для сегментации биомедицинских изображений [5]. Она была адаптирована для разных задач, таких как поиск дорог на изображении [4], локализации нарушений целостности цифровых изображений [6], распознавании растительности [7] и многих других.

Рассматриваемая нейросетевая модель имеет основные элементы классической архитектуры U-Net, как показано на рис. 1. Она состоит из двух основных блоков: кодировщика и декодировщика. В кодировщике изображение последовательно пропускается через сверточные слои, чтобы извлечь информативные признаки на различных уровнях абстракции. В декодировщике происходит обратная операция для восстановления изображения из полученных ранее признаков.

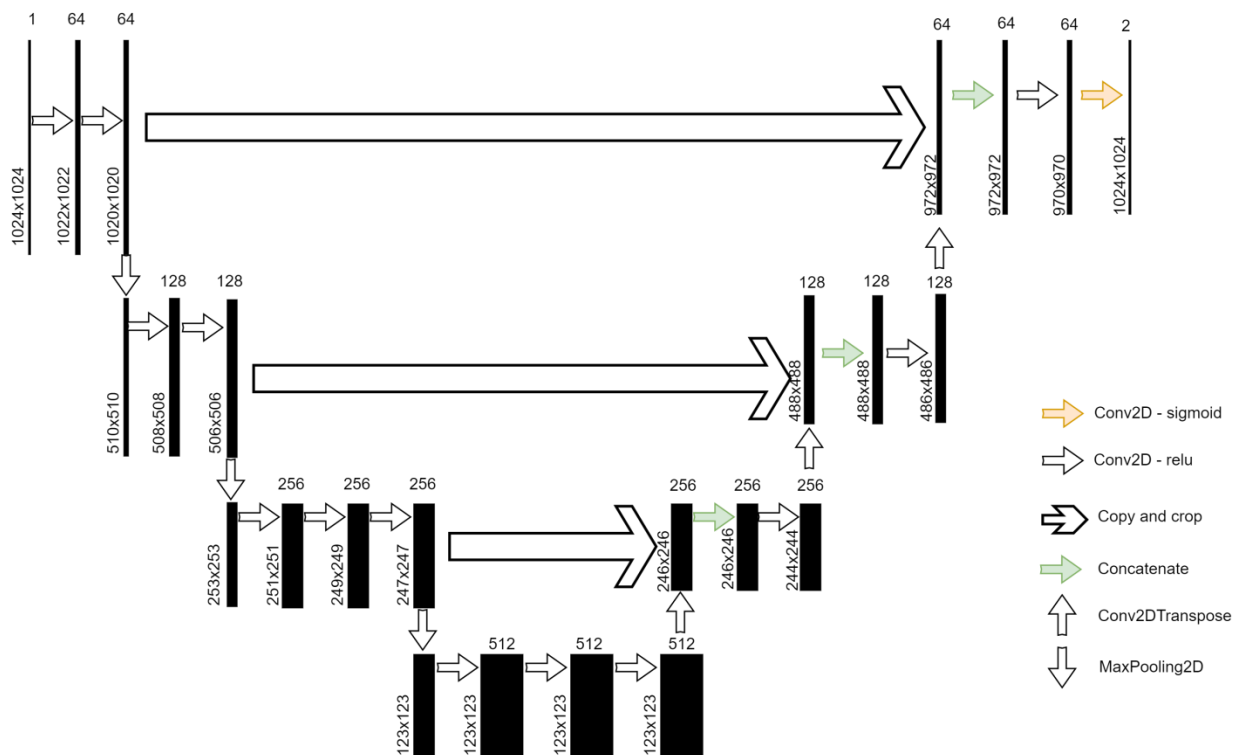
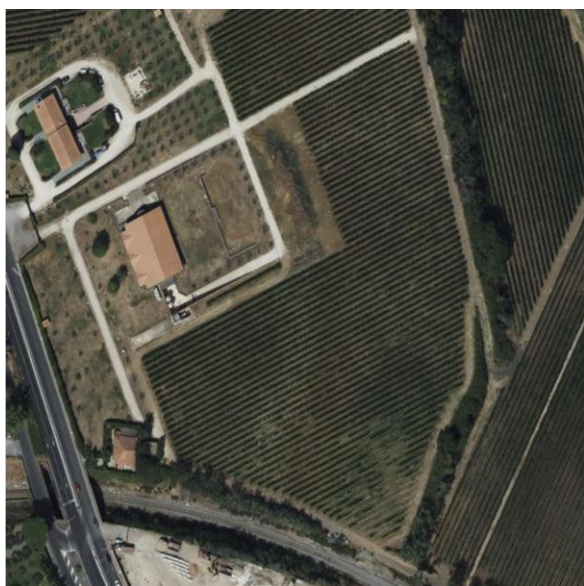
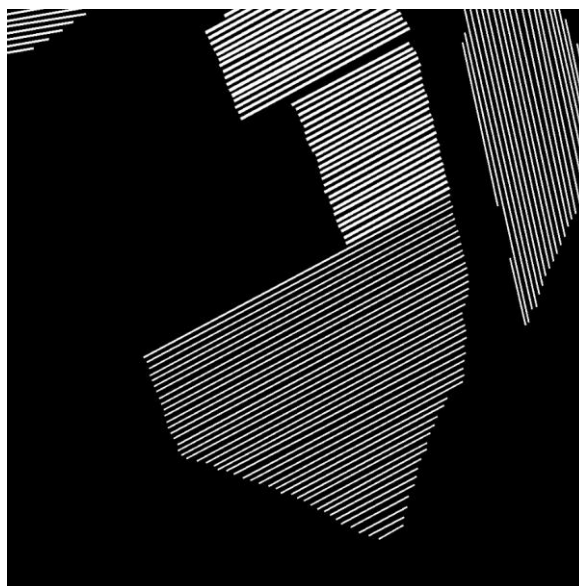


Рис. 1. Схема нейросетевой модели на основе архитектуры U-Net



а) аэрофотоснимок виноградника (Италия)



б) размеченные ряды виноградника

Рис. 2. Пример аэрофотоснимка виноградника из обучающей выборки и результат разметки рядов

### III. ПОДГОТОВКА НАБОРА ДАННЫХ

По данным картографического сервиса OpenStreetMaps, который содержит введённую пользователями информацию о местоположении и типах сельскохозяйственных земель, были отобраны виноградники, расположенные на территории Италии. Данная страна была выбрана, поскольку согласно исследованиям [8, 9] занимает первое место в мире по производству вина. Координаты отобранных виноградников были использованы для получения несколько тысяч спутниковых снимков, из которых

вручную были отобраны 108 для составления датасета. На попавших в датасет изображениях были вручную размечены ряды виноградников. Размер изображений в датасете  $1024 \times 1024$  пикселей.

Набор данных разбит на обучающую выборку из 93 изображений, тестовую из 8 и валидационную из 7. Пример исходного спутникового изображения виноградника и результат его ручной разметки показаны на рис. 2а и 2б.

## IV. ОПИСАНИЕ МЕТРИК

Для оценки качества обучения были использованы метрики: accuracy, precision, recall, F1-score.

Метрика recall вычисляется по формуле:

$$recall = \frac{TP}{PPos + e};$$

Метрика recall измеряет долю истинно положительных результатов (TP) относительно всех возможных положительных случаев (PPos) в данных, с учетом эpsilon (e) для избежание деления на ноль.

Метрика precision вычисляется по формуле:

$$precision = \frac{TP}{PP + e};$$

Метрика precision измеряет долю истинно положительных результатов (TP) относительно всех предсказанных положительных случаев (PP), с учетом эpsilon (e) для избежание деления на ноль.

Метрика F1-score вычисляется по формуле:

$$F1\_score = 2 * \frac{precision * recall}{precision + recall};$$

Данная метрика представляет собой среднее гармоническое между полнотой (recall) и точностью (precision).

## V. РЕЗУЛЬТАТ ОБУЧЕНИЯ НЕЙРОННОЙ СЕТИ

Максимальные значения метрик точности и F1-score на обучающей выборке достигаются на 100-й эпохе, как видно на рис. 3. Однако, максимальные значения на валидационной выборке достигаются на 46 эпохе: F1\_score = 0.69, Accuracy = 0.9.

Следовательно, модель нужно обучать только до 46 эпохи, так как дальше обучение неэффективно и происходит переобучение. Значения метрик точности и F1-score обученной в течение 46 эпох модели на тестовой выборке представлены на рис. 4. Минимальное значение точности равно 0.83, а F1 – 0.31.

Результат обработки одного из спутниковых снимков виноградника из тестовой выборки с помощью обученной модели показан на рис. 5.

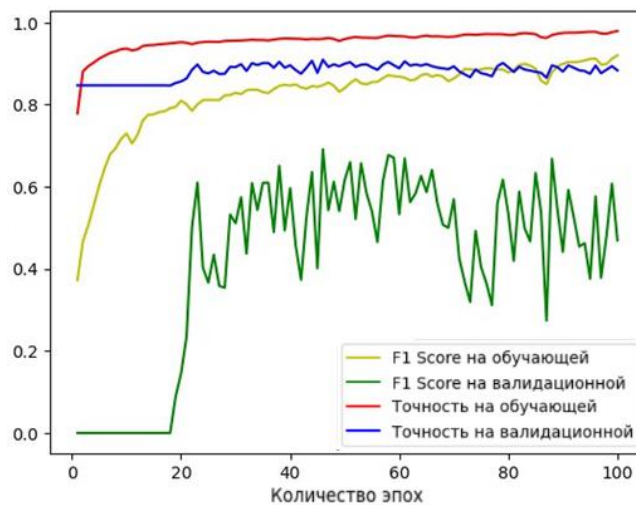


Рис. 3. Результат обучения нейронной сети

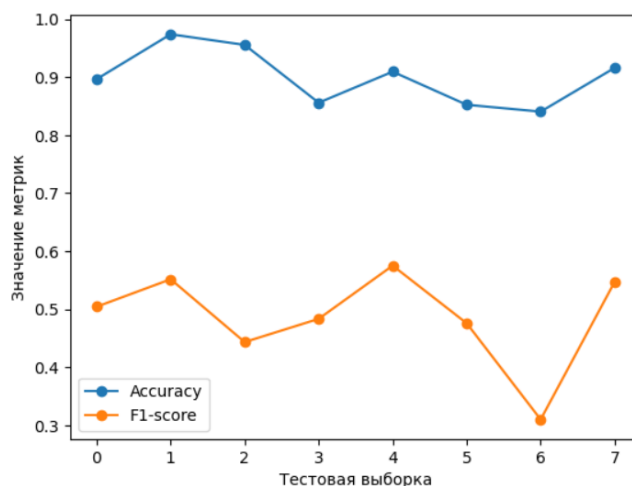
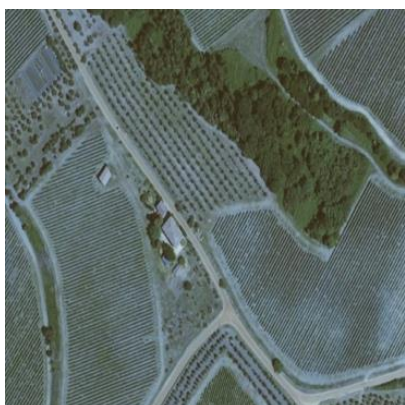


Рис. 4. Значения метрик на изображениях тестовой выборки



а) изображение из тестовой выборки



б) ручная разметка рядов виноградника



в) предсказание модели U-Net по исходному изображению

Рис. 5. Результат работы нейросетевой модели

## VI. ЗАКЛЮЧЕНИЕ

Построенная модель демонстрирует высокую эффективность в выделении рядов виноградников на спутниковых снимках. Результаты показывают, что предложенный подход успешно справляется с задачей сегментации и позволяет точно определить расположение виноградных рядов.

Полученные результаты могут быть использованы в областях, связанных с виноградным хозяйством и сельским хозяйством в целом.

Будущие исследования могут быть направлены на расширение обучающего набора данных, улучшение алгоритмов обработки изображений.

## СПИСОК ЛИТЕРАТУРЫ

- [1] Арестова Н.О., Рябчун И.О. Распространенность бактериальных болезней винограда в агроценозе Ростовской области // Плодоводство и виноградарство Юга России, 2020. 293-311 с. DOI: 10.30679/2219-5335-2020-4-64-293-311
- [2] Гагерова А.В., Гармашова Е.П. Обзор рынка виноградарства и виноделия Российской Федерации на фоне общемировых тенденций // Экономика. Информатика, 2020. 501-511 с. DOI:10.18413/2687-0932-2020-43-3-501-511
- [3] Смирнов, К.В. Виноградарство / под ред. К.В. Смирнова. Изд. МСХА, 1998. 511 с.
- [4] Abderrahim N.Y.Q., Abderrahim S., Rida A. Road segmentation using u-net architecture [2020 IEEE international conference of moroccan geomatics (Morgeo)]. Casablanca, Morocco, IEEE, 2020. с. 1-4.
- [5] Ronneberger O., Fischer P., Brox T. U-net: Convolutional networks for biomedical image segmentation [Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18]. Springer International Publishing, 2015. с. 234-241.
- [6] Абдуллина К.М., Спивак А.И. Применение нейронной сети архитектуры U-Net для задачи локализации нарушений целостности цифровых изображений // Научно-технический вестник информационных технологий, механики и оптики, 2020. 3 (127), с. 425-431.
- [7] Ганченко В.В., Дудкин А.А., Шелег С.В. Распознавание сельскохозяйственной растительности на изображениях земной поверхности на основе сверточной нейронной сети U-Net, 2021.
- [8] La produzione di vino nel mondo 2018 – prima stima OIV. Доступно по адресу: <http://www.inumeridelvino.it/2018/11/la-produzione-di-vino-nel-mondo-2018-prima-stima-oiv.html> (дата обращения 26.04.2024).
- [9] Italia prima assoluta per vitigni autoctoni, ecco i dati dei vari stati. Доступно по адресу: <https://giornalevinocibo.com/2017/06/03/italia-prima-assoluta-per-vitigni-autoctoni-ecco-i-dati-dei-vari-stati/> (дата обращения 26.04.2024).

# Глубокое обучение для всестороннего улучшения состояния сетчатки: помимо автоматизированной фундоскопии – классификация, сегментация и реконструкция изображения сетчатки

Ali Sultan Maaya

Санкт-Петербургский государственный электротехнический университет  
«ЛЭТИ» им. В.И. Ульянова (Ленина)

Alimayya1357@gmail.com

**Аннотация.** В этом исследовании представлен подход к глубокому обучению, направленный на расширение сферы анализа состояния сетчатки за рамки стандартной фундоскопии, с акцентом на задачи классификации и сегментации. Результаты свидетельствуют об эффективности таких моделей, как Inception, ResNet и Mobile Net, в точной классификации особенностей сетчатки, при этом Inception демонстрирует превосходную производительность (точность: 96 %, прецизионность: 90 %). Несмотря на то, что точность MobileNet несколько ниже, ее легкая архитектура оказывается подходящей для приложений реального времени, таких как фундоскопия смартфона (точность: 91 %). Кроме того, в исследовании предлагается использовать методы сегментации, основанные на глубоком обучении, такие как GAN, автоэнкодеры и U-Net, для решения проблем, связанных с сегментацией наборов данных, без указания достоверности. Эти результаты вносят значительный вклад в развитие автоматизированного анализа изображений сетчатки для эффективной диагностики глазных заболеваний, предоставляя ценную информацию для будущих исследований и клинического применения.

**Ключевые слова:** анализ состояния сетчатки, автоматизированная фундоскопия, глубокое обучение, классификация, трансферное обучение, сравнительный анализ, сегментация под наблюдением, генеративные состязательные сети (gans)

## I. ВВЕДЕНИЕ

Традиционная фундоскопия имеет свои ограничения, особенно в таких специализированных областях, как офтальмология. По данным Всемирной организации здравоохранения, во всем мире насчитывается около 2,2 млрд человек с нарушениями зрения/слепотой, из которых по меньшей мере 1 млрд поддается профилактике или лечению при раннем выявлении и вмешательстве [1]. Кроме того, традиционные офтальмологические осмотры требуют физического присутствия пациента у офтальмолога, что может быть проблемой для многих людей – лиц, проживающих в отдаленных районах с ограниченным доступом к

специалистам (особенно в развивающихся странах), или лиц с ограниченной мобильностью. С другой стороны, носимые биотехнологические устройства открывают новые горизонты для телемедицины, позволяя дистанционно оценивать состояние здоровья пациента. Одним из наиболее перспективных направлений применения является разработка носимых устройств для анализа глазного дна. [2]

Смартфонная фундоскопия, это прорыв в диагностике глазных заболеваний, поскольку камеры смартфонов используются для исследования глазного дна, включая сетчатку. Однако к недостаткам Вышеуказанного метода относятся низкое качество изображения, возможные диагностические ошибки и необходимость привлечения квалифицированных специалистов. Для решения этой проблемы предлагается носимый Аппаратно-Программный комплекс (НАПК) для автоматизированного обследования сетчатки. Данное устройство автоматизирует фундоскопию с помощью смартфона, упрощая процесс и позволяя автоматически диагностировать такие глазные заболевания, как ретинопатия и глаукома [3]. НАПК автоматизирует фундоскопию со смартфона, которая включает в себя точные перемещения объектива для фиксации особенностей сетчатки. Мы стремимся разработать модель, которая распознает частичные участки сетчатки в режиме реального времени во время движения объектива, что позволяет реконструировать полные изображения сетчатки. Методы семантической сегментации помогут распознать особенности сетчатки, несмотря на нехватку аннотированных наборов данных для сегментации изображений сетчатки [4].

Наша система мультиклассификации предназначена для того, чтобы отличать изображения сетчатки с определенными признаками от изображений без них, определяя, является ли изображение частичным или полным. Полные изображения сетчатки позволяют избежать дополнительной обработки и перейти к диагностике. В докладе представлено сравнение потенциала методов, таких как autoencoders, U-Nets, и GANs, для восстановления изображений сетчатки с

использованием наборов данных без указания достоверности. Этот подход улучшает автоматизированный анализ изображений сетчатки, способствуя быстрой диагностике и лечению глазных заболеваний. Был использован набор данных APTOS, содержащий изображения сетчатки, оценённые по степени тяжести диабетической ретинопатии. Кроме того, был включён набор данных MS COCO, предназначенный для решения различных задач компьютерного зрения. Для того чтобы уменьшить возможные дисбалансы из-за различий в размерах наборов данных, случайным образом была выбрана часть набора данных COCO, а изображения сетчатки были загружены вручную из банка изображений сетчатки. Выбранная стратегия предотвращает перегрузку модели и чрезмерное время вычислений, поскольку добавление дополнительных фотографий не окажет существенного влияния на производительность модели. [5, 6, 7].

Данное исследование направлено на расширение области исследований, направленных на применение глубокого обучения и сверточных нейронных сетей (CNN) для автоматического определения особенностей сетчатки на медицинских изображениях. Повышая точность и скорость распознавания изображений на сетчатке, мы стремимся изучить различные архитектуры глубокого обучения, включая трансферное обучение, для выполнения этой задачи. В нескольких исследованиях подчеркивалась важность методов глубокого обучения, в частности CNN, в задачах классификации. Некоторые из этих исследований продемонстрировали эффективность моделей, основанных на CNN, в классификации диабетической ретинопатии на изображениях сетчатки [8], достигая высокого уровня точности за счет обучения с переносом. В данном исследовании рассматривается несколько из этих методов классификации особенностей сетчатки, включая ResNet, MobileNet, Inception, а также построение простых нейронных сетей с нуля [9]. После классификации частичных изображений сетчатки следующей задачей является количественная оценка сегментированных областей. Это включает в себя определение степени присутствия элементов сетчатки на частичных изображениях, что может включать в себя расчет процентной доли изображения, занимаемой этими элементами, или других соответствующих показателей. Поскольку в нашем наборе данных отсутствуют точные аннотации для характеристик сетчатки, необходимы методы сегментации, которые не опираются на такие аннотации. Использование небольших наборов данных для анализа медицинских изображений остается сложной задачей, особенно для задач сегментации. Однако применение методов увеличения объема данных показало многообещающие результаты в повышении эффективности сегментации с использованием небольших наборов данных [10, 11].

Методы сегментации с глубоким обучением, такие как GAN, автоэнкодеры и U-Net, могут быть использованы для задач сегментации без необходимости в аннотациях к базовым данным в наборе данных. Структура данного исследования заключается в разработке концептуальной блок-схемы системы и построении мульти модели, что способно обнаруживать

особенности сетчатки. Мы сравним четыре предварительно обученные модели с передачей данных с помощью простой структурированной нейронной сети. Впоследствии мы представим проблему сегментации и предложим три возможных решения. Такие методы сегментации, основанные на глубоком обучении, такие как GAN, автоэнкодеры и U-Net, могут быть использованы для задач сегментации без необходимости в аннотациях к базовым данным в наборе данных, которые будут проверены для будущих рекомендаций. [12–15].

## II. МАТЕРИАЛЫ И МЕТОДЫ

### A. Наборы данных изображений сетчатки

В данном исследовании использовались два основных набора данных: набор данных APTOS и набор данных MS COCO. Набор данных APTOS состоит из 3662 изображений глазного дна и сетчатки, каждое из которых оценивается врачами на предмет тяжести диабетической ретинопатии по шкале от 0 до 4. Учитывая, что наша цель не заключалась в классификации стадий ретинопатии, все изображения были объединены в единый набор, включающий изображения как здоровой, так и пораженной сетчатки. Кроме того, был включен набор данных MS COCO, известный своими разнообразными функциями и обширной коллекцией, поддерживающей различные задачи компьютерного зрения. Для предотвращения дисбаланса данных из набора данных COCO были случайным образом отобраны 10 000 изображений, что обеспечило оптимальную эффективность и быстродействие модели. Тем более, набор данных был дополнен за счет ручного получения примерно 500 изображений сетчатки из банка изображений сетчатки, полученных с помощью фундоскопии со смартфона. Вышеупомянутый шаг особенно важен, учитывая конечную цель применения результатов исследования для автоматизированной фундоскопии со смартфона в НАПК. Главной целью этого процесса является эффективное использование автоматизированного устройства для фундоскопии со смартфона.

### A. Предварительная обработка данных

С целью расширить набор данных, были сгенерированы частичные изображения сетчатки путем случайного обрезания 1–4 фрагментов из 1200 изображений, полученных из набора данных APTOS, и 300 изображений сетчатки, полученных со смартфона. Каждый кадр был тщательно отобран, чтобы обеспечить разнообразие в разных областях сетчатки и избежать избыточности. Процесс обрезки был случайным образом распределен в пределах от 40% до 90% полной фотографии сетчатки. Важно учитывать, что 40%-ный урожай может не дать достаточной информации для последующего использования. Наглядное представление приведено на рис. 1.



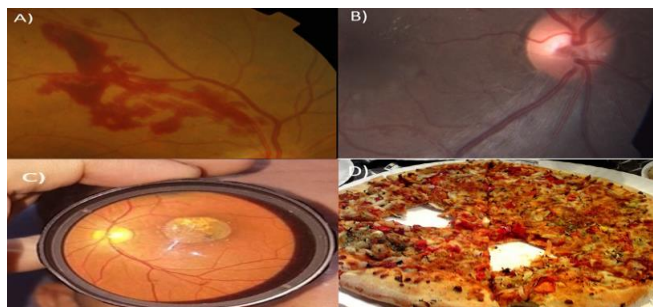


Рис. 1. А) Нездоровое частичное изображение сетчатки. В) Здоровое частичное изображение сетчатки, диска зрительного нерва. С) Полное изображение сетчатки со смартфона. Д) Пример несвязанной фотографии

К изображениям были применены этапы предварительной обработки, включая адаптивное выравнивание гистограммы с ограниченным контрастом (CLAHE) для повышения контрастности и устранения шума, а также гамма-коррекцию для устранения любых цветовых оттенков. Во избежание от дисбаланса в наборе данных, были использованы такие методы, как дублирование существующих выборок и создание синтетической выборки с использованием таких методов, как SMOTE (Synthetic Minority Over-sampling Technique)

Наконец, набор данных был разделен на три класса: полные изображения сетчатки, частичные изображения сетчатки и несвязанные изображения. Вышеописанная стратегия обеспечивает четкость и позволяет проводить целенаправленный анализ для решения задачи классификации [16].

### В. Подготовка наборов данных

После предварительной обработки набор данных был дополнен с помощью ImageDataGenerator для создания новых изображений со случайными преобразованиями, такими как поворот, сдвиг и переворачивание. Это было сделано для увеличения размера и разнообразия обучающего набора данных и уменьшения переобучения. Затем расширенный набор данных был разделен на обучающий и проверочный наборы в соотношении 80:20 [17].

## III. МОДЕЛИ И МЕТОД

В этом разделе мы рассмотрим модели и методологии, используемые для автоматизированного анализа изображений сетчатки. Для исследования были выбраны две основные модели, предназначенные для обнаружения областей сетчатки и оценки их присутствия на изображениях. Первая модель, *RetinaClassifier*, предназначена для классификации изображений по трем категориям: с полной сетчаткой, с частичной сетчаткой и без нее. Рассматриваются различные архитектурные решения, включая ResNet, MobileNet, Inception и простые нейронные сети, созданные с нуля. Для оценки производительности каждой архитектуры. Используются две различные функции потерь: *Weighted Cross-Entropy Loss* и *Categorical Cross-Entropy Loss*. С помощью этого сравнительного анализа мы стремимся определить

наиболее эффективный и точный подход к классификации изображений сетчатки.

Переходя ко второй модели, мы сосредоточимся на выявлении областей сетчатки в частичных изображениях сетчатки и оценке их занятости. Предлагается применить методы сегментации к классифицированным частичным изображениям, используя три предложенных метода, которые не основаны на аннотациях к основным данным. Количественно оценивая степень выраженности особенностей сетчатки на изображениях, мы проводим всесторонний анализ характеристик сетчатки, облегчая последующие диагностические процедуры. Благодаря тщательной оценке и сравнению этих моделей, наше исследование направлено на продвижение в области автоматизированного анализа изображений сетчатки. Внося свой вклад в разработку эффективных диагностических инструментов в офтальмологии, мы стремимся улучшить диагностику и лечение заболеваний сетчатки.

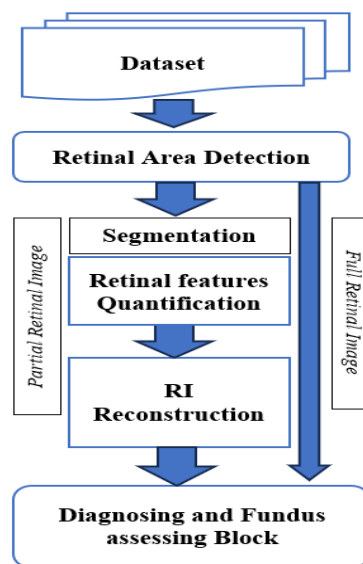


Рис. 2. Принципиальная схема (Flow-chart) предлагаемого способа

Технологическая схема (Flow-chart) предлагаемой модели начинается со сбора и предварительной обработки данных. Сбор данных включает в себя аннотирование набора данных для идентификации областей сетчатки в пределах частичных изображений сетчатки и стандартизацию размеров изображений и цветовых каналов. Переходя к первой модели, нами исследуются различные архитектуры для классификации изображений сетчатки, и каждая из них обучается и оценивается, чтобы определить ее эффективность при классификации изображений на полные и частичные изображения сетчатки или несвязанные

После классификации, если результатом первой модели является полное изображение сетчатки, оно напрямую передается в блок диагностики и оценки состояния глазного дна для анализа без дальнейшей обработки. Однако если фотография представляет собой частичное изображение, каждое из них, классифицированное как частичное изображение сетчатки, переносится в блок сегментации. Здесь применяются методы сегментации для количественной

оценки занимаемой области сетчатки в изображении. Целью всего этого процесса является эффективное использование автоматизированного устройства для фундоскопии со смартфона. Это устройство работает на основе специальной оптической системы, использующей два двигателя, которые перемещаются по осям x и y для получения множества частичных изображений сетчатки. Впоследствии полученные изображения подвергаются анализу и реконструкции для создания целостного представления всей сетчатки. После сегментации результаты суммируются и передаются в блок реконструкции. В этом блоке частичные изображения сетчатки восстанавливаются для получения полного изображения сетчатки. Наконец, восстановленное изображение направляется в блок диагностики и оценки состояния глазного дна для всестороннего анализа и оценки достоверности. Этот итеративный процесс позволяет автоматизировать анализ изображений сетчатки и помогает в диагностике заболеваний сетчатки.

#### IV. РЕЗУЛЬТАТЫ

Оценка моделей глубокого обучения для классификации признаков сетчатки. Был проведен сравнительный анализ с использованием ResNet, MobileNet, Inception и простых нейронных сетей, созданных с нуля. Кроме того, для каждой модели использовались две различные функции потерь: в Weighted Cross-Entropy (CE) Loss и Categorical Cross-Entropy (CE) Loss. Нашей целью было определить наиболее эффективный и точный подход к классификации изображений сетчатки. Все результаты представлены в табл. 1.

ТАБЛИЦА 1. РЕЗУЛЬТАТЫ СРАВНИТЕЛЬНОГО АНАЛИЗА ПРЕДВАРИТЕЛЬНО ОБУЧЕННЫХ МОДЕЛЕЙ

Модель	Функция потерь	Точность	Recall	Precision
ResNet101	WC-E	93%	88%	89%
	CC-E	92%	87%	86%
MobileNet V3	WC-E	91%	90%	87%
	CC-E	88%	87%	84%
Inception v3	WC-E	94%	90%	90%
	CC-E	96%	90%	91%
Simple CNN	WC-E	85%	82%	81%
	CC-E	84%	81%	80%

При оценке были рассмотрены несколько метрик производительности, включая точность, Recall и Precision. Точность измеряется общей правильностью классификации, в то время как Recall and Precision оценивают способность модели правильно идентифицировать положительные экземпляры и ее точность в их предсказании, соответственно. Наблюдалось, что модель Inception в целом превосходилась другими моделями по точности и Precision, достигая 96 % и 91 % соответственно. Можно сделать вывод, что ее сложная архитектура позволяет эффективно извлекать признаки и классифицировать их, что приводит к превосходной производительности. Тем не менее, ResNet и MobileNet также продемонстрировали конкурентоспособную производительность, особенно в сценариях с ограниченными вычислительными

ресурсами. С другой стороны, простые нейронные сети, созданные с нуля, показали сравнительно более низкую точность [9, 18].

MobileNet, хотя и имеет немного меньшую точность - 91% по сравнению с 93% у ResNet и 96% у Inception, благодаря своей легковесной архитектуре, оптимизированной для развертывания на устройствах с ограниченными ресурсами, таких как смартфоны, остается жизнеспособным вариантом. Несмотря на жертву небольшого процента точности, MobileNet обеспечивает более быстрые времена вывода и сниженный объем памяти. В реальном времени, в системах, таких как автоматизированная смартфонная фундоскопия, скорость и эффективность являются критически важными, что делает приемлемым компромисс MobileNet. Более того, разница в точности между MobileNet и другими моделями минимальна, что указывает на приемлемую производительность. Решение о приоритизации MobileNet в конечном итоге зависит от конкретных требований системы и компромиссов между точностью и эффективностью [18].

В данной статье рассматривается вызов, представленный наборами данных, требующими сегментации, но не имеющими аннотаций или размеченных данных. Предлагается использовать методы сегментации на основе глубокого обучения, включая GANs, автокодировщики и U-Net, в качестве многообещающих альтернатив. Упомянутые методы предлагают потенциал для повышения эффективности анализа изображений сетчатки и облегчения разработки эффективных диагностических инструментов в офтальмологии. Путем использования данных техник мы можем преодолеть ограничения традиционных подходов к сегментации, которые сильно зависят от размеченных данных, тем самым расширяя область применения автоматизированных систем анализа изображений сетчатки [12–15].

#### V. ЗАКЛЮЧЕНИЕ

В итоге, данное исследование в области глубокого обучения моделей для автоматизированного анализа ретинальных изображений привнесло многообещающие результаты. Особенно следует отметить выдающуюся производительность модели Inception, обусловленную ее сложной архитектурой, которая превосходит в извлечении признаков и классификации, а также модель ResNet-50, оказавшуюся мощным инструментом для классификации изображений благодаря своей глубокой структуре слоев и пропускам. Однако они часто требуют больших вычислительных ресурсов. Необходимо также отметить практическую эффективность модели MobileNet, что делает ее привлекательным выбором для приложений в реальном времени, таких как автоматизированная смартфонная фундоскопия, благодаря ее легкой архитектуре CNN, разработанной специально для мобильных и встроенных приложений компьютерного зрения. В нашем исследовании мы применили сильные стороны этих моделей для революционизации диагностики ретинальных заболеваний путем быстрого и точного анализа изображений. Наши дальнейшие исследования направлены на улучшение нашего подхода и

подтверждение его эффективности в клинических условиях. Кроме того, мы подчеркиваем важность методов сегментации на основе глубокого обучения, таких как GANs, автокодировщики и U-Net, которые предлагают ценные возможности в сегментации изображений без необходимости аннотаций. Наша статья представляет концептуальную блок-схему, описывающую наш подход, и вводит инновационные решения для сегментационных задач, заложив основу для будущих достижений в анализе ретинальных изображений.

#### СПИСОК ЛИТЕРАТУРЫ

- [1] Eye care, vision impairment and blindness [Electronic resource]. URL: <https://www.who.int/health-topics/blindness-and-vision-loss> (accessed: 04.05.2024).
- [2] A foundation model for generalizable disease detection from retinal images | Nature [Electronic resource]. URL: <https://www.nature.com/articles/s41586-023-06555-x> (accessed: 04.05.2024).
- [3] Corr R.H. Fundoscopy in the smartphone age: current ophthalmoscopy methods in neurology // *Arq. Neuropsiquiatr.* 2023. Vol. 81, № 5. P. 502–509.
- [4] Maya, A. (2024). Enhancing telemedicine with wearable biotechnologies: remote fundus assessment for improved visual health monitoring using smartphones and AI. In 79th Scientific and Technical Conference of the St. Petersburg Scientific Research Radio Technical Institute named after A.S. Popov, Section 10 "Radioelectronics in Medicine and Biology", April 22–26, 2024. Saint Petersburg: Saint Petersburg Electrotechnical University "LETI".
- [5] Tymchenko, B., Marchenko, P., & Spodarets, D. (2020). Deep Learning Approach to Diabetic Retinopathy Detection: arXiv:2003.02261. arXiv, 2020.
- [6] COCO - Common Objects in Context [Electronic resource]. URL: <https://cocodataset.org/#home> (accessed: 04.05.2024).
- [7] Home - Retina Image Bank [Electronic resource]. URL: <https://imagebank.asrs.org/> (accessed: 04.05.2024).
- [8] Abramoff, M.D., Lavin, P.T., Birch, M., Shah, N., & Folk, J.C. (2018). Pivotal trial of an autonomous AI-based diagnostic system for detection of diabetic retinopathy in primary care offices. *npj Digital Medicine*, 1, 39.
- [9] Khattab, R., Abdelmaksoud, I. R., & Abdelrazek, S. (2023). Deep Convolutional Neural Networks for Detecting COVID-19 Using Medical Images: A Survey. *New Gener. Comput.*, 41(2), 343–400.
- [10] Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. In *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, pp. 234–241. doi: 10.1007/978-3-319-24574-4\_28.
- [11] Oliveira, A., Pereira, S., & Silva, C.A. (2018). Retinal vessel segmentation based on Fully Convolutional Neural Networks. *Expert Systems with Applications*, 112, 229–242. doi: 10.1016/j.eswa.2018.06.034.
- [12] Souly, N., Spampinato, C., & Shah, M. (2017). Semi Supervised Semantic Segmentation Using Generative Adversarial Network. pp. 5688–5696.
- [13] You, A., et al. (2022). Application of generative adversarial networks (GAN) for ophthalmology image domains: a survey. *Eye Vis.*, 9(1), 6.
- [14] Sedai, S., Mahapatra, D., Hewavitharanage, S., Maetschke, S., & Garnavi, R. (2017). Semi-supervised Segmentation of Optic Cup in Retinal Fundus Images Using Variational Autoencoder. In *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2017*, pp. 75–82. URL: [https://link.springer.com/chapter/10.1007/978-3-319-66185-8\\_9](https://link.springer.com/chapter/10.1007/978-3-319-66185-8_9) (accessed: 04.05.2024).
- [15] Ullah, Z., Usman, M., Latif, S., Khan, A., & Gwak, J. (Year). SSMD-UNet: Semi-supervised multi-task decoders network for diabetic retinopathy segmentation. *Scientific Reports*. URL: <https://www.nature.com/articles/s41598-023-36311-0> (accessed: 04.05.2024).
- [16] Chawla, N.V., et al. (2002). SMOTE: Synthetic Minority Over-sampling Technique. *J. Artif. Intell. Res.*, 16, 321–357.
- [17] Arora, P., Kapse, V.M., Sinha, S., & Gera, S. (2021). An analytical study for Pneumonia Detection towards building an intelligent system using Image Data Generator. In 2021 9th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO), pp. 1–4. doi: 10.1109/ICRITO51393.2021.9596434.
- [18] Tallapragada, V.V.S., et al. (2023). Improved Accuracy in Image Tampering Detection using MobileNet. In 2023 7th Int. Conf. Trends Electron. Inform. ICOEI. Tirunelveli, India: IEEE, pp. 1337–1343.

# Проверка наличия музыкального образования по фотографиям лиц с помощью сверточных нейронных сетей и Vision Transformer

Я. О. Сениченкова, М. Д. Поляк

Санкт-Петербургский государственный университет аэрокосмического приборостроения

markpolyak@gmail.com

**Аннотация.** В работе исследуется возможность применения методов машинного обучения в патогномике. Рассматривается гипотеза о влиянии профессии человека на лицо. В публикации рассматривается частный случай с музыкантами симфонического оркестра. Был собран датасет, включающий себя фотографии музыкантов симфонического оркестра, и нем музыкантов. Для решения задачи бинарной классификации были использованы нейронные сети VGG-Face, CLIP-ViT, Google ViT. При помощи этих моделей, были извлечены эмбединги из изображений, на которых были обучены машины опорных векторов, которые выполняют роль классификатора. В данной работе сравнивается информативность эмбедингов вышеперечисленных моделей.

**Ключевые слова:** классификация; распознавание изображений; VGG-Face; Vision Transformer; CLIP-ViT; Google ViT; SVM; эмбединги; машина опорных векторов

## I. ВВЕДЕНИЕ

Современные технологии распознавания лиц способны извлекать дополнительную информацию о человеке. Существуют нейросетевые модели, определяющие сексуальную ориентацию [1], политические взгляды [2] путем анализа одних лишь лиц. Первая технологии показала достойные результаты: 91% точности по пяти фотографиям для мужчин и 83% точности по пяти фотографиям для женщин. Заявленная точность определения политических взглядов по фотографии лица – 71%. Помимо вышеперечисленного, существуют модели, которые могут извлечь из фотографии лица информацию о наличии/отсутствии как генетических заболеваний [3], так и психических, таких как шизофрения [4].

Модель Transformer [5] стала революцией среди решения задачи трансдукции последовательностей. До появления данной нейронной сети для решения задачи использовались рекуррентные нейронные сети [6]. Основной их недостаток заключается в том, что вычисления, осуществляемые данными нейросетями, невозможно распараллелить, что ведет к медленной скорости работы модели. Transformer нашел активное применение в задачах нейролингвистического программирования (NLP) [7], однако Transformer можно также применять для задачи классификации изображений [8], в том числе и для распознавания лиц

[9]. В последней задаче модель ViT-P12S8 продемонстрировал точность 99,8% на датасете Labeled Faces In The Wild.

Определение сексуальной ориентации, политических предпочтений, а также других личных качеств тесно связано с физиогномикой, которую принято считать лженаукой. Определение эмоций, напротив, можно рассматривать как задачу патогномики. Отличие между ними в том, что патогномика оперирует отразившимися на лице следами переживаний, последствиями образа жизни, профессии, социального статуса [10], т.е. патогномика базирует свои утверждения на биологических процессах, таких как движение лицевых мышц, формирование морщин, проблемы со здоровьем, и т.п. Классическая физиогномика, напротив, предполагает, что черты характера и поведение людей можно определить по форме и особенностям лица без какого-либо научного обоснования и без биологической интерпретации.

В данной работе рассматривается гипотеза о том, что музыкальное образование и профессия музыканта симфонического оркестра оставляют биологический след на человеческом лице. С точки зрения авторов этой статьи задача определения профессии человека по фотографии его или ее лица относится к задаче патогномики и, следовательно, не должна приводить к возникновению вопроса о «лженаучности». Профессия музыканта симфонического оркестра была выбрана для данного исследования специально не только потому, что весьма затруднительно представить какие-либо обвинения в дискриминации, но еще и потому, что фотографии сотрудников симфонических оркестров находятся в открытом доступе.

## II. МОДЕЛЬ

В данной работе рассматривается частный случай задачи определения профессии человека, а именно, является ли человек музыкантом симфонического оркестра (далее – «музыкант»). Для решения данных задач были использованы предобученные модели VGG-Face[11], Google ViT [8], CLIP-ViT [12]. Также ранее уже была рассмотрена модель FaceNet [13]. Веса использованных моделей находятся в открытом доступе в сети Интернет. В данной работе сравнивается информативность эмбедингов указанных моделей.

На вход экстрактора признаков (feature extractor), который реализован в виде одной из трех перечисленных моделей, подается изображение размером 224x224 пикселей, которое, после прохождения сверточных слоев, преобразуется в вектор, содержащий извлеченные из изображения информативные признаки. Затем вектор признаков, называемый эмбедингом, передается на вход классификатора, реализованного с помощью машины опорных векторов. Данный классификатор вычисляет вероятность принадлежности фотографии одному из двух классов: «музыкант симфонического оркестра» или «не музыкант». Предлагаемая модель для решения задачи бинарной классификации приведена на рис. 1.

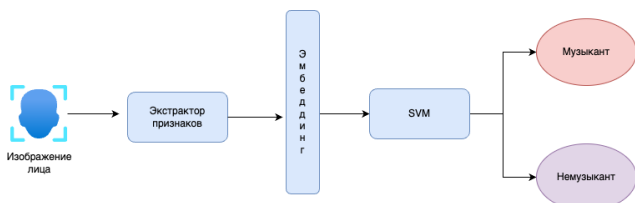
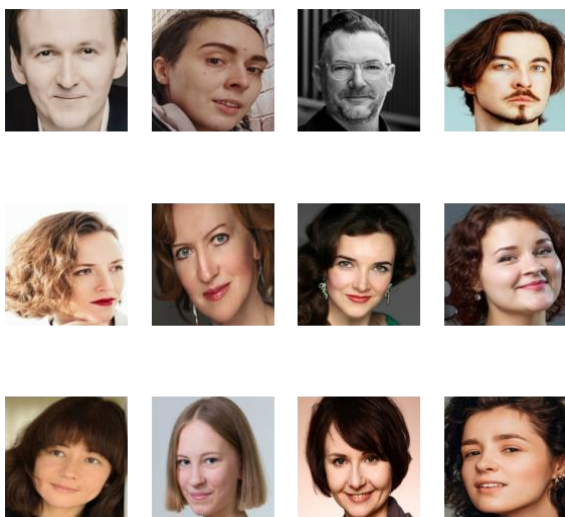


Рис. 1. Модель для бинарной классификации изображений с помощью машины опорных векторов по эмбедингам, полученных из экстрактора признаков

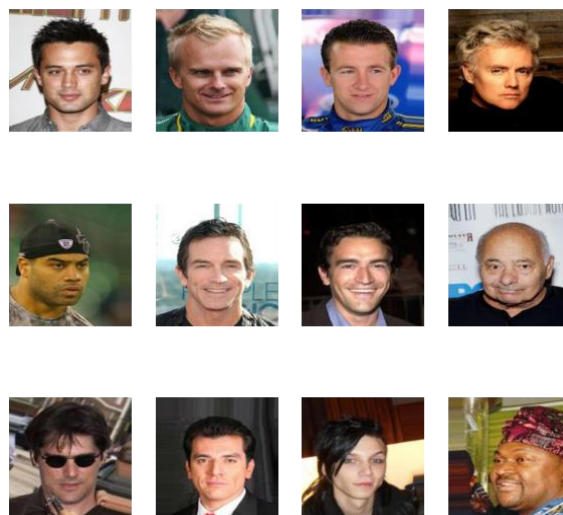
### III. НАБОР ДАННЫХ

Для обучения и тестирования классификатора были собраны фотографии музыкантов с сайтов симфонических оркестров в интернете, а также фотографии медийных личностей из набора данных CelebA, фотографии преподавателей высшей школы. Набор данных включает 1720 изображений, среди них 860 фотографий музыкантов, 860 фотографий немужиков. Тестовая выборка состоит из 258 изображений, где 129 изображений принадлежат к классу «музыкант», а 129 – к классу «немужик»

Примеры фотографий из набора данных приведены на рис. 2.



а)



б)

Рис. 2. Примеры изображений в собранном наборе данных: а) музыканты; б) не музыканты.

### IV. ОБУЧЕНИЕ КЛАССИФИКАТОРА

Для подбора оптимального значения гиперпараметра классификатора на основе машины опорных векторов использовался поиск по решетке (grid search). В качестве функции ядра использовалась радиально-базисная функция (RBF). Поскольку набор данных сбалансирован, для оценки качества обучения классификатора использовалась метрика точности (accuracy).

Результаты обучения классификатора на эмбедингах VGG-Face, Google ViT, CLIP-ViT приведены в таблице 1.

ТАБЛИЦА I. ПОЛУЧЕННЫЕ НА ТЕСТОВОЙ ВЫБОРКЕ МЕТРИКИ КЛАССИФИКАТОРОВ, ОБУЧЕННЫХ НА ЭМБЕДИНГАХ РАЗЛИЧНЫХ МОДЕЛЕЙ

Модель	Метрика		
	Accuracy	F1-score	roc-auc score
VGG-Face	0,713	0,678	0,713
Google ViT	0,867	0,863	0,864
CLIP-ViT	0,872	0,874	0,872

Матрица ошибок для тестовой выборки по предсказаниям классификатора, обученного на эмбедингах VGG-Face, приведена в табл. 2.

ТАБЛИЦА II. МАТРИЦА ОШИБОК ДЛЯ ТЕСТОВОЙ ВЫБОРКИ ПО ПРЕДСКАЗАНИЯМ КЛАССИФИКАТОРА, ОБУЧЕННОГО НА ЭМБЕДИНГАХ VGG-FACE

Истинный класс	Предсказанный класс	
	Музыкант	Немузыкант
Музыкант	106	23
Немузыкант	51	78

Матрица ошибок для тестовой выборки по предсказаниям классификатора, обученного на эмбедингах Google ViT, приведена в табл. 3.

ТАБЛИЦА III. МАТРИЦА ОШИБОК ДЛЯ ТЕСТОВОЙ ВЫБОРКИ ПО ПРЕДСКАЗАНИЯМ КЛАССИФИКАТОРА, ОБУЧЕННОГО НА ЭМБЕДИНГАХ GOOGLE ViT

Истинный класс	Предсказанный класс	
	Музыкант	Немузыкант
Музыкант	113	16
Немузыкант	19	110

Матрица ошибок для тестовой выборки по предсказаниям классификатора, обученного на эмбедингах CLIP-ViT, приведена в табл. 4.

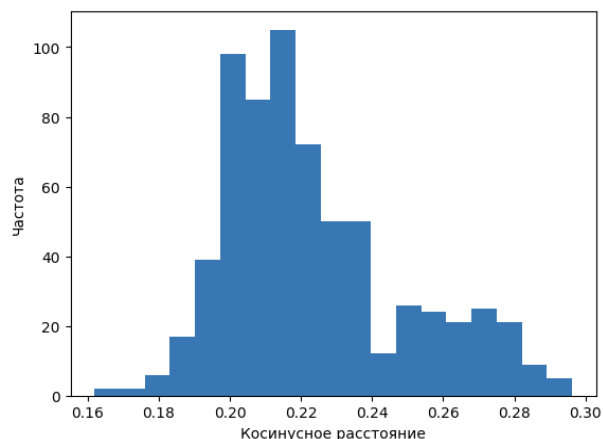
ТАБЛИЦА IV. МАТРИЦА ОШИБОК ДЛЯ ТЕСТОВОЙ ВЫБОРКИ ПО ПРЕДСКАЗАНИЯМ КЛАССИФИКАТОРА, ОБУЧЕННОГО НА ЭМБЕДИНГАХ CLIP-ViT

Истинный класс	Предсказанный класс	
	Музыкант	Немузыкант
Музыкант	111	18
Немузыкант	15	114

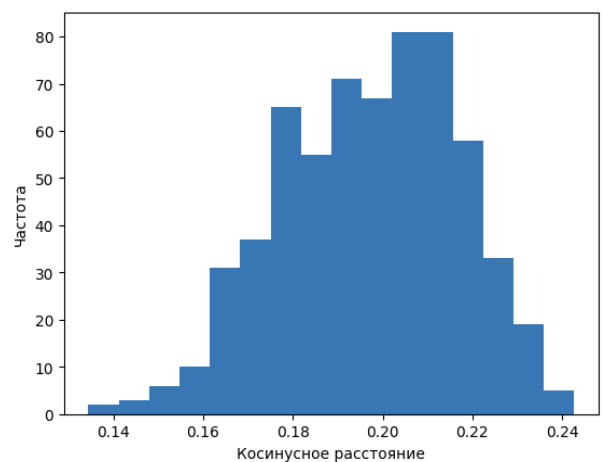
## V. ИСПОЛЬЗОВАНИЕ CLIP-ViT

Особенность модели CLIP-ViT заключается в способности сопоставлять изображения и тексты в одном и том же векторном пространстве. Авторами данной работы была выдвинута гипотеза, что эмбединг словосочетания «symphony orchestra musician» (музыкант симфонического оркестра) и эмбединг изображения музыканта симфонического оркестра будут схожи, о чем будет свидетельствовать косинусное расстояние, близкое по модулю к единице; в случае с эмбедингами, полученными из изображений немусыкантов, косинусное расстояние будет близко по модулю к нулю, что будет свидетельствовать о том, что текст и изображение не соответствуют друг другу. Были получены эмбединги изображений обоих классов и эмбединг словосочетания «symphony orchestra musician». Между каждым эмбедингом изображения и эмбедингом словосочетания было рассчитано косинусное расстояние. Гистограмма распределения косинусных мер сходства между эмбедингом словосочетания и эмбедингами изображений класса «Музыкант» приведена на рис. 3.

Согласно гистограммам, приведенным на рис. 3, значения косинусных мер для обоих классов не превышают по модулю значение 0,3, из чего можно сделать вывод, что благодаря сравнению эмбедингов изображений с эмбедингом фразы «symphony orchestra musician» невозможно однозначно определить, является ли человек музыкантом или нет.



а)



б)

Рис. 3. Гистограмма распределения косинусных мер между эмбедингом фразы «symphony orchestra musicians» и эмбедингами изображений класса а) «Музыкант» б) «Немузыкант»

Согласно гистограммам, приведенным на рис. 3, значения косинусных мер для обоих классов не превышают по модулю значение 0,3, из чего можно сделать вывод, что благодаря сравнению эмбедингов изображений с эмбедингом фразы «symphony orchestra musician» невозможно определить, является ли человек музыкантом или нет.

## VI. МОДЕЛЬ GOOGLE-ViT

Модели архитектуры ViT имеют патч-эмбединг слой и энкодер трансформер. Слой патч-эмбединга принимает на вход изображение и разбивает его на сетку патчей. После этого последовательность патчей поступает на вход энкодера. Далее преобразованные изображения поступают на вход классификатора [8].

В данной работе модель Google-ViT была использована в качестве экстрактора признаков, т. е. она преобразовывала изображение в вектор из 768 элементов, который впоследствии поступает на вход описанного ранее классификатора.

## VII. ЗАКЛЮЧЕНИЕ

С использованием алгоритмов машинного обучения были построены три классификатора, позволяющие с точностью 0,713, 0,867, 0,872 соответственно по одной фотографии лица определить, является ли профессией человека выступление в составе симфонического оркестра. Нейронные сети группы Transformer показали наилучшие результаты в рамках решения поставленной задачи. Данное исследование показывает, что на изображениях лиц существуют незаметные для человеческого взгляда особенности, которые с легкостью могут быть найдены алгоритмами машинного обучения.

Полученные результаты свидетельствуют о том, что технологии распознавания лиц представляют реальную угрозу конфиденциальности, поскольку в некоторых случаях профессия человека может быть определена по фотографии его или ее лица без предварительного на то согласия. Данное исследование может иметь практическое применение в сфере маркетинга для формирования персональных рекомендаций (рекламы) товаров и услуг, например, в магазинах, торговых центрах и т. д., за счет обработки изображений лиц посетителей, получаемых с существующих камер видеонаблюдения.

## СПИСОК ЛИТЕРАТУРЫ

- [1] Wang Y., Kosinski M. Deep neural networks are more accurate than humans at detecting sexual orientation from facial images //Journal of personality and social psychology. Т. 114, № 2, 2018. 246 с.
- [2] Kosinski M. Facial recognition technology can expose political orientation from naturalistic facial images. Scientific reports, Т. 11. №. 1, 2021. 100 с.
- [3] Hallgrímsson B. et al. Automated syndrome diagnosis by three-dimensional facial imaging. Genetics in Medicine. Т. 22, 2020.
- [4] Zhang X. et al. Recognizing schizophrenia using facial expressions based on convolutional neural network. Brain and Behavior, Т. 13, №. 5, 2023. e3002 с..
- [5] Vaswani A. et al. Attention is all you need. Advances in neural information processing systems. Т. 30, 2017.
- [6] Graves A. Sequence transduction with recurrent neural networks. arXiv preprint arXiv:1211.3711, 2012.
- [7] Gillioz A. et al. Overview of the Transformer-based Models for NLP Tasks. 2020 15th Conference on Computer Science and Information Systems (FedCSIS), IEEE, 2020. 179-183 с.
- [8] Dosovitskiy A. et al. An image is worth 16x16 words: Transformers for image recognition at scale. arXiv preprint arXiv:2010.11929, 2020.
- [9] Zhong Y., Deng W. Face transformer for recognition. arXiv preprint arXiv:2103.14803, 2021.
- [10] Bendel O. The uncanny return of physiognomy. 2018 AAAI Spring Symposium Series, 2018.
- [11] Parkhi O., Vedaldi A., Zisserman A. Deep face recognition. BMVC 2015-Proceedings of the British Machine Vision Conference 2015, British Machine Vision Association, 2015.
- [12] Radford A. et al. Learning transferable visual models from natural language supervision. International conference on machine learning, PMLR, 2021. 8748-8763 с.
- [13] Senichenkova Y., Polyak M. Finding Traces of Music Education on Facial Images. XXVI International Conference on Soft Computing and Measurements (SCM), 2023, pp. 167-170.

# Особенности нейронных сетей в задаче прогнозирования уровня аварийности на дорогах общего пользования

А. Ю. Абаляев, Л. В. Грунская

Владимирский государственный университет имени Александра Григорьевича и Николая Григорьевича Столетовых

ice\_aya@bk.ru, grunsk@vlsu.ru

**Аннотация.** Использование статистических методов построения прогнозов не всегда обеспечивает высокую достоверность результатов прогнозирования, так как не могут в полной мере учесть влияние факторов, имеющих относительно слабую корреляцию с уровнем прогнозируемого явления, однако оказывающих влияние на человека, его психо-физиологическое состояние, реакции на внешние раздражители, и, в конечном счете, на вероятность в данных, конкретных, условиях создание аварийной ситуации.

**Ключевые слова:** рекуррентные нейронные сети, прогнозирование, средняя аварийность, гео- и гелиофизические факторы, повышение точности прогнозов

## I. ВВЕДЕНИЕ

Прогнозирование аварийности на дорогах общего пользования в краткосрочной перспективе, от одних суток до недели или двух, является актуальной задачей, поскольку позволяет определить оперативную потребность экстренных служб в средствах спасения, первой помощи, дежурных экипажах и т. п. Использование статистических методов построения прогнозов не всегда обеспечивает высокую достоверность результатов прогнозирования, так как не могут в полной мере учесть влияние факторов, имеющих относительно слабую корреляцию с уровнем прогнозируемого явления, однако оказывающих влияние на человека, его психо-физиологическое состояние, реакции на внешние раздражители, и, в конечном счете, на вероятность в данных, конкретных, условиях создание аварийной ситуации.

## II. МЕТОДЫ ИССЛЕДОВАНИЯ

Вопросы прогнозирования аварийности отечественными и зарубежными исследователями изучаются довольно давно. Если первоначально речь шла об использовании статистических методов, то с развитием технологий искусственного интеллекта расширяется их использование для оценки безопасности дорожного движения. Достаточно полный анализ методов прогнозирования уровня аварийности на дорогах общего пользования приведен в работе [1].

Следует отметить, что в практике применения методов искусственного интеллекта, и, в частности, искусственных нейронных сетей для прогнозирования аварийности существует два подхода. Первый, предполагающий анализ дорожной обстановки и

построение прогнозов ее развития на конкретных участках дорог, например, регулируемых и нерегулируемых перекрестках. Второй позволяет оценить зависимости от влияния различных факторов.

Для обучения искусственных нейронных сетей и оценки качества прогноза использовались данные по суточному уровню аварийности на региональных дорогах общего пользования за период 2001–2010 гг. [2]. Общий объем выборки составил 3652 отсчета. Анализ данных показал большое влияние случайной составляющей на величину суточного уровня аварийности. Так, в результате мультипликативной декомпозиции временного ряда [3] были получены оценки мощности трендовой  $FT = 0,265$  и сезонной  $FS = 0,019$  составляющих, что свидетельствует о достаточно слабой сезонности. В тоже время нельзя говорить и о сильном влиянии трендовой составляющей уровня аварийности, что позволяет сделать вывод об определяющем влиянии случайной составляющей временного ряда уровня аварийности. Поэтому, с целью уменьшения влияния случайной составляющей на результаты прогнозирования, проводилось осреднение используемых данных об уровне аварийности на дорогах общего пользования методом скользящего среднего с интервалом осреднения, равным 7 суток.

Используемый временной ряд разбивался на последовательности из 30 значений среднего уровня аварийности, используемых в качестве входа нейронной сети, и 7 «прогнозных» значений, применяемых в качестве выходов при обучении и валидации результатов прогнозирования. Данные о среднем уровне аварийности на региональных дорогах общего пользования предварительно подвергались нормализации.

В качестве обучающей выборки использовались три четверти из числа полученных последовательностей, а оставшаяся четверть – в качестве контрольной выборки.

На уровень аварийности оказывают влияние различные факторы. Одним из таких факторов, влияющих на психофизиологическое состояние человека, является уровень солнечной активности. Число Вольфа, характеризующее число солнечных пятен, вычисляется как

$$W = 10N_g + N_s,$$

где  $N_g$  – число групп солнечных пятен,  $N_s$  – число солнечных пятен в момент наблюдения.



В большинстве случаев числа Вольфа, определенные различными наблюдателями, несколько различаются. Это связано с опытом наблюдателя, его остроты зрения, особенностей используемого оборудования и условий наблюдения. Для того, чтобы уменьшить влияние случайных факторов на значение числа Вольфа при решении практических задач используется международный индекс числа солнечных пятен, определяемый по результатам измерений порядка 80 станций наблюдения по всему миру [4]. Результаты многолетних наблюдений показывают ярко выраженный 11-летний цикл солнечной активности.

Результаты наблюдения за Солнцем показывают, что число Вольфа для северного и южного полушарий звезды существенно различаются. Однако, подобные наблюдения стали регулярными начиная с 2015 г. [4].

Поэтому, для получения непрерывной последовательности чисел Вольфа для полушарий Солнца необходимо определить относительные доли числа солнечных пятен за период наблюдения в один день для северного:

$$n_i = N_i / (S_i + N_i),$$

где  $N_i$  – число пятен в северном полушарии Солнца,  $S_i$  – число пятен в южном полушарии Солнца,  $i$  – день одиннадцатилетнего цикла солнечной активности текущего периода, и южного полушарий Солнца:

$$s_i = S_i / (S_i + N_i).$$

Затем, полученные значения  $n_i$  и  $s_i$  используются для определения числа солнечных пятен за предыдущий период, для которого такие данные отсутствуют, в северном

$$R_{Ni} = n_i \cdot R_i,$$

где  $R_i$  – общее число солнечных пятен,

и южном полушариях

$$R_{Si} = s_i \cdot R_i.$$

Необходимо иметь ввиду выполнение тождества

$$R_{Ni} + R_{Si} = R_i.$$

Особенностью ряда данных, описывающих числа Вольфа, является его достаточно большая волатильность (рис. 1).

Эта особенность не позволяет с достаточной точностью использовать данные непосредственных наблюдений для задач прогнозирования как активности Солнца активности, так и для исследования влияния солнечной активности на происходящее на Земле процессы. С целью уменьшения влияния случайной составляющей ряда чисел Вольфа на результаты прогнозирования рекомендуется [4] использовать сглаженные кривые.

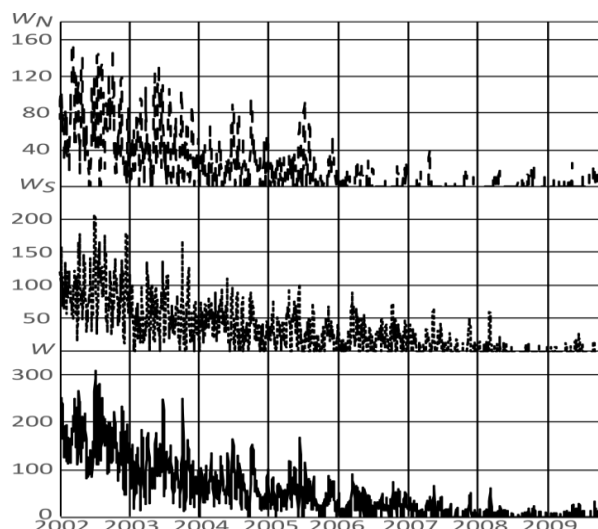


Рис. 1. Число Вольфа за рассматриваемый период – несглаженные данные

Результат сглаживания ежедневных данных о числах Вольфа в северном и южном полушариях Солнца методом скользящего среднего с периодом 13 месяцев приведены на рис. 2.

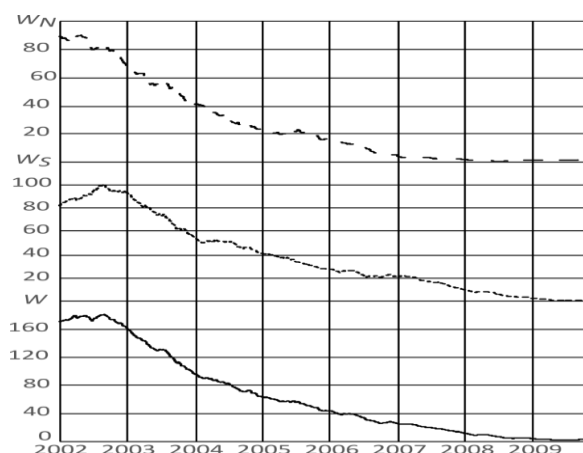


Рис. 2. Число Вольфа за рассматриваемый период – осредненные данные

Сравнительный анализ приведенных на рис. 1 данных показывает, что для наблюдаемого числа Вольфа волатильность для несглаженных данных действительно оказывается достаточно велика, что, в свою очередь, во-первых, может оказать отрицательное влияние на результаты прогнозирования, а, во-вторых, такая величина числа Вольфа сама по себе не поддается точному прогнозированию для будущих периодов. Последнее становится особенно актуальным в случае прогнозирования развития ситуации со средним уровнем аварийности для временных интервалов, лежащих за пределами периода регистрации солнечной активности. В отличие от данных, представленных на рис. 1, сглаженные данные обладают существенно меньшей волатильностью. Для них возможна как интерполяция, так и экстраполяция ряда значений за пределы периода регистрации чисел Вольфа с достаточной для практических целей точностью.

Числа Вольфа, характеризующие солнечную активность в период, соответствующий 2001–2010 гг., были определены на основании данных [4]. При этом, с целью обеспечения равноточности, по рекомендации [4], проводилось сглаживание данных с шириной окна, равной интервалу осреднения временного ряда уровня аварийности. Для обеспечения более высокого уровня достоверности в процессе обучения и валидации прогнозов, полученных с использованием нейронных сетей, было принято решение учитывать количество пятен в северном и южном полушариях солнца по отдельности.

Дополнительным фактором, позволяющим увеличить точность прогнозирования, является не просто учет абсолютного значения фактора, но и скорость изменения указанного значения. Последняя равняется первой производной значения фактора по времени и вычисляется достаточно просто, например, с использованием разностных схем первого порядка точности, таких как разности назад или разности вперед, либо второго порядка точности – центральные разности. В настоящей работе использована центрально-разностная схема второго порядка точности.

Всего, для исследования влияния различных факторов на точность прогноза среднего уровня аварийности на дорогах общего пользования, были сформированы 7 различных наборов данных (табл. 1).

ТАБЛИЦА I. НАБОРЫ ДАННЫХ ДЛЯ ОБУЧЕНИЯ И ВАЛИДАЦИИ РНС ДЛЯ ПРОГНОЗИРОВАНИЯ СРЕДНЕГО УРОВНЯ АВАРИЙНОСТИ НА ДОРОГАХ ОБЩЕГО ПОЛЬЗОВАНИЯ

№	Уровень ДТП	Первая производная от уровня ДТП	Число солнечных пятен в северном полушарии	Первая производная от числа солнечных пятен в северном полушарии	Число солнечных пятен в южном полушарии	Первая производная от числа солнечных пятен в южном полушарии
1	X	X				
2	X	X	X			
3	X	X		X		
4	X	X	X	X		
5	X	X			X	
6	X	X				X
7	X	X			X	X

Для прогнозирования среднего уровня аварийности на дорогах общего пользования использовались рекуррентные нейронные сети с использованием LSTM-нейронов [5]. Для оценки влияния архитектуры рекуррентной нейронной сети (РНС) и ее емкости на точность получаемого прогноза использовалась нейронная сеть, состоящая из: двух слоев LSTM-нейронов, слоя нормализации (Batch Normalization Layer), скрытого и выходного слоев обратного распространения, в дальнейшем обозначаемой как РНС1. Для исследования влияния емкости нейронной сети на точность прогнозирования была использована РНС с дополнительным скрытым слоем обратного распространения, в дальнейшем обозначаемый как РНС2.

В процессе исследования была выявлена склонность РНС2 к переобучению вследствие достаточно высокой емкости сети и относительно небольшого набора данных для обучения. Для того, чтобы избежать переобучения нейронной сети, между скрытыми слоями обратного распространения был добавлен слой отбрасывания (Dropout).

### III. РЕЗУЛЬТАТЫ ИССЛЕДОВАНИЯ И ИХ ОБСУЖДЕНИЕ

Результаты оценки ошибки прогноза среднего уровня аварийности, полученного с помощью рекуррентной нейронной сети на тестовой выборке, приведены в табл. 2.

ТАБЛИЦА II. ОШИБКА ПРОГНОЗА РНС, ОБУЧАЕМОЙ НА ПРОТЯЖЕНИИ 10 ЭПОХ

Вариант	1	2	3	4	5	6	7
РНС1	0,5446	0,5627	0,5670	0,5623	0,5554	0,5439	0,5316
РНС2	0,6098	0,6016	0,5811	0,7104	0,5625	0,5825	0,5798

Анализ представленной в табл. 2 информации показывает, что включение слоя отбрасывания приводит к некоторому ухудшению точности прогноза. При этом ошибка прогноза достаточно сильно зависит от используемого набора данных. Так, для РНС1 наименьшую ошибку валидации показал набор данных 7 (табл. 1), включающий в свой состав как количество пятен в южном полушарии солнца, так и скорость их появления. А для РНС2, обладающей большей емкостью и включающей в свой состав слой отбрасывания для избежания переобучения, наилучшим оказался набор данных 5, сформированный без учета скорости появления солнечных пятен в южном полушарии. Набор данных 7 также показал относительно низкую ошибку валидации, превышающую минимальное значение всего на 3%. Такой результат может быть связан с рандомизацией обучающей выборки. Внедрение слоя отбрасывания, ожидаемо, привело к снижению точности прогнозирования с использованием РНС, однако, увеличенная емкость нейронной сети с двумя скрытыми слоями обратного распространения позволяет более точно оценить тенденцию изменения среднего уровня аварийности на дорогах общего пользования с использованием РНС2, нежели РНС1 (рис. 3).

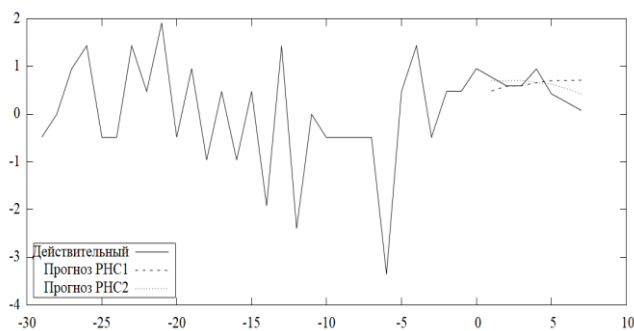


Рис. 3. Сравнение результатов прогнозирования РНС1 и РНС2

Анализ представленной на рис. 3 информации показывает, что несмотря на в среднем меньшую оценку ошибки прогнозирования для РНС1, прогноз, полученный с помощью РНС2, обладающей большей емкостью благодаря включению дополнительного

скрытого слоя обратного распространения, позволяет точнее оценить тенденцию изменения среднего уровня аварийности на дорогах общего пользования, хотя и не полностью учитывает особенности изменения ситуации, в частности, наличие локальных экстремумов на кривой наблюдаемого (действительного) среднего уровня аварийности.

#### IV. ЗАКЛЮЧЕНИЕ

В целом, исходя из изложенного выше, можно сделать следующие выводы:

1. Для прогнозирования среднего уровня аварийности на дорогах общего пользования можно использовать рекуррентные нейронные сети, содержащие LSTM-нейроны.

2. Применение РНС с двумя скрытыми слоями обратного распространения, несмотря на склонность такой нейронной сети к переобучению, позволяет спрогнозировать не только уровень аварийности в краткосрочной перспективе, но и тенденцию его изменения. РНС с одним скрытым слоем не обладают такой возможностью.

3. Использование в процессе обучения РНС информации о дополнительных природных факторах, оказывающих влияние на психофизиологическое

состояние человека, позволяет увеличить точность прогнозирования среднего уровня аварийности на дорогах общего пользования.

4. Существенное влияние на точность прогнозирования оказывает не только информация об изменении прогнозируемой величины и дополнительных факторов с течением времени, но и информация о скорости их изменения.

#### СПИСОК ЛИТЕРАТУРЫ

- [1] A.Mohammed, "Classification of Traffic Accident Prediction Models", A Review Paper, International Journal of Advances in Science Engineering and Technology, 2018, Vol 6, pp. 35-38.
- [2] Показатели состояния безопасности дорожного движения [Электронный ресурс], <http://stat.gibdd.ru> (23.04.2024 г.).
- [3] R.J. Hyndman and G. Athanasopoulos, "Forecasting: Principles and Practice", OTexts, 2021, 442 p.
- [4] M. Astrid Veronig, J. Shantanu, T. Podladchikova, W. Pötzi, and F. Clette, "Hemispheric Sunspot Numbers 1874-2020", Astronomy & Astrophysics, 2020, Vol. 652, pp. 56-57.
- [5] A.Y. Abalyaev and L.V. Grunskaya, "Development of Recurrent Neuron Net System for Forecasting of an Accident Rate Mean Level on General Use Roads", 2022 International Conference on Engineering Management of Communication and Technology (EMCTECH), Vienna, Austria, 2022, pp. 1-4.

# Оценка выполнения физических упражнений методами глубокого обучения

М. Д. Кораблева, Я. А. Бекенева

Санкт-Петербургский государственный электротехнический университет  
«ЛЭТИ» им. В.И. Ульянова (Ленина)

marinkakd.ru23322@gmail.com, yabekeneva@etu.ru

**Аннотация.** В последние годы большую популярность приобрели приложения, предоставляющие пользователям программы физической активности. Такие программы включают себя комплексы упражнений, направленные на определенные группы мышц, кардиотренировки и т. п. Пользователь не привязан к определенному залу и может выполнять упражнения в любом удобном месте. Однако при этом отсутствует контроль со стороны тренера. Целью работы является создание средств анализа качества выполнения упражнений пользователем. На основе изображения, которое фиксируется с помощью камеры пользователя за определенное время, создавая видеопоток, производится сравнение с эталонной техникой исполнения тренера, заранее добавленное в систему. В ходе тренировки пользователю постоянно показывается неправильное положение ключевой точки (в виде подвижных частей тела), обеспечивая возможность качественно выполнить упражнение.

**Ключевые слова:** компьютерное зрение, обработка изображений, определение поз

детекции ключевых точек даёт ощутимое преимущество перед обычными моделями обнаружения объектов, которые гарантируют на выходе лишь получение области, ограничивающей сам объект на изображении (boundary box). Детекция частей тела может быть полезна в самых различных ситуациях (медицинской диагностике, робототехнике, технологиях виртуальной реальности и многих других), в нашем случае – в анализе качества выполнения физических упражнений.

В случае анализа качества выполнения физических упражнений, детекция частей тела может помочь в оценке правильности позы, движений и техники выполнения упражнений. Такой подход позволяет самостоятельно контролировать свои движения и корректировать их для улучшения результатов. В целом, использование моделей детекции ключевых точек или скелета в области анализа физических упражнений может значительно повысить эффективность тренировок и предотвратить возможные травмы.

## I. ВВЕДЕНИЕ

Видеоуроки, а в последние годы онлайн курсы и специальные мобильные приложения предоставляют возможность заниматься физкультурой в домашних условиях в удобное время. Неоспоримое преимущество онлайн программ состоит в отсутствии временных и финансовых затрат на дорогу до спортивного центра. Согласно статистическому исследованию [1], число пользователей приложений для фитнеса выросло вдвое за период с 2018 по 2021. Существенным недостатком является отсутствие контроля за фактом выполнения упражнений, и качеством их выполнения. Неверные движения могут не принести ожидаемой пользы для здоровья, а напротив, нанести вред и способствовать закреплению некорректных двигательных привычек. Существуют онлайн курсы, где требуется снимать видео и отправлять файл тренеру для проверки, однако в этом случае отсутствует контроль в режиме реального времени, и качество выполнения упражнений оценивается после завершения тренировки. Следовательно, существует необходимость в разработке инструментов контроля в режиме реального времени.

Обнаружение объектов – это обширная область компьютерного зрения, которая используется во множестве повседневных действий. С развитием этой области знаний появилась необходимость в обнаружении ключевых точек, будь то тело человека полностью или отдельно: детекция конечностей или лица. В некоторых случаях использование моделей

## II. РЕЛЕВАНТНЫЕ РАБОТЫ

Подавляющее большинство фитнес приложений в магазинах приложений на ПК и смартфонах – приложения, в которые не внедрен искусственный интеллект [2]. По сути, они представляют собой простые трекеры активности: они записывают время выполнения тренировки и примерное количество сожженных калорий (которые заранее рассчитаны на каждое упражнение). Сам процесс тренировки представляет собой набор упражнений, которые необходимо выполнить за назначенное время.

Принцип работы таких приложений очень простой, что является как плюсом, так и минусом. Все, что касается любого рода физической нагрузки является травмоопасным: бывают случаи, что во время таких одиночных тренировок пользователь может что-то себе повредить, так как не была соблюдена техника выполнения упражнения. Также такие приложения не могут гарантировать результат вследствие невозможности отследить действия человека.

На рынке приложений их не так много, но можно выявить следующих представителей: Infigro [3], DeepSport [4] и ALPHA AI Coach [5]. Все они представляют собой персонализированные фитнес приложения с использованием возможностей искусственного интеллекта. Infigro предлагает тренировки с отзывами о выполнении техники в режиме реального времени, а также интерактивные тренировки

для предотвращения травмоопасных ситуаций в режиме реального времени. Для них пользователю необходимо иметь под рукой телефон (для ПК он не предназначен). Приложение предлагает большую базу в виде сотни тренировок различных категорий, поэтому каждый найдет то, что ему по душе.

Во время проведения тренировки на экране отображается видео с техникой выполнения упражнения, при этом параллельно камера фиксирует движения пользователя. В приложении не добавлена функция наложения точек скелета поверх тела пользователя (будучи встроенной для осуществления анализа), что в определенных обстоятельствах может быть менее удобным. Если обнаружена ошибка в технике, голосом воспроизводится комментарий, способствующий ее улучшению. Также на экране отображается процент выполнения конкретного упражнения, а после окончания тренировки пользователь может оценить, насколько точно он следовал инструкциям, основываясь на отображенном проценте, а также получить информацию по времени тренировки и приблизительном количестве сожженных калорий.

DeepSport нацелен на домашние тренировки, не требующие специального оборудования. В основе тренировок лежит распознавание позы человека, попавшего в кадр, при этом выполнение упражнений сопровождается комментариями и звуковыми эффектами.

При первом скачивании приложения пользователю предлагается оценить свои возможности в спорте в тесте «на пригодность», необходимо выполнить ряд несложных упражнений. Это полезно, так как по окончании тестирования производится анализ полученных результатов (показатель выносливости и силовой характеристики), на основе которых пользователю предлагают более подходящий ему комплекс упражнений. Во время проведения тренировки на экране отображается видео с техникой выполнения упражнения, для напоминания, а также видео с камеры телефона человека с наложением поверх него ключевых точек скелета. После физической активности пользователю предлагают ознакомиться со своим результатом в виде процента попадания в технику, времени тренировки и примерного количества сожженных калорий.

Alpha AI Coach предлагает тренировки, направленные на различные цели: потеря веса, фитнес и подобные. База пока маленькая, насчитывается около 10 упражнений на все цели. Выполнение упражнений сопровождается анализом видео с камеры пользователя в режиме реального времени. Искусственный интеллект предоставляет звуковые и визуальные рекомендации по улучшению техники.

Приложение предлагает выбрать «проблемные» зоны пользователя и количество тренировок в день. Во время выполнения упражнений нет эталона, на который необходимо ориентироваться, что довольно неудобно, есть только минимальные указания «поднимитесь», «опуститесь» при этом в приложении как таковой непрерывной тренировки нет: необходимо после каждого выполненного подхода подходить к телефону,

чтобы выбрать другое упражнение. Плюсом можно считать отображение поверх видеопотока пользователя его ключевых точек. Приложение оценивает количество (время тренировки и сожженные калории), а не качество (проценты, характеризующие точность исполнения техники), что является огромным недостатком.

### III. РАЗРАБОТКА ПРОГРАММНОГО ПРОТОТИПА

Обнаружение объектов в рамках выполнения работы ориентировано на определение позы человека, попавшего в объектив камеры. Возьмем на рассмотрение модели глубокого обучения YOLO [6] и Mediapipe [7], так как именно их используют часто для решения этой задачи, а также рассмотрим, на основе какого датасета они обучались, а именно на COCO [8].

COCO (Common Object in Context) – это крупномасштабный набор данных для обнаружения, сегментации объектов и создания подписей. Датасет предоставляет более 330 000 изображений и 1,5 миллиона экземпляров объектов.

COCO широко используется в исследованиях в области компьютерного зрения и используется для обучения и оценки многих современных моделей. Он предоставляет множество решений: обнаружение объектов, различные виды сегментации, обнаружение ключевых точек, задача «плотной позы» (оценивает трехмерную позу объектов или людей на изображении).

YOLO (You Only Look Once) – метод идентификации и распознавания объектов на фотографиях и видео, который впервые был предложен в 2015 году и опубликован в 2016 году Джозефом Рэдмондом. Его скорость, точность распознавания объектов и простота в обучении сделала YOLO одним из самых используемых и распространенных алгоритмов в области компьютерного зрения.

Для некоторых моделей YOLO характерна возможность переключаться между процессорами вычислительного устройства. Таким образом, при наличии мощного GPU можно использовать его для ускорения обучения и выполнения моделей YOLO, что существенно сократит время обработки изображений или видео. При этом, если доступа к графическому процессору нет – можно запустить фреймворк на CPU. Это делает YOLO более гибким и универсальным.

Существует более 10 версий YOLO, выпущенные разными авторами, наибольший интерес для реализации продукта представляет YOLOv7, который на данный момент является единственной моделью среди всей линейки продуктов, способной осуществлять детекцию скелета. Обнаружение позы YOLO выполняет для всех кадров.

Mediapipe – решение, предложенное компанией Google, которое внедряли разработчики, начиная с 2012 года. Изначально он был разработан для анализа видео и аудио в режиме реального времени на YouTube. Постепенно Mediapipe был интегрирован во многие другие продукты компании. Публичный релиз состоялся в 2019 году, дав возможность развиваться самостоятельным разработчикам в области компьютерного зрения.

В качестве фреймворка, в котором заложены функции распознавания объектов, выбран Mediapipe, разработанный компанией Google. Основными функциями фреймворка являются распознавание лиц (face mesh), определение позы человека (pose estimation), а также частей тела. Программное решение будет представлять собой пользовательское приложение, которому для запуска и работы достаточно мощностей графического процессора, также оно, в теории, может стать кроссплатформенным, поэтому реализация на YOLO могла бы иметь затруднения в осуществлении этой затеи.

Пользование приложением для оценки качества выполнения упражнений носит индивидуальный характер (не групповой). Большинство таких программ подразумевает заполнение параметров человека таких как: рост, вес, предпочтения – поэтому распознавание в кадре нескольких людей просто не имеет смысла. Также сам процесс обработки у Mediapipe более «легкий» по отношению к YOLO, поэтому во время тренировок у пользователя должно возникать меньше проблем с потерей кадров. Это обеспечит большее удобство и возможность контролировать свои движения в режиме, приближенном к реальному времени (так как минимальное отставание все еще может присутствовать).

Наличие чуть большего количества точек у Mediapipe Pose может обеспечить обработку положения ступней и пальцев, в отличие от YOLOv7. Если рассматривать упражнения из йоги, то во многих упор делается еще и на эти менее заметные для обывателя части тела. Это важно, так как именно во время неправильной постановки ступни можно получить растяжение в том или ином месте. В данном контексте Mediapipe будет более востребованным.

Модель Mediapipe для предсказания поз содержит информацию о 33 точках тела человека. Разметка ключевых точек представлена на рис. 1.

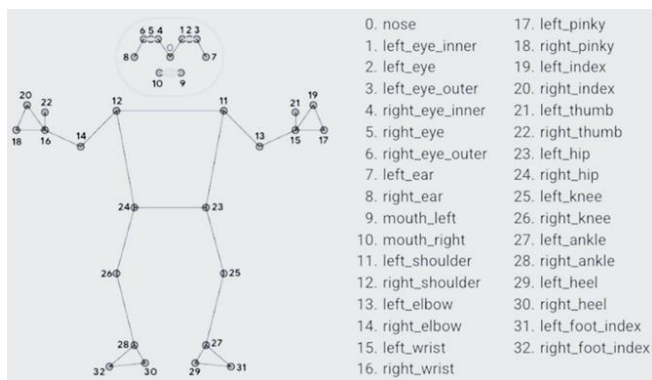


Рис. 1. Представление ключевых точек человека

Каждая из представленных точек имеет характеристики:

- Координата  $x$  – горизонтальное положение точки на изображении. Значение координаты может быть любым числом в пределах ширины изображения, где 0 обычно соответствует крайней левой точке, а максимальное значение – крайней правой точке.

- Координата  $y$  – вертикальное положение точки на изображении. Значение координаты может быть любым числом в пределах высоты изображения, где 0 обычно соответствует верхней точке, а максимальное значение – нижней точке.
- Координата  $z$  – глубина точки (ее удаленность от камеры). Значение может использоваться для определения трехмерного положения точки в пространстве. Значение координаты может быть положительным или отрицательным числом.
- Параметр  $visibility$  (видимость точки) равен 0, если указанной точки не видно на изображении, 1 – если она присутствует.

Используя возможности данного фреймворка, можно создать приложение, способное сравнивать позу пользователя, считываемую с изображения с камеры, с эталонным исполнением позы.

Для сравнения двух поз на изображениях нельзя просто сравнивать координаты ( $x, y, z$ ) всех 33 точек, так как при небольшом сдвиге все точки выделены как некорректные даже при идеальном попадании в эталонную позу. Чтобы этого избежать, необходимо сравнивать полученные модели, следуя алгоритму, приведенному ниже:

1. Из моделей необходимо изъять следующие точки и их координаты ( $x, y$ ): левое плечо, правое плечо, левое бедро, правое бедро (точки 1, 2, 3 и 4 соответственно).
2. Необходимо найти точки  $C_{12}$  и  $C_{34}$  таким образом, чтобы точка  $C_{12}$  была центром точек 1 и 2, а  $C_{34}$  – центром точек 3 и 4.
3. Перебираем все ключевые точки модели. Каждой точке пользователя сопоставляем такую же точку эталонной модели.

Правило: при сравнении верхней части тела (голова, руки, плечи) используется точка  $C_{12}$ , при сравнении нижней –  $C_{34}$ . Поскольку как начало отрезка выбрана точка  $N$ , относящаяся к верхней части тела, то концом отрезка будет точка  $C_{12}$ . Таким образом получим две прямые, проходящие через отрезки  $(C_{12} N)_п$  и  $(C_{12} N)_э$ .

Имея две прямые, можем найти угол, которые образуются между ними –  $\alpha$ . Сравним этот угол с углом нормального отклонения –  $\text{angular\_deviation}$  (максимальное отклонение =  $15^\circ$ ), можем сделать вывод, в правильной ли позиции находится ключевая точка или нет.

На основе описанных выше решений был реализован программный продукт, который оценивает качество исполнения пользователем асан из йоги. Приложение постоянно отслеживает положение ключевых точек у пользователя, чтобы при заметном расхождении их с эталоном, они подсвечивались, и пользователь мог это заметить на расстоянии и поправить свое положение тела. Для демонстрации работы приложения можно обратиться к рис. 2 и рис. 3.

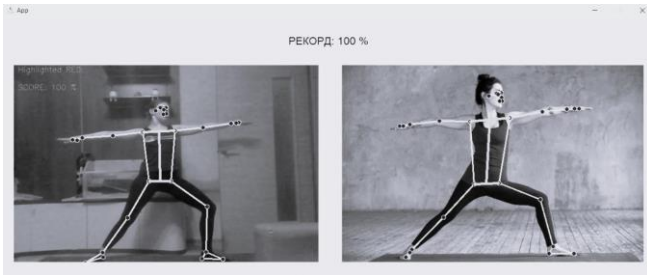


Рис. 2. Правильное исполнение асаны

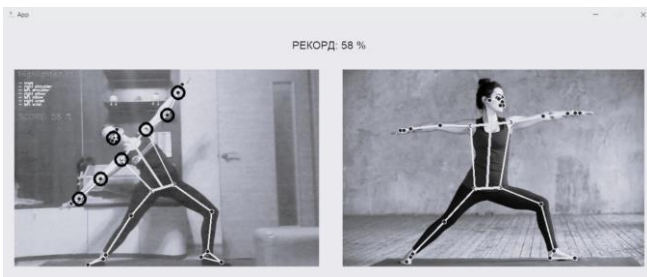


Рис. 3. Неправильное исполнение асаны

#### IV. ЗАКЛЮЧЕНИЕ

В результате работы было собрано приложение, которое способствует росту физической активности пользователя за счет своей простоты в использовании (нужна всего лишь рабочая камера), мгновенной обратной связи (при неправильном выполнении техники приложение мгновенно укажет пользователю на ошибку) и гибкости графика, ведь его выбирает сам пользователь.

#### СПИСОК ЛИТЕРАТУРЫ

- [1] Kalgotra P., Raja U., Sharda R. Growth in the development of health and fitness mobile apps amid COVID-19 pandemic // *Digital Health*. – 2022. Т. 8. С. 20552076221129070.
- [2] Lee J. C., Lin R. The continuous usage of artificial intelligence (AI)-powered mobile fitness applications: the goal-setting theory perspective // *Industrial Management & Data Systems*. 2023. Т. 123. №. 6. С. 1840-1860.
- [3] InfiGro Infivolve, URL: <https://www.infivolve.com>\_(дата обращения 03.05.2024)
- [4] DeepSport Athletic – AI Athlete Development - DeepSport, URL: <https://deepsportapp.com>
- [5] ALFA AI: Artificial Intelligence Fitness Coach, URL: <https://www.alfa-ai.com>
- [6] YOLO: Real-Time Object Detection, URL: <https://pjreddie.com/yolo>
- [7] Mediapipe, URL: <https://developers.google.com/mediapipe>
- [8] COCO - Common Objects in Context, URL: <https://cocodataset.org>

# Применение гистограммы длин хорд с целью повышения точности распознавания лёгочных узлов на компьютерных томограммах

Д. С. Смирнова<sup>1,2</sup>, Е. И. Пчицкая<sup>1</sup>, В. С. Чуканов<sup>1,2</sup>

<sup>1</sup>Лаборатория анализа биомедицинских изображений и данных  
Санкт-Петербургский политехнический университет Петра Великого

<sup>2</sup>Высшая школа прикладной математики и вычислительной физики  
Санкт-Петербургский политехнический университет Петра Великого

daria.welt@bk.ru

**Аннотация.** Развитие технологий компьютерной томографии в области медицинской диагностики легочных узлов приводит к необходимости разработки эффективных методов классификации и сегментации. В данном исследовании оценивается применение гистограммы длин хорд (CLDH) в сочетании с методами машинного обучения, такими как SVR, случайный лес и градиентные бустинги, для анализа трехмерной геометрии легочных узлов. Полученные результаты свидетельствуют о том, что CLDH является перспективным дескриптором формы, обеспечивая высокую точность классификации легочных узлов. В сравнении с нейросетевыми подходами, метод, основанный на CLDH, эффективен даже при ограниченном объеме обучающей выборки, при этом потребляя меньше вычислительных ресурсов. Предложенный подход к описанию морфологии легочных узлов продемонстрировал хорошую разделимость данных и точность классификации 0.8 на наборе данных LUNA16. Полученные результаты подчеркивают значимость использования классических методов машинного обучения в анализе трехмерной геометрии для повышения точности диагностики рака легких.

**Ключевые слова:** компьютерная томография; легочные узлы; сегментация; гистограмма длин хорд; классификация; пространство признаков; представление признаков; рак легких

## I. ВВЕДЕНИЕ

Рак легких – серьезная социальная и медицинская проблема, поскольку это один из самых часто встречающихся типов злокачественных опухолей. Несмотря на наблюдаемую тенденцию снижения общей смертности от рака и, в частности, от рака легких, рак по-прежнему остается второй основной причиной смерти, при этом рак легких продолжает занимать ведущее положение по доле летальных исходов среди онкологических заболеваний [1]. Для эффективного лечения требуется применение инновационных подходов для раннего выявления заболевания и точной диагностики. В последние годы появление передовых технологий медицинской визуализации открыло новые возможности для точной характеристики легочных патологий [2–4]. Особое внимание в данной задаче

уделяется легочным узелкам, определяемым как небольшие округлые образования размером от 2 до 30 мм, которые обнаруживаются на снимках компьютерной томографии (КТ) при рутинной визуализации. Эти структуры представляют собой серьезную проблему при анализе КТ-снимков из-за их потенциальной связи со злокачественными новообразованиями, которая коррелирует с размером узелков [5, 6]. В случае небольших размеров легочного узелка задача определения злокачественности образования усложняется из-за их визуальной схожести с доброкачественными образованиями. Более того, строгого математического критерия злокачественности узелков не существует, и для их классификации применяются методы машинного и глубокого обучения, способные «выучить» данный критерий неявно, на основе данных [7, 8].

Пересечение технологий медицинской визуализации и вычислительного анализа представляет собой благоприятные условия для междисциплинарного сотрудничества, объединяя экспертизу в области науки о данных, машинного обучения, математики и биологии. В данном контексте перспективным направлением представляется анализ и разработка новых дескрипторов формы для анализа трехмерного представления узелков. Это не только соответствует более широкой задаче повышения точности диагностики, но и отвечает потребностям в неинвазивных и эффективных методах классификации легочных узелков. Традиционные подходы к классификации легочных уголков часто не позволяют учесть все тонкости их геометрии и формы. Усовершенствованные дескрипторы формы способны раскрыть морфологические закономерности, которые могут быть недоступны при использовании традиционных методов диагностики [9, 10]. Внедрение таких дескрипторов позволит преодолеть данные ограничения и достичь более качественного описания морфологии узлов, выявить тонкие вариации формы, которые могут свидетельствовать о злокачественности исследуемого образования, что в конечном итоге приведет к повышению точности классификации, и будет способствовать раннему выявлению рака и, как следствие, улучшению прогноза лечения легочных патологий.

Работа поддержана грантом в рамках государственного задания FSEG-2024-0025.



Ранее исследователями предлагалось использовать данный метод описания формы в сочетании с другими статистическими методами в качестве входных данных нейронной сети [11, 20]. Данное исследование рассматривает гистограмму длин хорд (chord length distribution histogram, CLDH) в качестве самостоятельного признака для классификации и его эффективность на одном из популярных наборов данных LUNA16.

Как правило для моделей оценки риска развития рака легких используют такие характеристики как возраст пациента, диаметр узелка, информация является ли человек курящим и было ли в истории его семьи это заболевание, пол человека и положение узелка. Среди таких моделей – VA модель, модель Брока и модель Свенсона, которые могут классифицировать узелки с точностью до 80 % [21]. Гистограмма длин хорд в свою очередь способна отразить морфологические особенности узлов, что позволяет провести качественный анализ данных, выделяя характерные черты и закономерности, которые могут быть невидимы для человеческого взгляда или традиционных методов.

Настоящее исследование направлено на разработку метода анализа трехмерной геометрии легочных узелков с помощью метода гистограммы длин хорд [11] с целью повышения точности классификации новообразований на КТ-изображениях методами машинного обучения. Мы показали, что на открытом наборе данных LUNA16 [12] данный метод с высокой точностью выявляет легочные узелки на КТ снимках и превосходит нейросетевой подход по качеству и ресурсоэффективности. Описание алгоритмов, использованных при описании данных КТ снимков, приведено в разделе «Методы» и включает в себя алгоритмы сегментации и получения значений гистограмм длин хорд. Исследование качества используемого метода для описания данных и его сравнение с нейросетевым подходом представлено в разделе «Описание экспериментов». Выводы приведены в разделе «Обсуждение результатов».

## II. МЕТОДЫ

Для вычисления дескриптора формы объекта по КТ снимку необходимо произвести его сегментацию и реконструировать его форму по полученной трехмерной воксельной сетке.

### A. Сегментация легочных узелков

В данном исследовании разработан алгоритм сегментации узелков, основанный на алгоритме водоразлива [13] для автоматизированного извлечения полигональной сетки из стопки изображений по z плоскостям. Первым шагом является удаление шума на изображении с помощью медианного фильтра с размером ядра 3. Далее по изображению вычисляется порог бинаризации алгоритмом Ли [14] и производится его бинаризация (рис. 1а, б).

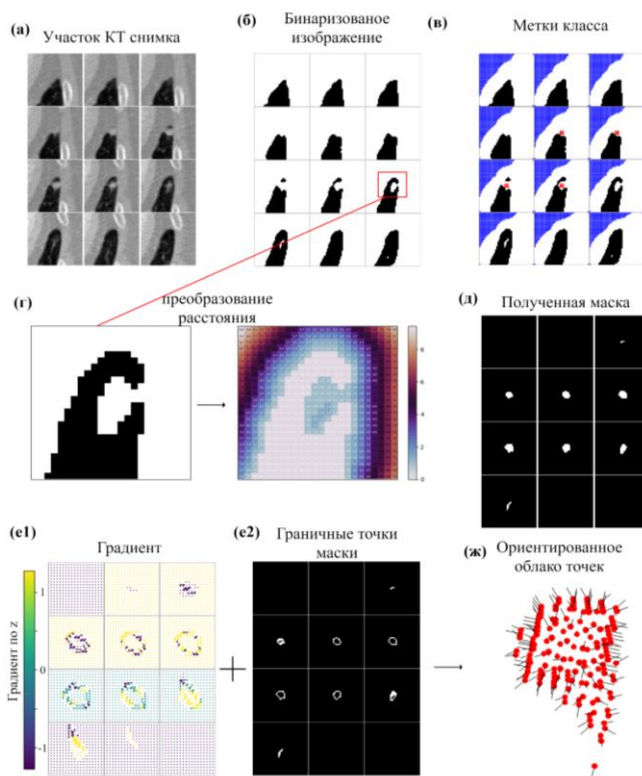


Рис. 1. Сегментация образования на КТ снимке LUNA16; (а) - исходная стопка снимков легочного образования изображенная по z плоскостям; (б) - результат бинаризации исходного снимка; (в) - расположение меток классов образования (красная метка) и фона (синяя метка) на бинарном изображении; (г) - преобразование расстояния на бинарном изображении для получения матрицы расстояний; (д) - маска легочного образования, полученная методом водораздела с использованием матрицы расстояния, бинарного изображения и меток классов; (е1) и (е2) - преобразования маски легочного образования: получение градиента по бинарной маске и граничных точек маски соответственно; (ж) - результат сегментации снимка: ориентированное облако точек для получения поверхности, в котором вершины являются граничными точками маски, а нормали вычисляются по градиенту по маске

По бинарному изображению вычисляется преобразование расстояния, при котором формируется матрица, где каждое значение соответствует расстоянию до ближайшей граничной точки, то есть до точки с нулевым значением (рис. 1г) [15]. Вычисление матрицы расстояний позволяет сегментировать узелки, которые образованы на стенках легких и по яркости сливаются с изображением стенки.

Изображения для сегментации получены таким образом, что отмеченная экспертом точка интереса, принадлежащая легочному узелку, расположена в центре, поэтому в центральную точку устанавливается метка класса для рассматриваемого новообразования, которая будет играть роль маркера для алгоритма водораздела. Для отделения узелков от стенок легких и других ярких объектов в крайние точки изображений и в точки, в которых расстояние до границы больше ширины легочного узелка, установлены маркеры класса фона, другими словами объекта «не узелок» (рис. 1в).



Рис. 2. Алгоритм получения гистограммы длин хорд поверхности

Полученные данные: матрица расстояний до границы и матрица меток классов передаются в алгоритм водораздела, который позволяет получить сегментированное изображение. В изображении пиксели со значением равным метке класса узелков являются пикселями, принадлежащими легочному узелку, по этим значениям извлекается маска для легочного узелка (рис. 1д).

### В. Извлечение признаков объекта по маске

Получение поверхности по стопке бинарных изображений, представляющих собой маску легочного узелка, производится алгоритмом реконструирования поверхности. Граничные воксели маски представляют собой облако точек (рис. 1е2, ж)), из которых алгоритмом Пуассона извлекается полигональная сетка [16]. Нормали в данных точках для передачи их в алгоритм Пуассона вычисляются по данным градиента, полученного по трехмерной маске рассматриваемого узелка (рис. 1е1).

В качестве комплексного дескриптора формы предлагается использовать гистограмму длин хорд поверхности [11]. Ранее исследователями был предложен алгоритм получения длин хорд изображения силуэта, в котором хорды извлекаются по набору углов наклона к оси  $x$  и смещений по оси  $y$  [17]. Также вычисление метрики CLDN с использованием случайной выборки хорд использовалось для определения сходства фигур на плоскости [18].

Наша реализация метода направлена на работу с трехмерными фигурами и использует стохастический подход. Из набора случайно взятых по объекту хорд извлекается распределение длин отрезков хорд, содержащихся внутри исследуемого легочного узелка, и представляется в виде гистограммы (рис. 2). Таким образом, алгоритм расчета параметризуется количеством извлекаемых хорд и количеством столбцов выходной гистограммы. Хорды представляют собой отрезок, соединяющий две случайно выбранных точки поверхности воксельной сетки, все точки которого принадлежат воксельной сетке, то есть лежат внутри формы. После получения требуемого количества хорд вычисляются их длины, далее в массиве длин хорд находится максимальное значение, и длины хорд нормализуются делением на него. Таким образом, унифицируется распределение длин хорд различных объектов.

В случае невыпуклой поверхности, случайно выбранные точки на поверхности могут образовать отрезок, содержащий один или несколько участков, находящихся вне рассматриваемой фигуры. Такие отрезки разбиваются на подотрезки, лежащие строго внутри фигуры, которые впоследствии рассматриваются как отдельные хорды.

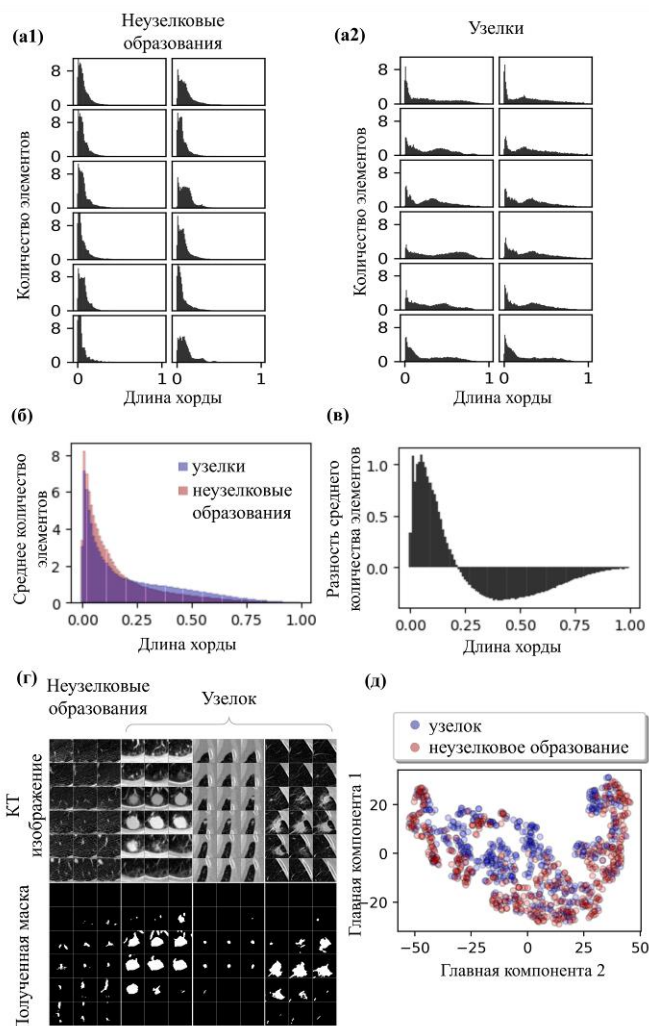


Рис. 3. Анализ данных датасета LUNA16 методом гистограммы длин хорд; (a1) и (a2) - гистограммы длин хорд элементов данных для образований не являющихся узелками и для узелков соответственно; (б) - усредненные по классу гистограммы длин хорд для образований не являющихся узелками и узелков; (в) - график результата вычитания усредненной гистограммы длин хорд неузелковых образований из усредненной гистограммы для класса узелков; (г) - Элементы набора данных LUNA16 с полученными алгоритмом сегментации масками, изображенные по  $z$  плоскостям; (д) - расположение элементов данных в координатах t-SNE анализа для  $n=2$

Гистограмма длин хорд нормализована таким образом, чтобы представлять собой плотность вероятности распределения, что позволяет проводить попарное сравнение морфологий: более вытянутые формы с наличием цилиндрических участков будут иметь большее количество коротких хорд, в то время как объекты, близкие к шару, будут состоять преимущественно из длинных хорд.

## III. ОПИСАНИЕ ЭКСПЕРИМЕНТОВ

### А. Описание набора данных

Для проведения экспериментов и оценки прикладного значения гистограммы длин хорд в области классификации легочных новообразований использован датасет LUNA16 [12], который содержит более 1 тыс. КТ-изображений с легочными узлами, проанализированными рентгенологами. Для извлечения

поверхностей и расчета данных по дескриптору формы были отфильтрованы наиболее качественные маски, полученные алгоритмом водораздела и представляющие форму узлов повоксельно, а не приближительную область локализации образования. Было получено 1200 полигональных сеток легочных узлов, наилучшим образом отражающих как образования, являющиеся легочными узелками, так и ложные (неузелковые) образования.

По извлеченным полигональным сеткам вычислены гистограммы длин хорд. Усредненное по классам представление гистограмм показывает, что легочные узелки имеют большую концентрацию длинных хорд по сравнению с ложными образованиями (рис. 3).

#### В. Результаты классификации легочных образований

Для оценки эффективности рассматриваемого дескриптора формы произведено сравнение классической модели классификации по 100 признакам, отвечающим значениям частоты в гистограмме, и нейросетевой модели. Набор данных из 1200 наблюдений был разбит на два подмножества – тренировочные данные составили 80 % общего числа данных, тестовые данные – 20 %.

Для классификации образований с использованием гистограммы длин хорд были использованы методы:

- случайный лес;
- XGBoost;
- машина опорных векторов с различными ядерными функциями.

Параметры моделей подбирались по результатам кросс-валидации. На рассматриваемом наборе данных значение точности достигло 0,80 для модели случайного леса. XGBoost также показал высокую точность классификации (табл. 1).

ТАБЛИЦА I. СРАВНИТЕЛЬНЫЙ АНАЛИЗ ЭФФЕКТИВНОСТИ МОДЕЛЕЙ КЛАССИФИКАЦИИ

SVC laplacian	0,79	0,76	0,87	0,67	1,82 ± 0,13	0,022 ± 0,005
SVC gauss	0,77	0,75	0,84	0,67	1,55 ± 0,41	0,023 ± 0,003
SVC linear	0,61	0,60	0,61	0,59	13,50 ± 1,07	0,021 ± 0,001
SVC PCA (n=2)	0,70	0,71	0,69	0,72	2,31 ± 0,16	0,021 ± 0,004
XGBoost	0,79	0,77	0,70	0,86	2,61 ± 0,88	0,004 ± 0,001
Random Forest	0,80	0,79	0,84	0,74	1,49 ± 0,06	0,025 ± 0,004
ResNet3d	0,73	0,76	0,69	0,85	15 960	54,910 ± 3,120

Для сравнения метода классификации по гистограммам хорд была взята нейросетевая модель ResNet50 зарекомендовавшая себя в качестве модели классификации как изображений общего содержания, так и непосредственно медицинских [19]. Обучение модели на тех же данных и том же разбиении на тренировочную и тестовую выборки позволило достичь точности классификации 0,73 и потребовало в 4 раза больше времени чем подготовка гистограмм длин хорд и

обучение машины опорных векторов. Наиболее ресурсоемкой операцией при использовании дескриптора CLDH в задаче классификации является подготовка данных, в то время как обучение модели заняло 2–13 секунд для рассматриваемого набора данных, что составляет менее 0,1 % от времени, затраченного на подготовку данных. Сравнение эффективности вычислений для моделей производилось на машине с процессором 11th Gen Intel (R) Core (TM) i7-1165G7 2.80GHz и с 32Гб ОЗУ.

## IV. ЗАКЛЮЧЕНИЕ

### А. Обсуждение результатов экспериментов

Исследование гистограммы длин хорд (CLDH) в контексте анализа трехмерной геометрии легочных узлов выявило ее высокий потенциал в качестве дескриптора формы для классификации легочных узлов. Рассмотренный дескриптор гистограммы длин хорд представил данные легочных узлов в оптимальной форме, позволяя провести классификацию образований на узелковые и не узелковые с результатами, не уступающими методам глубокого обучения. Нормализация, осуществляемая на заключительном этапе вычисления CLDH, обеспечивает однородное представление разнообразных выборок длин хорд, соответствующих различным объектам.

Таким образом, представление трехмерных форм в форме гистограммы длин хорд позволяет получить векторное представление объекта заданной размерности.

Это открывает возможность вместо ресурсоемких методов глубокого обучения использовать классические менее ресурсозатратные методы машинного обучения, такие как метод ближайших соседей, метод опорных векторов и другие. При этом векторное представление морфологии в виде гистограммы не только обеспечивает высокую точность классификации, но также предоставляет возможность произвести визуальную оценку и проинтерпретировать результаты, установив связь полученных данных с предметной областью. В области анализа легочных узлов имеются методы, позволяющие описывать морфологические особенности образований в текстовой форме [22]. Гистограмма длин хорд требует участие человека для оценки формы, но при этом обеспечивает более широкий и гибкий спектр возможностей для интерпретации данных.

Использование метода CLDH в сочетании с классификацией гистограмм моделью случайного леса продемонстрировало наилучшие результаты на датасете LUNA16, превосходя по точности определения достоверности наблюдения легочных узлов нейронную сеть с архитектурой Resnet-50 и классический метод классификации – машину опорных векторов с различными ядрами. Это может быть особенно значимо, учитывая, что объем данных обучающей выборки может быть недостаточным для полноценного обучения нейронных сетей, в то время как метод, основанный на гистограмме длин хорд, оказывается эффективным даже на относительно небольшом объеме данных. При этом использование методов машинного обучения позволяет производить обучение модели и предсказание класса более ресурсоэффективно по сравнению с подходами,

основанными на нейронных сетях. Это позволило получить более точный результат за меньшее время, что может быть особенно актуально в условиях ограниченности обучающих данных и временных ресурсов.

#### *В. Дальнейшие перспективы использования гистограммы длин хорд для описания легочных узлов*

В современной литературе активно исследуются различные методы для обнаружения легочных узлов для повышения качества диагностики. Большая часть работ посвящена совершенствованию нейросетевых методов сегментации и классификации [21]. Последние наиболее точные модели сосредоточены на улучшении качества за счет введения подобранных вручную фильтров и признаков в систему диагностики [23], а также использования дополнительных моделей, которые компенсируют недостатки основной системы [24]. Таким образом, наибольшую эффективность показывают комплексные системы обнаружения и диагностики легочных узелков, в которые внедряются дополнительные модели и векторные признаки. Одним из таких признаков мы рассматриваем гистограмму длин хорд, которая может эффективно использоваться в качестве отдельного дескриптора в сочетании с методами машинного обучения. Преимуществом данного подхода является меньшая требовательность системы к размеру обучающего набора данных и вычислительным ресурсам, что делает возможным относительно недорогое внедрение метода в существующую систему в качестве одного из элементов ансамбля моделей.

Использование гистограммы длин хорд в качестве дескриптора формы в данной работе успешно решает задачу классификации легочных узлов, предоставляя перспективное решение в области анализа трехмерной геометрии для целей медицинской диагностики. В сочетании с новейшими подходами к сегментации легочных узелков на КТ снимках, таким как Multi-crop CNN [25], демонстрирующим высокую точность, данный дескриптор может быть полезен для автоматизированной диагностики.

#### СПИСОК ЛИТЕРАТУРЫ

- [1] R. L. Siegel, K. D. Miller, N. W. Sandeep, W. Mbbs, and A. Jemal, "Cancer statistics, 2023," *CA Cancer J Clin*, vol. 73, pp. 17–48, 2023, doi: 10.3322/caac.21763.
- [2] G. Zhang, S. Jiang, Z. Yang, L. Gong, X. Ma, Z. Zhou, C. Bao, Q. Liu., "Automatic nodule detection for lung cancer in CT images: A review," *Comput Biol Med*, vol. 103, pp. 287–300, Dec. 2018, doi: 10.1016/J.COMPBIOMED.2018.10.033.
- [3] F. Binczyk, W. Prazuch, P. Bozek, and J. Polanska, "Radiomics and artificial intelligence in lung cancer screening," *Transl Lung Cancer Res*, vol. 10, no. 2, pp. 1186–1199, Feb. 2021, doi: 10.21037/tlcr-20-708.
- [4] Y. Gu, J. Chi, J. Liu, L. Yang, B. Zhang, D. Yu, Y. Zhao, X. Lu., "A survey of computer-aided diagnosis of lung nodules from CT scans using deep learning," *Comput Biol Med*, vol. 137, p. 104806, Oct. 2021, doi: 10.1016/J.COMPBIOMED.2021.104806.
- [5] M.K. Gould, J. Donington, W.R. Lynch, P.J. Mazzone, D.E. Midthun, D. P. Naidich, R. S. Wiener, "Evaluation of Individuals With Pulmonary Nodules: When Is It Lung Cancer?: Diagnosis and Management of Lung Cancer, 3rd ed: American College of Chest Physicians Evidence-Based Clinical Practice Guidelines," *Chest*, vol. 143, no. 5, pp. e93S-e120S, May 2013, doi: 10.1378/CHEST.12-2351.
- [6] I. Vlahos, K. Stefanidis, S. Sheard, A. Nair, C. Sayer, and J. Moser, "Lung cancer screening: nodule identification and characterization," *Transl Lung Cancer Res*, vol. 7, no. 3, p. 288, Jun. 2018, doi: 10.21037/TLCR.2018.05.02.
- [7] S. Fra-Fernández, L. Gorospe-Sarasúa, A. Cabañero-Sánchez, G. Muñoz-Molina, U. Caballero-Silva, and N. Moreno-Mata, "Subsolid pulmonary nodules: why not 'watch and wait'?" // *Ann Transl Med*, vol. 12, no. 1, pp. 3–3, Feb. 2024, doi: 10.21037/ATM-23-1794.
- [8] B.-G. Kim and S.-W. Um, "A narrative review of the clinical approach to subsolid pulmonary nodules," *Ann Transl Med*, vol. 11, no. 5, pp. 217–217, Mar. 2023, doi: 10.21037/ATM-22-5246.
- [9] H. Laga, "A survey on nonrigid 3d shape analysis," *Academic Press Library in Signal Processing*, Volume 6, pp. 261–304, 2018.
- [10] A. K. Dhara, S. Mukhopadhyay, A. Dutta, M. Garg, and N. Khandelwal, "A combination of shape and texture features for classification of pulmonary nodules in lung CT images," *J Digit Imaging*, vol. 29, pp. 466–475, 2016.
- [11] A. A. Meldo and L. V Utkin, "A new approach to differential lung diagnosis with CT scans based on the Siamese neural network," *J Phys Conf Ser*, vol. 1236, no. 1, p. 012058, Jun. 2019, doi: 10.1088/1742-6596/1236/1/012058.
- [12] "Lung Nodule Analysis 2016 (LUNA16) Grand Challenge." Accessed: Mar. 21, 2024. [Online]. Available: <https://luna16.grand-challenge.org/Home/>
- [13] A. Kornilov, I. Safonov, and I. Yakimchuk, "A Review of Watershed Implementations for Segmentation of Volumetric Images," *Journal of Imaging 2022*, Vol. 8, Page 127, vol. 8, no. 5, p. 127, Apr. 2022, doi: 10.3390/JIMAGING8050127.
- [14] C. H. Li and P. K. S. Tam, "An iterative algorithm for minimum cross entropy thresholding," *Pattern Recognit Lett*, vol. 19, no. 8, pp. 771–776, Jun. 1998, doi: 10.1016/S0167-8655(98)00057-9.
- [15] M. Manduhu and M. W. Jones, "A Work Efficient Parallel Algorithm for Exact Euclidean Distance Transform," *IEEE Transactions on Image Processing*, vol. 28, no. 11, pp. 5322–5335, Nov. 2019, doi: 10.1109/TIP.2019.2916741.
- [16] M. Kazhdan and H. Hoppe, "Screened poisson surface reconstruction," *ACM Trans Graph*, vol. 32, no. 3, pp. 1–13, Jun. 2013, doi: 10.1145/2487228.2487237.
- [17] Y. Mingqiang, K. Kidiyo, and R. Joseph, "Chord Context Algorithm for Shape Feature Extraction," in *Object Recognition, InTech*, 2011. doi: 10.5772/14979.
- [18] S. P. Smith and A. K. Jain, "Chord distributions for shape matching," *Computer Graphics and Image Processing*, vol. 20, no. 3, pp. 259–271, Nov. 1982, doi: 10.1016/0146-664X(82)90084-3.
- [19] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, Jun. 2016, pp. 770–778. doi: 10.1109/CVPR.2016.90.
- [20] L. Utkin, A. Meldo, M. Kovalev, and E. Kasimov, "An Ensemble of Triplet Neural Networks for Differential Diagnostics of Lung Cancer," *Conference of Open Innovation Association, FRUCT*, pp. 346–352, Nov. 2019, doi: 10.23919/FRUCT48121.2019.8981542.
- [21] C. de Margerie-Mellon and G. Chassagnon, "Artificial intelligence: A critical review of applications for lung nodule and lung cancer," *Diagn Interv Imaging*, vol. 104, no. 1, pp. 11–17, Jan. 2023, doi: 10.1016/J.DIII.2022.11.007.
- [22] A. Meldo, L. Utkin, M. Kovalev, and E. Kasimov, "The natural language explanation algorithms for the lung cancer computer-aided diagnosis system," *Artif Intell Med*, vol. 108, p. 101952, Aug. 2020, doi: 10.1016/J.ARTMED.2020.101952.
- [23] H.-H. Chang, Cheng-Zhe Wu, Audrey, and H. Gallogly, "Pulmonary Nodule Classification Using a Multiview Residual Selective Kernel Network," *Journal of Imaging Informatics in Medicine 2024 37:1*, vol. 37, no. 1, pp. 347–362, Jan. 2024, doi: 10.1007/S10278-023-00928-4.
- [24] J. Zhang, W. Zou, N. Hu, B. Zhang, and J. Wang, "S-Net: an S-shaped network for nodule detection in 3D CT images," *Phys Med Biol*, Feb. 2024, doi: 10.1088/1361-6560/AD2B96.
- [25] B.C. Sweetline, C Vijayakumaran, A Samyudurai, "Overcoming the Challenge of Accurate Segmentation of Lung Nodules: A Multi-crop CNN Approach," *Journal of Imaging Informatics in Medicine 2024*, pp. 1–20, Feb. 2024, doi: 10.1007/S10278-024-01004-1.

# Модель оценки качества воды, основанная на AI: данные длительного наблюдения

Е. А. Алтай, Л. Базарбай

*Satbayev University*

*г. Алматы, Казахстан*

lashyn\_7754@mail.ru

**Аннотация.** В настоящей статье представлены результаты разработки модели оценки качества воды с использованием метода искусственного интеллекта. Представленная модель основана на линейной регрессии, которая при оценке выявила статистически значимую взаимосвязь между совмещенными показателями качества воды. Выявлено, что среди измеренных показателей наиболее влиятельным предиктором на кислотность, проводимость, мутность и окислительно-восстановительный потенциал является температура воды. Обнаруженная взаимосвязь между измеренными показателями в основном ассоциирована с влиянием температуры на физические и химические процессы, происходящие при повышении и понижении температуры речной воды.

**Ключевые слова:** контроль качества воды; искусственный интеллект; линейная регрессия; температура; кислотность; мутность; кондуктивность

## I. ВВЕДЕНИЕ

Вода является одним из основных жизненно важных элементов в живых системах и ресурсом для наземных и водных экосистем. Влияние различных факторов, а именно антропогенных и экологических, может оказывать существенное изменение биологических, физических и химических свойств воды, снижая ее качество. Учитывая это, в настоящее время контроль качества воды осуществляется кратковременно в лабораторных условиях и на постоянной основе с помощью специализированных измерительных систем, включающих в себя устройства интернета вещей (IoT) для дистанционной передачи измеренных данных. Среди перечисленных видов контроля качества воды именно длительный контроль может выявить ряд значимых изменений, которые не могут быть своевременно обнаружены при кратковременном контроле. В связи с этим длительный контроль качества воды в режиме реального времени является важной научной задачей, требующей комплексного подхода к решению.

Базируясь на принципах формирования комплексного подхода, в данной работе для разработки модели оценки качества воды осуществляется длительный контроль с использованием современных датчиков, позволяющих измерять температуру ( $t$ ), мутность (ЕМФ), кислотность (рН), электропроводность (С) и окислительно-восстановительный потенциал (ОВП,  $e$ ). Важно отметить, что совместное измерения пяти показателей для контроля качества воды ранее не получили должной проработки, что подтверждается

научными результатами, опубликованных в работах [1]–[15]. В основном это можно связать с тем, что для совместной оценки качества воды используют датчики кислотности, мутности и проводимости, а другие исследователи совмещают только температуру, без датчика ОВП. Совместная оценка показателей воды вместе с ОВП позволит определить содержание кислорода и разложение загрязняющих веществ [1, 7]. Таким образом, оценка показателя ОВП вкупе с перечисленными выше показателями, а также их комбинирование для контроля качества воды являются важными и необходимыми, так как позволят комплексно оценить показатели и выявлять значимые ассоциации между ними при разработке модели оценки качества воды.

Разработка модели оценки качества воды при контроле показателей осуществляется с применением методов искусственного интеллекта. Известно, что использование методов искусственного интеллекта дает возможность повысить достоверность и разработать прогностическую модель контролируемого объекта при оценке качества воды. Согласно опубликованным данным [2, 8] среди методов искусственного интеллекта для разработки прогностической модели оценки качества воды используются линейная и нелинейная регрессия, а также их комбинация с методом эмпирической модовой декомпозиции. Такая гибридная реализация позволяет разработать математическую модель, характеризующую связь физических и химических показателей воды. Вместе с тем следует отметить, что метод эмпирической модовой декомпозиции и его различные модификации при разложении измеренных данных могут вызывать эффект смешивания мод, обусловленный «краевыми эффектами» [16, 17], что может существенно усложнить процесс разработки модели оценки на основе метода искусственного интеллекта.

В рассматриваемой статье предлагается модель оценки качества воды на основе метода искусственного интеллекта – линейной регрессии. Данный метод является эффективным и высокочувствительным к анализируемому данным, не требующим специального специфического обучения выборок из различных объектов контроля [18, 19]. Таким образом, целью исследования рассматриваемой статьи заключается в разработке математической модели регрессионной зависимости, прогнозирующей качество воды и ассоциацию совокупных показателей.

## II. СБОР ДАННЫХ СОВМЕЩЕННЫХ ИЗМЕРЕНИЙ

В данном разделе статьи для совместной оценки показателей разработана аппаратная часть измерительной системы, реализующая совмещенное измерение температуры, мутности, кислотности, электропроводности и окислительно-восстановительного потенциала воды на платформе микроконтроллера ATmega2560 и обеспечивает беспроводную передачу измеренных данных на специальное мобильное приложение Blynk через облачный сервис и IoT-устройства. Результат разработки измерительной системы для сбора и дистанционной передачи данных совмещенных показателей воды представлены на рис. 1.



Рис. 1. Система для сбора и передачи данных совмещенных измерений.

Важно отметить, что измерительная система, представленная на рис. 1 в отличие от известных аналогов, позволяет проводить совмещенное измерение пяти показателей воды из различных объектов контроля. При апробации данной системы сбор данных измерений осуществлялся на основе трехмесячного наблюдения за показателями воды в акватории реки Алматы на территории города Алматы. Важно отметить, что первое наблюдение проводилось с 1 сентября по 30 сентября, второе наблюдение проводилось с 1 октября по 31 октября, а третье наблюдение проводилось с 1 по 30 ноября, соответственно. Каждое из наблюдений состоит более чем  $n = 10000$  измерений, и имеют различие в значениях измерения. Сформированный таким образом набор данных совмещенных измерений способствует разработке математической модели и выявлению изменений качества воды, из которой они были получены.

Известно, что при разработке регрессионной модели наиболее чувствительным и прогностическим показателем процесса изменения свойств веществ и материалов является среднее значение результатов измерений. В связи с этим в рассматриваемой работе для разработки линейной регрессионной модели оценки качества воды обоснован выбор именно среднего значения результатов, состоящих более чем  $n = 10000$  измерений. Результат разработки модели представлен в следующем разделе данной статьи.

## III. РЕГРЕССИОННАЯ МОДЕЛЬ ОЦЕНКИ КАЧЕСТВА ВОДЫ

Модель оценки качества воды при регрессионном анализе в зависимости от результатов измерения совмещенных показателей разделяют на линейный, нелинейный и логистический анализ [8]. Перечисленные методы анализа дают возможность разработать модель оценки качества воды с определенной степенью

достоверности, которая может быть количественно охарактеризована. В соответствии с результатами исследования [18, 19], достоверность модели при данном анализе может быть охарактеризована через коэффициент детерминации  $R^2$ , уровень значимости  $p$ -value и связь между наблюдаемой и прогнозируемой величинами через коэффициент корреляции  $r$ .

Результаты расчета продемонстрировали, что среднее значение трехмесячных показателей линейно возрастает и убывает в зависимости от качества речной воды. Полученный результат позволяет обосновать выбор именно линейную регрессию с целью построения регрессионной зависимости и оценки ассоциации совмещенных показателей методом наименьших квадратов. На рис. 2–5. представлены результаты регрессионной зависимости показателей качества воды.

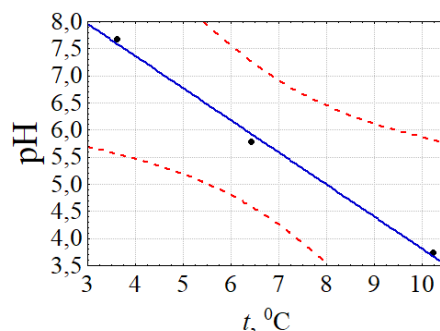


Рис. 2. Зависимость, характеризующая влияние температуры на кислотность.

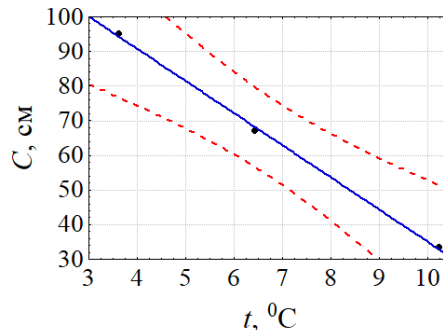


Рис. 3. Зависимость, характеризующая влияние температуры на conductivity.

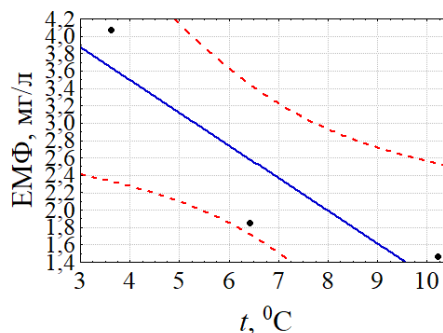


Рис. 4. Зависимость, характеризующая влияние температуры на мутность.

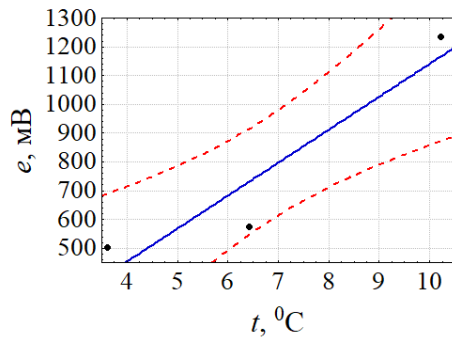


Рис. 5. Зависимость, характеризующая влияние температуры на ОВП.

Из представленных результатов зависимостей на рис. 2–5. следует, что в силу повышения температуры анализируемых проб речной воды убывает значения показателей кислотности, проводимости и мутности воды, следовательно, увеличивается значение окислительного восстановительного потенциала воды в зависимости от времени наблюдения.

Регрессионная модель, характеризующая влияния предиктора – температуры на показатели кислотности, проводимости, мутности и ОВП анализируемых проб речной воды описана линейным уравнением регрессии и рассчитаны коэффициенты детерминации  $R^2$  и корреляции  $r$ , а также уровня значимости  $p$  модели, как:

$$pH = -0,5922 \cdot t + 9,732 \text{ при } R^2 = 0,995; r = -0,997; p = 0,0414 \text{ (см. рис. 2);}$$

$$C = -9,283 \cdot t + 127,885 \text{ при } R^2 = 0,998; r = -0,999; p = 0,022 \text{ (см. рис. 3);}$$

$$EMФ = -0,376 \cdot t + 5,003 \text{ при } R^2 = 0,791; r = -0,889; p = 0,302 \text{ (см. Рис. 4);}$$

$$e = 114,571 \cdot t + 4,678 \text{ при } R^2 = 0,884; r = 0,940; p = 0,220 \text{ (см. рис. 5);}$$

Полученный результат регрессионного анализа позволяет спрогнозировать о том, что при увеличении температуры анализируемых проб речной воды на 1 градус Цельсия следует ожидать уменьшение кислотности воды на 99,58 %, проводимости на 99,87 %, мутности на 79,10 %, а также повышение ОВП на 88,49 %, соответственно. Установленная ассоциация довольно близка к линейному изменению и устойчива к влиянию температуры речной воды. Важно отметить, что на долю влияющих прочих факторов при построении регрессионной модели, а именно для кислотности приходится – 0,42 %, проводимости – 0,13 %, мутности 20,9 %, ОВП – 11,51 %.

Важно отметить, что величина коэффициента детерминации  $R^2$  разработанной модели объясняет наблюдаемую дисперсию предиктора (показателя в фокусе данного исследования) при контроле показателей качества воды. Полученные коэффициенты модели с  $p$ -значением, меньше, чем для  $p = 0,05$ , следовательно, считается статистически значимыми для доверительного интервала с надежностью  $P=0,95$ , соответственно. То есть связанные совмещенные показатели могут быть значимыми предикторами контролируемых процессов, происходящих в воде с вероятностью надежности

доверительного интервала. На сегодняшний день, накопленный экспериментальный материал по оценке качества воды подтверждает возможность применение выбранных показателей для проведения предиктивной аналитики состояния проб воды [8].

#### IV. ЗАКЛЮЧЕНИЕ

В настоящей статье представлены результаты разработки математической модели регрессионной зависимости, прогнозирующей качество воды и ассоциацию совокупных показателей. Результаты показали, что по мере увеличения температуры речной воды наблюдалась тенденция к уменьшению взаимосвязи между кислотностью, проводимостью и мутностью при коэффициенте корреляции от  $r = -0,889$  до  $r = -0,999$ . Также следует отметить, что по мере возрастания температуры воды повышается значение показателя ОВП при коэффициенте корреляции  $r = 0,940$ , соответственно.

Полученные результаты частично согласуются с результатами зарубежных исследований, в которых дистанционные измеренные показатели, предложенные в настоящей работе, конкурировали за роль наиболее влиятельных предикторов для модели оценки качества воды.

#### СПИСОК ЛИТЕРАТУРЫ

- [1] Tariq M. O. An Open Source Water Quality Measurement System for Remote Areas // Engineering Proceedings. 2021. № 1. P. 50.
- [2] Zhao Y. Retrieval of water quality parameters based on near-surface remote sensing and machine learning algorithm // Remote Sensing. 2022. №. 21. P. 5305.
- [3] Wood R. Implementation of the WHO core components of an infection prevention and control programme in two sub-saharan African acute health-care facilities: a mixed methods study // Antimicrobial Resistance and Infection Control. 2024. № 1. P. 4.
- [4] Mathur R. Modeling of two-stage anaerobic onsite wastewater sanitation system to predict effluent soluble chemical oxygen demand through machine learning // Scientific Reports. 2024. №. 1. P. 1835.
- [5] Wang Z. Research on water environmental indicators prediction method based on EEMD decomposition with CNN-BiLSTM // Scientific Reports. 2024. №. 1. P. 1676.
- [6] Shah M. I., Javed M. F., Abunama T. Proposed formulation of surface water quality and modelling using gene expression, machine learning, and regression techniques // Environmental Science and Pollution Research. 2021. №. 28. P. 13202-13220.
- [7] de Camargo E.T. Low-cost water quality sensors for IoT: A systematic review // Sensors. 2023. №. 9. P. 4424.
- [8] Essamlali I., Nhaila H., El Khaili M. Advances in machine learning and IoT for water quality monitoring: A comprehensive review // Heliyon. 2024.
- [9] Fadel A. A., Shujaa M. I. Water Quality Monitoring System Based on IOT Platform // IOP Conference Series: Materials Science and Engineering. 2020. №. 3. P. 032054.
- [10] Bogdan R. Low-cost internet-of-things water-quality monitoring system for rural areas // Sensors. 2023. №. 8. P. 3919.
- [11] Lakshmikantha V. IoT based smart water quality monitoring system // Global Transitions Proceedings. 2021. №. 2. P. 181-186.
- [12] Geetha S., Gouthami S. Internet of things enabled real time water quality monitoring system // Smart Water. 2016. №. 2. P. 1-19.
- [13] Shah M. I., Javed M. F., Abunama T. Proposed formulation of surface water quality and modelling using gene expression, machine learning, and regression techniques // Environmental Science and Pollution Research. 2021. №. 28. P. 13202-13220.
- [14] Chowdury M.S. IoT based real-time river water quality monitoring system // Procedia computer science. 2019. №. 155. P. 161-168.

- [15] Sabari M. Water quality monitoring system based on IoT // 2020 IEEE 5th International Conference on Devices, Circuits and Systems. 2020. P. 279-282.
- [16] Altay Y.A. Cascade Notch Filter with a Unity Feedback and Improved Transient Response // V International Conference on Control in Technical Systems. 2023. P. 217-220.
- [17] Y.A. Altay, A. S. Kremlev, "Signal-to-Noise Ratio and Mean Square Error Improving Algorithms Based on Newton Filters for Measurement ECG Data Processing," in Proc. EIConRus, 2021, paper 21.09.04, pp. 1590-1595.
- [18] Fedorov A.V., Altay Y.A., Stepanova K.A., Kuzivanov D.O. The effect of signal-to-noise-ratio value on the error in measuring acoustic emission parameters: statistical assessment. Scientific and technical journal of information technologies, mechanics and optics, 2022, V. 22, pp. 1205-1215.
- [19] Y.A. Altay, A.S. Kremlev, "On the use of the statistical methods for biomedical signals and data processing" in Proc. EIConRus, 2019, paper 19.04.03, pp. 1129-1134.



# Measuring the Impact of Strategic Vigilance in Strategic Intelligence

Zafar Nasser Hussain

*Business Administration Dept.*

*University of Babylon*

Babylon, Iraq

bus.dhafar.naser@uobabylon.edu.i

q

ORCID: 0009-0001-5846-5316

Riyadh Hussein Obyes

*Business Administration Dept.*

*University of Kerbala*

Kerbala, Iraq

Riyadh.h@uokerbala.edu.iq

ORCID: 0000-0002-9498-4658

Alaa Hussein Fadhil

*Business Administration Dept.*

*University of Kerbala*

Kerbala, Iraq

alaa.hussein@uokerbala.edu.iq

ORCID: 0000-0002-5418-5464

Ahmed Abdullah Amanah

*Business Administration Dept.*

*University of Kerbala*

Kerbala, Iraq

Ahmed.a@uokerbala.edu.iq

ORCID: 0000-0001-5092-391X

Mohammed Faez Hasan

*Finance and Banking Sciences Dept.*

*University of Kerbala*

Kerbala, Iraq

alaa.hussein@uokerbala.edu.iq

ORCID: 0000-0002-4579-3214

**Abstract**— This research addressed the role which play strategic vigilance in strategic Intelligence, The rapidly development in economic, political, technological and etc, imposes on business organizations makes several of continuous changes both in organizational structure and business activities, for achieve the necessary sustainability and rapidly respond to all newest things. The problem's research expressed in questions about range of effect strategic vigilance and it's dimensions in strategic Intelligence within simple from employees in Al-Furat General Company for Chemical Industries, The research aims to achieve some of goals according to the main and sub hypothesis, and used questionnaire to collect data and information about this simple, which distributed on simple formulate in ( 200 ) persons included staff ( employee , division manager and department manager ) who works in this company, and for process this data used several statistical methods and represented by a factor of linear correlation of Pearson as well as linear regression. The statistical methods shown a number of results from the most prominent is having a significant effect between strategic vigilance and strategic Intelligence. In the light of discussion the results showed the big importance to strategic vigilance and strategic Intelligence concepts, This will need more care from the company in strategic vigilance because it's effect in strategic Intelligence.

**Keywords**— *strategic vigilance, strategic intelligence, organizations, styling, insert*

## I. INTRODUCTION

Organizations in our modern world need new business values, ways of thinking, new standards and behavior patterns which in turn can lead or promote new practices and ways of working in line with the external and internal environment constantly changing. When the signs of change and development emerge, it turns into a challenge because the organization itself or its members resist this change because of organizational or individual reasons that prevent them from accepting it.

## II. LITERATURE REVIEW

### A. Strategic Vigilance (SV)

Strategic vigilance is not just a limited act of simple observation of the environment, it is a voluntary act. It works continuously and seeks to anticipate in a timely fashion. There are several definitions to Strategic vigilance, Strategic vigilance is an ongoing process of information management and decision support for the development and sustainability of the institution (Gujil, 2012: 56). Strategic vigilance has been defined as an organized process that includes research, analysis and the appropriate selection of information that enables the organization to obtain competitive advantages (Faltah, 2013: 32). The vigilance of the institution towards its changing world may also be a reflection of the institution's sensitivity to it (Boubaya, 2014: 78).

#### Characteristics of Strategic Vigilance

Strategic vigilance has a set of characteristics associated with its concept to pursue a set of objectives.

A. Strategy: It helps decision-making and is therefore related to non-recurring decisions, which are not any model models examined and value by experience, and reflect decisions that have a very significant impact on the competitiveness and survival of the Organization.

B. Volunteerism: Strategic vigilance cannot be passive and limited to simple follow-up and observation of the environment, because it is an innovative goal. On the contrary, it is voluntary, by requiring to go to the expected information interface with keen attention, stimulating all senses and exploring information.

C. Collective Intelligence: Collective intelligence means the presence of a group of individuals that observe signs in the environment to compare them to a particular meaning, which represents the goal of collective action whereby the members of the group or team are in contact and interaction in all appropriate forms.

D. Surrounding: The organization's surrounding is not just a concept, it is composed of several influential factors, and thus its impact will be practical, especially when targeting strategic vigilance.

E. Innovation: Strategic vigilance - taking into account the characteristics of the information concerned and the inquiry - early warning signals that relate to the element of Innovation. The relevant information in question does not describe any of the events or actions performed previously but allows the formulation of hypotheses and the creation of a voluntary vision.

F. Expectation: It is the information that has its own predictive properties, which must provide the explanations as a light for the future and it is not important to reflect the past or the present. Strategic vigilance is the process that binds the organization to its environment in such a way that it is ready to face sudden changes and pre-emptively, which requires strategic decision makers to understand the nature, interactions, relationship, and influence of each other. (Aziz, 2015: 94).

#### Types of Strategic Vigilance

The types of strategic alertness are divided according to the field of target activity. There are four major types of strategic vigilance: business vigilance, competitive vigilance, technological vigilance and surrounding vigilance, which are discussed below (M. AKLI, 2008: 47).

A. Business Vigilance: The research, processing and dissemination of information related to the market of the institution, and allows to identify new markets, and propose new products to the customers of the institution.

B. Competitive Vigilance: research, processing, and dissemination of information related to the organization's competitor, It is the control of direct and indirect competitors, both current and potential. This vigilance enables the determination of the business direction and the identification of the most threatening competitors.

C. Technological Vigilance: Technological vigilance is a systematic and disciplined effort by the organization to monitor, receive, analyze, publish and retrieve accurate information about certain events in economic, technological, social or commercial information that is important because it can involve an opportunity or threat (Dan SĂVESCU, 2014: 58).

D. Surrounding Vigilance: Is the process of research, collection and processing of information related to the macro environment of the institution, law, political, and social developments.

#### Stages of Strategic Vigilance

Strategic Vigilance is an information context that begins with the search for information and ends with the translation and use of information. In order for the organization to look at its environment and determine its location in it.

Phase I / Information Search: The stage of searching for information is the first and fundamental step in the process of strategic vigilance and therefore it needs to recruit all energies and possibilities. The most important steps of this stage are to target, track and collect vigilant information.

Phase II / Analysis and Information Processing: The process of strategic vigilance does not make sense if the information that has been monitored is not analyzed and interpreted. This information is often weak, fragmented or vague. It needs to be addressed to be useful and usable by decision makers. The process included Information selection, analysis, interpretation and storage of information.

Phase III / Use of Information: The previous stages are the core of strategic vigilance. It is often thought that strategic vigilance is limited to previous processes, but strategic vigilance cannot make sense without it being applied to the necessary decisions or it will only be excessive costs and loss for the institution. (Hassan Popaat, 2014: 46).

#### Strategic Intelligence (SI)

Strategic intelligence is an operational tool used by organizations that seek to maximize performance by adapting to the nature of changes in the business environment (Keikha and Hadadi, 2016: 264). Strategic intelligence is a revolution in modern thinking and adaptation practiced by organizations to ensure the acquisition of opportunities to survive, grow and develop performance. It is the most important concept that deals with the expected strategic change and response to any position that can sustain the organization (ALzuabi, 2016: 233). Strategic intelligence is a means of creating capacities for the systematic collection of information, which in turn will support strategic decision-making within organizations (Maccoby, 2014: 4).

#### Role of Strategic Intelligence and Its Impact on Management

Strategic intelligence focuses on the long-term goals of organizations, groups and governments. It deals with the most important goals and comprehensive plans, while it may be required or not required in the jobs performed in the operating environment and close to the work performed daily, depending on the level of assigned strategic tasks. Strategic intelligence can provide senior management and executive managers a deep insight and understanding of a range of areas including: (Mcdowel, 2009: 24).

- A. Modern and current trends.
- B. Changes in the troubled environment.
- C. Threats affecting the safety of others.
- D. Opportunities for task control and program development to address threats.
- E. Potential changes in the political, strategic, programs and laws.

#### Dimensions of Strategic Intelligence

Strategic intelligence includes three dimensions (business intelligence, competitive intelligence, and knowledge management) that will be addressed below (Keikha and Hadadi, 2016: 264-265).

A. Business Intelligence: Business Intelligence refers to business management used to describe applications and technology in the collection, presentation, access and

analysis of data in order to assist the organization and to achieve the best decisions in the field of business.

B. **Competitive Intelligence:** Competitive intelligence is a systematic process to gain information from competitors and to explore that information in order to facilitate organizational learning, development and monopolization in achieving industry, market and customer goals.

C. **Knowledge Management:** Knowledge management is an integrative and systematic approach aimed at identifying, managing and sharing all intellectual assets, including databases, documents, policies and experiences in the minds of individuals.

### III. METHODOLOGY

#### Research Problem

Strategic vigilance and strategic intelligence is an important issue for any modern organization that seeks to keep up with the dynamic developments in the business sector in which it operates. The research problem is reflected in the following questions:

A. Does the company investigated realize the importance of strategic vigilance in achieving strategic intelligence under the rapidly changing environmental conditions that sustainably adapt to ?

B. Is there an impact of strategic vigilance on the organization's strategic intelligence?

C. Does the company take into consideration the suggestions and ideas of its staff when conducting the strategic vigilance process?

#### Research Importance

In light of the economic, technological, political and other developments that have imposed on the organizations in general the method of change to adapt to the environment in which they operate in order to serve their objectives, we find that this change is an integrated system of internal and external information management by the company. To raise the interest of the organizations in general and the Organization in particular, in addressing such issues, which contribute to increasing their capacity and strengthen their competitive position in the future.

#### Research objectives

In light of the problem of research and its importance, the main objective of this research is to identify the role played by strategic vigilance in achieving the strategic intelligence of the company investigated, and this is reflected through the following sub-goals: A. Building a cognitive framework for the subject of strategic vigilance and strategic intelligence of the company by following the theoretical paths of specialized literature, and then presenting a conceptual framework describing the explanatory directions of these subjects.

B. Determine the impact of strategic vigilance in achieving the strategic intelligence of the investigated company.

C. Attempting to design and construct a detailed research plan and test it in light of the indicators and dimensions of the main and sub-research variables.

D. Provide a range of suggestions and recommendations to the company investigated and similar organizations about the extent to which strategic vigilance contributes to the achievement of strategic intelligence.

### IV. RESEARCH MODEL

In the light of the research problem and the objectives of research within its theoretical and applied framework, a virtual scheme was designed to express the logical relations between the variables of search as in Fig. 1. This diagram refers to two types of variables.

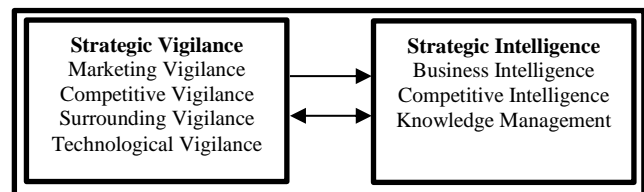


Fig. 1. Research Model

#### Research Hypothesis

In order to achieve the objectives of the research and test its hypothesis, the research was based on the main hypothesis (that there is a significant effect of strategic vigilance in the strategic intelligence of the company). The following sub-assumptions:

A. There is a significant significance of marketing vigilance (X1) in strategic intelligence (Y)

B. There is a significant effect of competitive vigilance (X2) on strategic intelligence (Y)

C. There is a significant effect of surrounding vigilance (X3) on strategic intelligence (Y)

D. There is a significant effect of technological vigilance (X4) on strategic intelligence (Y)

#### Research Boundaries

A. **Spatial boundaries:** The spatial boundaries of the research are a sample of employees, division managers and department managers working in the Al-Furat General Company for Chemical Industries

B. **Time boundaries:** The time boundaries of the research are extended for the period from 13/11/2017 to 17/12/2017

#### Research Tools

In order for this research to achieve its objectives, the data and information collection process was adopted on the questionnaire and included the following topics as shown in Table 1.

TABLE I. QUESTIONNAIRE TOPICS

First Section	Variables	Number Of Paragraphs
Information For Respondents	Gender, Age Group, Academic Qualification, Current Administrative Position, Years Of Service In The Organization, Number Of Training Courses Involved	6
Second Section	Variables	Number Of Paragraphs
First: Strategic Vigilance	Marketing Vigilance, Competitive Vigilance, Surrounding Vigilance , Technological Vigilance	10
Second: Strategic Intelligence	Business Intelligence, Competitive Intelligence , Knowledge Management	6
Total Sum		22

Distribution the questionnaire

Table 2 indicates the number of questionnaires distributed and retrieved by the respondents, as the number of questionnaires (200), and the number of questionnaires (200), which means that the response rate was (100 %).

TABLE II. NUMBER OF DISTRIBUTED AND RETRIEVED QUESTIONNAIRES

Research Community	Sample Size	Questionnaire		
		Distributed	Retrieved	Response Rate
Health Directorate Of The Holy Province Of Kerbala	300	200	200	100%
Sum		200	200	100%

Methods of statistical analysis used in research

In order to test the hypothesis of the research, a number of statistical methods were used, as well as the use of the social statistical program called SPSS 20. The statistical methods used are frequencies and percentages, simple regression coefficient simple regression coefficient,  $R^2$  identification factor, F test.

V. RESULTS

This section deals with the measurement of the effect of strategic vigilance in strategic intelligence, which is included in the main hypothesis that (there is significant relationship of strategic vigilance in strategic intelligence). The correlation matrix between the two research variables showed preliminary evidence of the existence of significant correlation between the search variables. However, the correlation needs to measure the level of effect between the two variables and between the independent and the dependent variables. In order to test the main hypothesis and its subsidiary assumptions, multiple regression analysis will be used. Table 3 shows the regression analysis of the research hypothesis. This table includes non-standard beta values ( $\beta$ ), the value of the interpretation factor ( $R^2$ ), and the calculated (F) value.

Test the main hypothesis

The second major hypothesis of the effect indicates that there is a significant effect between strategic vigilance and strategic intelligence. The Table 3 shows the results of the simple linear regression. For the purpose of assessing the impact of strategic vigilance in strategic intelligence.

TABLE III. VALUES THE REGRESSION MODEL FOR THE VARIABLES OF THE STUDY (N=200)

Strategic Vigilance / Strategic Intelligence	F Calculated	R <sup>2</sup>	$\beta$
	510.503	0.71	0.85

As shown in Table 3:

First: The value of (F) calculated for the estimated model reached (510.503) at the level of significance (0.01). Therefore, the hypothesis is accepted. This means that there is a statistically significant effect of strategic vigilance in strategic intelligence and with confidence (99 %).

Second: It is clear from the value of the coefficient of determination ( $R^2$ ) of (0.71) it is clear that strategic vigilance is able to explain (71 %) of changes in the strategic intelligence in the sample research. The percentage (29 %) is due to the contribution of other variables not included in the research model.

Third: It is evident through the value of the marginal slope coefficient ( $\beta$ ) of (0.85). Increasing the levels of strategic vigilance by one unit of standard deviations will increase the strategic intelligence by 85 % of one standard deviation unit if the hypothesis is accepted.

It is the main hypothesis and the following sub-assumptions emerge:

The first sub-hypothesis: There is a relationship of significant impact to the dimension of marketing vigilance in strategic intelligence. The results of the Table 4 show the following:

TABLE IV. REGRESSION MODEL FOR THE MARKETING VIGILANCE AND STRATEGIC INTELLIGENCE (N=200)

Marketing Vigilance / Strategic Intelligence	R <sup>2</sup>	B
	0.72	0.84

First: It is clear through the value of the coefficient of determination ( $R^2$ ) of (0.72) that the marketing vigilance is able to explain the percentage (72%) of changes in the variable strategic intelligence. The percentage (28%) is due to the contribution of other variables not included in the research model.

Second: It is determined by the value of the marginal slope coefficient ( $\beta$ ) of 0.84. An increase in marketing vigilance by one unit of standard deviations will increase the strategic intelligence by (84%) of one standard deviation unit and since the relation of the effect was significant, accept the hypothesis.

The second sub-hypothesis: There is a significant Impact between the competitive vigilance and the strategic intelligence in the research sample. The results of the Table 5 show the following:

TABLE V. REGRESSION MODEL FOR THE COMPETITIVE VIGILANCE AND STRATEGIC INTELLIGENCE (N=200)

Competitive Vigilance / Strategic Intelligence	R <sup>2</sup>	$\beta$
	0.32	0.57

First: The value of the coefficient of determination ( $R^2$ ) reached (0.32). That the competitive vigilance, 32 % of the changes in strategic intelligence are explained. And the rest

(68 %) contributed by other variables not included in the model.

Second: The value of the marginal inclination coefficient ( $\beta$ ) (0.57). In other words, the increase in the degree of competitive vigilance by one unit of standard deviation will increase the strategic intelligence by (57 %) from one standard deviation unit and since the relation of the effect is significant, accept the hypothesis.

The third sub-hypothesis: There is a relationship of significant Impact for the dimension of surrounding vigilance and strategic intelligence in the research sample, the results of Table (6) show the following:

TABLE VI. REGRESSION MODEL FOR THE SURROUNDING VIGILANCE AND STRATEGIC INTELLIGENCE (N=200)

Surrounding Vigilance	R <sup>2</sup>	B
Strategic Intelligence	0.10	0.32

First: The value of the coefficient of determination ( $R^2$ ) reached (0.10), that is, the surrounding vigilance explains the percentage (10%) of the changes in strategic intelligence. The remainder is contributed by other variables not included in the model.

Second: The value of the marginal slope coefficient ( $\beta$ ) (0.32). In other words, the increase in surrounding vigilance by one unit of standard deviations will increase the strategic intelligence by (32 %) of the unit of standard deviation and since the relation of the effect was significant, it accepts the hypothesis.

The fourth sub-hypothesis: There is a significant effect relationship of technological vigilance and strategic intelligence in the research sample. The results of Table 7 show the following:

TABLE VII. REGRESSION MODEL FOR THE TECHNOLOGICAL VIGILANCE AND STRATEGIC INTELLIGENCE (N=200)

Technological Vigilance	R <sup>2</sup>	B
Strategic Intelligence	0.53	0.73

First: The value of the coefficient of determination ( $R^2$ ) (0.53) means that technological vigilance explains positively (53 %) of changes in strategic intelligence. This indicates a positive reflection of the technological vigilance processes carried out by the management of the company in its work environment from the point of view of the surveyed sample.

Secondly, the value of the beta coefficient (0.73) indicates that any change of one unit technological vigilance leads to a change of (73 %) in strategic intelligence. This is a positive indicator. The higher the level of technological vigilance in the company, increased level of strategic intelligence.

## VI. CONCLUSIONS AND RECOMMENDATIONS

In this section, the conclusions will be presented in the light of the theoretical and practical procedures conducted by the researchers in the previous research paragraphs, and then submit the proposed recommendations to the organization, especially the research sample and other similar organizations.

## VII. CONCLUSIONS

1. The statistical results to describe the responses of the sample of the study on strategic vigilance showed that the investigated organization gives great attention to this variable because it has a big role in influencing the effectiveness of strategic intelligence programs. This has been demonstrated by what follows:

- Clear interest in marketing vigilance through which the organization has the real opportunity to know about the market conditions and what are the customers' ideas about the nature of the products.
- The obvious interest in competitive vigilance because of its great role in making the necessary adjustments to the strategies of the organization to monitor the nature of the methods used by competitors and excel in the market.
- The clear interest in surrounding vigilance, which contributes to an increase in the research, collection and processing of information on the organization's surroundings, allowing it to continuously renew its activities.
- The great emphasis on technological vigilance because it has an effective role in providing complete and correct information on technological development in terms of its application, timing of introduction and level of objectives.

2. The statistical results to describe the responses of the study sample on strategic intelligence showed that the investigated organization pays great attention to this variable as an operational tool used to maximize performance. This has been demonstrated by what follows:

- Pay good attention to the Business Intelligence Index by describing applications and technology in the collection and presentation of information to improve the decision-making process.
- Pay attention to the competitive intelligence index by actively seeking to gain systematic information about competitors and explore them in order to facilitate the process of organizational learning.
- The organization's interest in the knowledge management index in the formulation of its overall strategy through the exploitation and participation of all intellectual assets, including documents and databases to support internal and external knowledge.

## VIII. RECOMMENDATIONS

In the light of the above conclusions, and the theoretical and practical frameworks of the subject of the effect of strategic vigilance in achieving the strategic intelligence of the organization, and the findings (theoretical and practical), we review the following set of recommendations to the organization sample study as well as some general recommendations:

1. The organization's attention is focused on the issue of strategic intelligence because of its great role in increasing its competitiveness through the introduction of standards,

new patterns of behavior and ways of working to cope with the rapid change in the external environment.

2. The need to pay attention to the development of plans and strategies that will deal with high professionalism with strategic vigilance and research in the mechanisms of information management in order to support organizational decisions.

3. The selection of employees to accomplish the tasks commensurate with their abilities and their scientific and practical qualifications with emphasis on providing material and moral support and participation in decision-making.

4. The best exploitation of all the ideas and skills provided by working individuals and their attempt to be reflected in the organization's culture, processes and activities in order to achieve the desired organizational change.

5. The organization adopts the strategy of outstanding performance and work on the discovery of opportunities and exploitation of professionalism with the commitment to provide high-quality outputs to a wide range of customers.

6. It is necessary to prepare educational and cultural programs that contribute to the dissemination of a culture of awareness to the employees about the importance of strategic intelligence in improving and increasing the effectiveness of the overall performance of the organization.

#### REFERENCES

- [1] Aaker, D.A., V. kumar, G.S. Day and R. Leone, (A decision making perspective on marketing intelligence, in marketing research), 10th Edn., United Kingdom: John Wiley & Sons Ltd, 2009.
- [2] Al-Zu'bi, Hasan Ali (Aspects of Strategic Intelligence and its Role in Achieving Organizational Agility: An Empirical Investigation), International Journal of Academic Research in Business and Social Sciences, Vol. 6, No. 4, 2016.
- [3] Aziz, the relationship between strategic alertness and marketing culture and its impact in the direction of marketing management, doctoral thesis, Iraq 2015.
- [4] Balwinder Singh, Addl.Secretary, CVC, Vigilance as a Management Function – Leveraging Technology.
- [5] Brizon & Wybo, Vigilance: a process contributing to the resilience of organizations, France.
- [6] Cavalcanti, E.P. (The relationship between business intelligence and business success), Journal of Competitive Intelligence and Management, 2005.
- [7] Coyne, J. and P. Bell., (The role of strategic intelligence in anticipating transnational organised crime), A literary review. Science Direct, 2011.
- [8] Dan SĂVESCU, Aspects Regarding the Competitive Strategy in Companies with Technologic Vigilance, 2014.
- [9] Esmacili Mahmoud Reza (A Study on the Effect of the Strategic Intelligence on Decision Making and Strategic Planning), International Journal of Asian Social Science, 2014.
- [10] Gabber, H. (Competitive intelligence topology analyze for improved plan operation), Industrial Management and Data Systems 2007.
- [11] Gujail Nour El Abidine, The role of strategic vigilance in rationalizing communication between the institution and its surroundings, Master Thesis, Algeria, 2012.
- [12] Hassan Boubaya, The Effectiveness of Strategic Information Systems in Decision Making and Competitiveness Support, PhD, Algeria, 2014.
- [13] keikha, Abbas and hadadi, Ebrahim (Investigating effects of Strategic Intelligence of Managers on the performance of employees (Case Study: Private Banks in city of Zahedan), International Journal of Scientific & Engineering Research, Volume 7, Issue 3, March-2016.
- [14] Kuosa, T. (Different approaches of pattern management and strategic intelligence). Science Direct, 2011.
- [15] M. AKLI, Veale strategies en enterprise: Communication, Gouvernances et intelligence economic, Collogue International UFC 2008.
- [16] Maccoby, M. PhD, Tim Scudder, CPA, March article in Performance Improvement called "Strategic Intelligence: A Conceptual System of Leadership for Change", 2014.
- [17] Maymand, Mohammad Mahmoudi, Shayan, Ali and Kashani, Masoumeh (The Effect of Strategic Intelligence on the Organizational Citizenship Behavior of the Employees of Office of Economic and Financial Affairs of Hormozgan Province), I J A B E R, Vol. 13, No.6 (2015).
- [18] MCDOWELL, DON (Strategic Intelligence), A Handbook for Practitioners, Managers, and Users, by Scarecrow Press, Inc. 2009.
- [19] Sabrina, B., B. Laid and L.M. Jean, (Towards an information intelligence and knowledge management process in the context of information technologies), Journal of Organizational Knowledge Management, 2012.
- [20] Sadok M, Autre, Apports différentiels de l'internet pour la veille anticipative : Application au cas de réponse aux atteintes à la sécurité des réseaux d'entreprise, Actes du colloque AIM 2003 POSTED BY SATYENDRA ON MAY 8, 2014 IN MANAGEMENT
- [21] Salih, Ahmad Ali and Abdulrahman, Manal (The Role of Strategic Intelligence in the Development of Managers Competencies Portfolio a Study with Reference Jordanian Commercial Banks), International Journal of Economics, Commerce and Management, United Kingdom Vol. III, Issue 12, 2015.
- [22] Valtah, vigilance and its importance in strategic decision-making, doctoral thesis, Algeria 2013.

# Кластеризация нестационарных временных рядов

М. А. Калмыков

Санкт-Петербургский государственный электротехнический университет  
«ЛЭТИ» им. В.И. Ульянова (Ленина)

mica\_2011@mail.ru

**Аннотация.** Данная статья посвящена исследованию методов кластеризации нестационарных временных рядов, содержащих данные о пульсе человека. В работе рассматриваются методы K-Means, DBSCAN, Agglomerative Clustering и Random Forest для кластеризации таких временных рядов. Этот тип данных характеризуется изменяющимися статистическими характеристиками во времени и имеет важное значение в медицинском анализе. Основное внимание уделяется исследованию эффективности указанных методов на практике и анализу их применимости в медицинской диагностике. Полученные результаты имеют важное значение для развития методов анализа медицинских данных и прогнозирования на основе данных о пульсе человека.

**Ключевые слова:** временной ряд, пульс, кластеризация

## I. ВВЕДЕНИЕ

Исследование кластеризации временных рядов на основе данных о пульсовой волне человека является ключевой задачей в анализе физиологических данных. Пульсовые данные содержат информацию о сердечной активности, артериальном давлении и других физиологических параметрах, необходимую для медицинских исследований и диагностики. Рассматривается эффективность различных методов кластеризации, таких как K-Means, DBSCAN, Agglomerative Clustering и случайные леса, для анализа пульсовых данных. Цель исследования заключается в выборе наилучшего метода кластеризации временных рядов пульсовых данных с учетом их особенностей и применений в медицинских исследованиях и практике. Использованный набор данных о пульсовой волне содержит информацию о пульсовой частоте, важной для оценки физиологического состояния человека.

Каждый временной ряд в наборе данных представляет собой последовательность значений, отражающих изменения пульсовой частоты в определенные моменты времени. Набор данных содержит такие поля, как timestamp: временная метка, представляющая собой момент времени, когда были собраны данные; data: массив из 73 значений, представляющих пульсовую волну человека в соответствующий момент времени.

---

Работа выполнена при поддержке Минобрнауки России в рамках соглашения № 075-15-2021-291 от 15.04.2022г о предоставлении гранта в форме субсидий из федерального бюджета на осуществление государственной поддержки создания и развития научного центра мирового уровня «Павловский центр «Интегративная физиология – медицине, высокотехнологичному здравоохранению и технологиям стрессоустойчивости»

## II. МЕТОДЫ КЛАСТЕРИЗАЦИИ

Для анализа временных рядов пульсовых данных был применен ряд методов кластеризации. Каждый из этих методов обладает своими уникальными особенностями, которые позволяют эффективно выявлять структуры в данных и группировать их в соответствии с их сходством. Целью применения этих методов было сравнение результатов их работы и выбор оптимального по скорости работы и качеству метода для дальнейших исследований пульсовой волны.

### A. K-Means

K-Means разбивает данные на заранее заданное количество кластеров, оптимизируя сумму квадратов расстояний между каждой точкой данных и центром своего кластера. Этот метод прост в реализации и хорошо работает на больших объемах данных, но он требует заранее заданного числа кластеров и чувствителен к начальным условиям [1].

### B. DBSCAN

DBSCAN основан на плотности данных и способен обнаруживать кластеры произвольной формы. Он автоматически определяет число кластеров и обнаруживает выбросы. Этот метод хорошо работает с данными, содержащими шум и выбросы, но он может быть менее эффективен на данных с разной плотностью [2].

### C. Agglomerative Clustering

Agglomerative Clustering иерархически объединяет ближайшие кластеры на основе определенной меры расстояния. Этот метод создает дерево кластеризации, которое может быть визуализировано для анализа и интерпретации результатов. Он хорошо работает на данных с разной плотностью, но требует больше вычислительных ресурсов [3].

### D. Random Forest

Случайные леса – это ансамбль методов машинного обучения, основанный на деревьях решений. В контексте кластеризации временных рядов каждое дерево в случайном лесу может быть использовано для определения кластера для каждого временного ряда на основе его признаков. Этот метод хорошо работает с данными высокой размерности и способен обрабатывать большие объемы данных [4].

Каждый из этих методов имеет свои преимущества и недостатки, и выбор конкретного метода зависит от характеристик данных и целей исследования.

### III. РЕЗУЛЬТАТЫ

В этом разделе представлены результаты работы каждого из использованных методов кластеризации, а также показаны примеры кластеров, выделенных каждым методом.

В результате анализа временных рядов пульсовых данных каждым из 4 методов сформировано 10 кластеров. В целом, результаты работы алгоритмов схожи, однако имеют различия.

Пример кластера, выделенного методом K-Means, представлен на рис. 1. На графике изображены временные ряды пульсовых волн из данного кластера, выделенные цветом для наглядности. Этот пример демонстрирует результаты работы K-Means и структуру кластера, выявленную этим методом.

На рис. 1 видно, что метод выделил в один класс фрагменты волн, схожие по структуре. Большая часть фрагментов демонстрируют высокое сходство по форме, амплитуде и частотным характеристикам, но также, в класс попали фрагменты, достаточно сильно отличающиеся амплитудой.

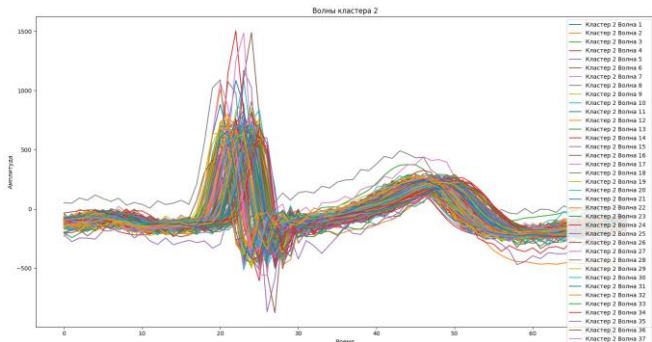


Рис. 1. Пример кластера, выделенного методом K-Means

Результаты работы метода DBSCAN представлены на рис. 2. Здесь показан пример кластера, выделенного этим методом, с учетом плотности данных.

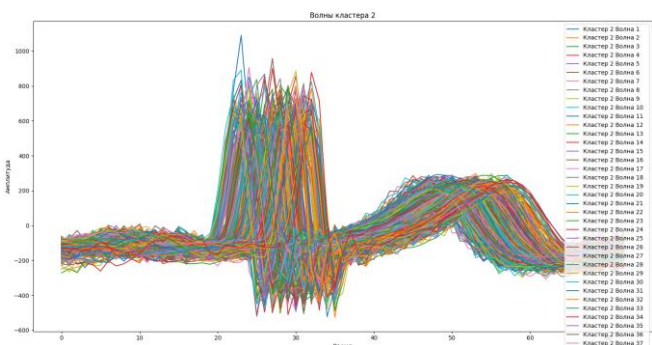


Рис. 2. Пример кластера, выделенного методом DBSCAN

На рис. 2 видно, что метод DBSCAN очень хорошо выделил схожие по структуре фрагменты. Хорошо заметны отличия результата от предыдущего метода.

На рис. 3 представлен пример кластера, выделенного методом аггломеративной кластеризации. Этот график демонстрирует иерархическую структуру кластеров, полученную этим методом, и визуализирует процесс объединения кластеров. Из рис. 3 видно, что данный

алгоритм хорошо справился с задачей, успешно выделив фрагменты со схожей структурой, несмотря на некоторое различие с результатами предыдущих методов.

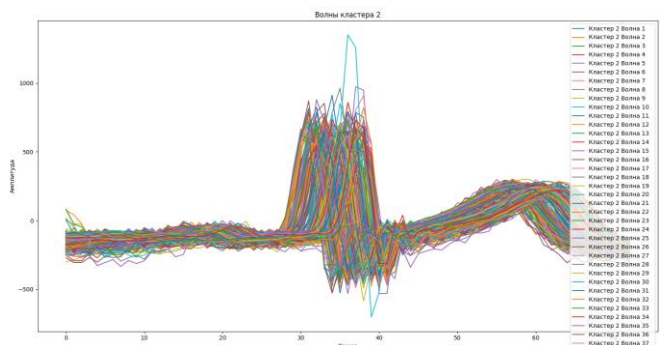


Рис. 3. Пример кластера, выделенного методом Agglomerative Clustering

Наконец, на рис. 4 показан пример кластера, выделенного методом случайных лесов. График отображает характеристики временных рядов пульсовых волн, соответствующих данному кластеру, и их распределение в пространстве признаков.

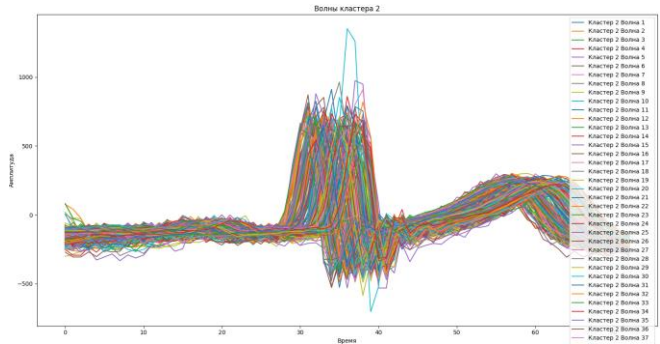


Рис. 4. Пример кластера, выделенного методом Random Forest

Результат, показанный на рис. 4, практически полностью совпадает с результатом метода аггломеративной кластеризации, что говорит о схожей эффективности работы этих методов.

Результаты сравнения методов показывают, что метод DBSCAN показал наилучшие результаты по сравнению с другими методами кластеризации. DBSCAN смог эффективно обнаружить кластеры с различными формами и плотностями, что подтверждает его эффективность в анализе временных рядов пульсовых данных.

Кроме того, было обнаружено, что кластеры, выделенные методом DBSCAN, имеют высокую степень схожести внутри кластера и различаются значительно по характеристикам пульсовых волн между кластерами. Это говорит о наличии различных физиологических состояний, которые могут быть выявлены с использованием анализа временных рядов пульсовых данных.

Стоит отметить, что наихудший результат показал метод K-Means, так как в кластеры, образованные этим методом, довольно часто попадали волны, совершенно не подходящие по форме и частоте.



Надо отметить, что использование ансамблевого метода кластеризации на основе случайных лесов и метода аггломеративной кластеризации также дало хорошие результаты. Результаты их работы, как видно из рис. 3 и рис. 4, сильно схожи, некоторые различия наблюдаются лишь в 2 кластерах.

Дальнейшие исследования в данной области могут привести к развитию новых методов анализа данных и повышению понимания физиологических механизмов организма.

#### IV. ОБСУЖДЕНИЕ

Результаты проведенного исследования демонстрируют эффективность применения методов кластеризации для анализа временных рядов пульсовых данных. Каждый из рассмотренных алгоритмов – K-Means, DBSCAN, Agglomerative Clustering и Random Forest – показал свои сильные стороны в выявлении структур в данных.

ТАБЛИЦА I. СРАВНЕНИЕ РЕЗУЛЬТАТОВ

Метод	Кластеры	Время	Точность	Масштабируемость
K-Means	10	8 сек.	0.79	Высокая
DBSCAN	10	5 сек.	0.92	Средняя
Agg. Clustering	10	16 сек.	0.81	Средняя
Random Forest	10	14 сек.	0.81	Высокая

Согласно результатам, приведенным в табл. 1, метод DBSCAN показал наилучшие результаты по сравнению с другими методами кластеризации временных рядов пульсовых данных. Несмотря на то, что методы K-Means и Random Forest продемонстрировали высокую масштабируемость, позволяющую эффективно применять их к большим объемам данных, DBSCAN смог наиболее эффективно обнаруживать кластеры с различными формами и плотностями.

DBSCAN смог выявить высокую степень схожести внутри кластеров и значительные различия между ними по характеристикам пульсовых волн. Это говорит о том, что данный метод способен эффективно обнаруживать различные физиологические состояния, что имеет важное значение для медицинской диагностики.

В то время как K-Means показал наихудший результат, часто включая в кластеры волны, не соответствующие по форме и частоте, DBSCAN продемонстрировал высокую точность и полноту кластеризации. Методы Random Forest и Agglomerative Clustering также показали хорошие результаты, но уступили DBSCAN по эффективности.

Таким образом, несмотря на высокую масштабируемость K-Means и Random Forest, метод DBSCAN является наиболее предпочтительным для кластеризации временных рядов пульсовых данных, учитывая его способность эффективно обнаруживать кластеры с различными характеристиками.

#### V. ПРАКТИЧЕСКОЕ ПРИМЕНЕНИЕ

Результаты данного исследования имеют важное практическое значение для медицинской диагностики и мониторинга состояния пациентов. Применение методов кластеризации к данным пульсовых волн может помочь в:

- выявлении патологических состояний на ранних стадиях по отклонениям пульсовых характеристик от нормы;
- оценке эффективности лечебных и реабилитационных мероприятий по динамике изменения пульсовых кластеров;
- прогнозировании развития заболеваний на основе анализа тенденций в переходах между кластерами пульсовых рядов;
- персонализации медицинских рекомендаций с учетом индивидуальных особенностей пульсовых характеристик пациента.

Для практического применения необходимо дальнейшее развитие методов анализа пульсовых данных, в том числе:

- создание эталонных баз данных пульсовых характеристик для различных возрастных групп и состояний здоровья;
- разработка алгоритмов автоматического определения патологических отклонений в пульсовых кластерах;
- интеграция методов кластеризации пульсовых данных в системы мониторинга состояния пациентов.

#### VI. ДАЛЬНЕЙШИЕ НАПРАВЛЕНИЯ ИССЛЕДОВАНИЯ

Результаты анализа временных рядов пульсовых данных с применением методов кластеризации представляют значимый вклад в область медицинского исследования. Одним из ключевых аспектов, требующих дальнейшего исследования, является автоматический подбор параметров модели DBSCAN. Этот метод, несмотря на свою эффективность, требует тщательной настройки параметров, что может быть улучшено с применением методов оптимизации, таких как сетки поиска или байесовская оптимизация [5].

Другим важным направлением исследования является работа с данными различной размерности. В реальных условиях длина пульсовых волн может варьироваться, и адаптация методов кластеризации к работе с данными произвольной размерности позволит расширить их применимость в медицинских системах.

Наконец, определение активностей человека на основе результатов кластеризации пульсовых данных представляет собой перспективное направление исследований. Это позволит не только выявлять различные физиологические состояния, но и использовать данные для распознавания активности человека. Путем сопоставления кластеров пульсовых рядов с активностями, зафиксированными в дневниках или с помощью акселерометров, можно разработать модель распознавания активности на основе пульсовых

данных, что будет полезно для мониторинга активности в повседневной жизни и реабилитационных программах. Эти направления исследования способствуют развитию методов кластеризации пульсовых данных и их широкому применению в медицинской практике, открывая новые возможности для диагностики и лечения.

## VII. ЗАКЛЮЧЕНИЕ

Проведенное исследование демонстрирует эффективность применения методов кластеризации для анализа временных рядов пульсовых данных. Особенно перспективным оказался алгоритм DBSCAN, который смог выявить кластеры с различными формами и плотностями, отражающие различные физиологические состояния организма.

Результаты работы имеют важное значение для развития методов анализа медицинских данных и прогнозирования на основе данных о пульсе человека. Практическое применение полученных результатов

может помочь в ранней диагностике заболеваний, оценке эффективности лечения и персонализации медицинских рекомендаций.

## СПИСОК ЛИТЕРАТУРЫ

- [1] Abiodun M. Ikotun, Absalom E. Ezugwu, Laith Abualigah, Belal Abuhajja, Jia Heming. K-means clustering algorithms: A comprehensive review, variants analysis, and advances in the era of big data //Information Sciences. 2023. Т. 622. С. 178-210.
- [2] Deng D. DBSCAN clustering algorithm based on density //2020 7th international forum on electrical engineering and automation (IFEEA). IEEE, 2020. С. 949-953.
- [3] Radovanović, Ana & Li, Junshi & Milanović, Jovica & Milosavljevic, Nina & Storchi, Riccardo. (2020). Application of Agglomerative Hierarchical Clustering for Clustering of Time Series Data. 640-644. 10.1109/ISGT-Europe47291.2020.9248759.
- [4] Liu, Zhenyu & Wen, Tao & Sun, Wei & Zhang, Qilong. (2020). Feature-Weighting and Clustering Random Forest. International Journal of Computational Intelligence Systems. 14. 10.2991/ijcis.d.201202.001.
- [5] Fan H., Mu J., Zhang W. Bayesian optimization with clustering and rollback for CNN auto pruning //European Conference on Computer Vision. Cham: Springer Nature Switzerland, 2022. С. 494-511.

# Организация процедуры обнаружения действий в видеоданных с использованием парадигмы MLOps

М. С. Куприянов<sup>1</sup>, Ю. А. Шичкина<sup>2</sup>, С. Е. Ильин<sup>3</sup>

Санкт-Петербургский государственный электротехнический университет  
«ЛЭТИ» им. В.И. Ульянова (Ленина)

<sup>1</sup> Mikhail.kupriyanov@gmail.com, <sup>2</sup> strange.y@mail.ru, <sup>3</sup> s.ilin2086@gmail.com

**Аннотация.** Обсуждаются особенности обращения к парадигме MLOps в контексте решения задачи по обнаружению действий в видеоданных. Анализируются результаты применения названной парадигмы на уровне планирования структуры проекта, при работе над его компонентами, а также над элементами в составе этих компонентов. Делается вывод о перспективности применения MLOps при организации процедуры обнаружения действий.

**Ключевые слова:** машинное обучение; парадигмы машинного обучения; Machine Learning Operations; MLOps; глубокое обучение; обнаружение действий

## I. ВВЕДЕНИЕ

Современный этап развития общества характеризуется активным обращением к технологиям генерации, сбора, анализа и редактирования видеоданных. Важное место в соответствующих процессах отводится интеллектуальным системам, призванным оптимизировать и автоматизировать работу с видеоданными. Исследовательское сообщество уделяет созданию и совершенствованию названных систем большое внимание. Растет количество вариантов их использования и круг решаемых с их помощью задач [1, с. 1–2; 8, с. 1–4; 15, с. 1].

К центральным задачам указанного спектра относится задача по автоматическому обнаружению действий в видеоданных (англ. action detection, action localization, action spotting) [9, с. 2]. Она допускает применение в ряде значимых областей общественной жизни (например, в работе с данными камер видеонаблюдения) и играет существенную роль в компьютерном зрении, автоматическом анализе видео и распознавании паттернов [16, с. 5]. Содержание задачи по обнаружению действий определяется по-разному в зависимости от количества агентов в кадре [10, с. 2]. В случае одного агента речь идет о выявлении в видеоданных моментов начала и завершения его действий с последующим их распознаванием (англ. temporal action detection). В случае, если агентов больше,

обнаружение действий представляет собой комбинацию задач по локализации этих агентов в кадре и распознаванию их действий. При этом распознавание действий в обоих случаях интерпретируется как отнесение действий агентов к одному из заранее определенных классов [10, с. 2].

По состоянию на сегодня наиболее эффективным путем решения указанной задачи признается обращение к машинному обучению (англ. Machine Learning, ML), и, конкретнее, к такому разделу машинного обучения, как глубокое обучение (англ. Deep Learning, DL) [16, с. 5]. При этом практическая значимость автоматического обнаружения действий делает перспективным направление соответствующей работы в русле парадигмы MLOps (англ. Machine Learning Operations), ориентированной на перенос в машинное обучение достижений из таких областей, как инженерия данных (англ. Data Engineering) и DevOps (англ. Development and Operations). Это открывает дополнительные возможности для повышения качества использования данных в контексте глубокого обучения, а также для применения практик непрерывной интеграции (англ. Continuous Integration, CI) и непрерывного развертывания программного обеспечения (англ. Continuous Delivery, Continuous Deployment, CD). В результате можно получить повышение качества работы моделей машинного обучения; рост контроля над использованием и распределением ресурсов, включая финансово затратные ресурсы; рост воспроизводимости экспериментов с моделями машинного обучения; упрощение и автоматизацию процессов дообучения и обслуживания моделей машинного обучения; и др. [11; 13; 14].

## II. ОБЗОР ЛИТЕРАТУРЫ

Исследования, связанные с применением парадигмы MLOps в контексте обнаружения действий, можно разделить на две основные группы. Первая группа объединяет работы, в которых обсуждается использование глубокого обучения при решении задачи обнаружения действий. В данном случае речь идет о применении таких парадигм машинного обучения, как обучение с учителем (англ. Supervised Learning), обучение без учителя (англ. Unsupervised Learning), слабо контролируемое обучение (англ. Weak Supervision) и др. [4, с. 53–66; 12, с. 2–3]. В зависимости от исследовательского подхода на передний план при

Работа выполнена при финансовой поддержке Министерства науки и высшего образования Российской Федерации, соглашение №075-15-2022-291 о предоставлении гранта в виде субсидий из федерального бюджета на осуществление государственной поддержки создания и развития Научного Центра Мирового Уровня Павловский центр "Интегративная физиология – медицине, высокотехнологичному здравоохранению и технологиям стрессоустойчивости"

решении задачи могут выдвигаться разные архитектуры искусственных нейронных сетей (сверточные сети, мультимодальные сети, трансформеры и пр.), а также методы обнаружения действий (разбивка видео на равные по времени отрезки с последующим их анализом; обнаружение действий с помощью запросов на естественном языке; и т. д.) [4, с. 30–52; 12, с. 4–13; 16, с. 6–8]. Исследования проводятся с принятием во внимание специфики поведения агентов, действия которых анализируются (люди, животные и т. д.), и охватывают несколько уровней обобщения данных (от создания интеллектуальных инструментов общего назначения до решения прикладных задач) [3; 10; и др.].

Вторая группа исследований включает в свой состав публикации, посвященные парадигме MLOps, особенностям ее использования, а также характерным для нее методам. Соответствующие источники фокусируются на содержании жизненного цикла моделей машинного обучения, интеллектуальных вызовах и практиках, направляющих работу в рамках парадигмы MLOps, а также на инструментах, призванных обеспечить приобщение к этим практикам [5, с. 1–2; 14, с. 3–6]. Обозначенные инструменты сопоставляются друг с другом по ряду критериев, таких как спектр поддерживаемых возможностей CI/CD, наличие контроля версий для датасетов, моделей, и/или пайплайнов, возможности оптимизации гиперпараметров моделей и др. [14, с. 6–7]. Предлагаются критерии для оценки уровня автоматизации систем на базе MLOps [5, с. 2–3; 11, с. 20]. Особенности парадигмы анализируются на примерах сценариев ее использования при создании систем машинного обучения [5, с. 7–11; 6, с. 5–8].

При обращении к указанным группам исследований особого внимания заслуживают публикации, ориентированные на непосредственное применение парадигмы MLOps в контексте автоматического обнаружения действий. Насколько удастся понять, они немногочисленны и затрагивают, в основном, отдельные аспекты темы. Речь может идти о решении задач, которые, в свою очередь, допускают рассмотрение в составе процедуры автоматического обнаружения действий, включая обнаружение объектов на изображениях и автоматическое распознавание действий [2; 5, с. 7–11]. При этом в фокусе исследования может находиться весь жизненный цикл моделей глубокого обучения [5] или отдельные его этапы с принятием в расчет тех аспектов MLOps, которые выносятся за границы исследования [2, с. 15]. Ряд обращенных к теме публикаций обнаруживается за пределами научного дискурса и дает примеры решения задачи с точки зрения практики [7].

Встреченные публикации наводят на мысль о перспективности дальнейшего сближения практик автоматического обнаружения действий и парадигмы MLOps в русле научного дискурса. В качестве задачи для исходного этапа соответствующей работы выбрана реализация процедуры обнаружения действий в составе видеоданных, полученных в условиях научного эксперимента при помощи одной неподвижной камеры для случая одного агента в кадре.

### III. МЕТОДОЛОГИЯ И РЕЗУЛЬТАТЫ

Поставленная задача решалась с помощью методов глубокого обучения, трансферного обучения и парадигмы обучения с учителем. Работа велась путем выполнения последовательности шагов, объединенных в три смысловых блока. Во-первых, шаги, направленные на импорт исходных данных. Во-вторых, шаги, направленные на обучение нейронных сетей. Наконец, в-третьих, шаги, направленные на использование этих сетей в рамках анализа новых данных. Обнаружение действий выступало целью указанного анализа и предполагало поиск в видеоданных фактов осуществления агентом конкретного действия из его поведенческого репертуара с выявлением кадров, на которых это действие представлено. С учетом специфики определения задачи в основу решения был положен метод двухклассовой классификации кадров видео по критерию наличия или отсутствия в них искомого действия агента.

Обращение к парадигме MLOps нашло выражение, по меньшей мере, на трех уровнях реализации решения. Прежде всего, на уровне структуры проекта потребовалось принятие во внимание технических компонентов, на базе которых выстраиваются системы машинного обучения с использованием MLOps. Конкретнее, речь идет о таких компонентах, как:

- хранилище исходных данных,
- репозиторий с исходным кодом проекта,
- инфраструктура для осуществления машинного обучения, включая необходимые вычислительные ресурсы,
- хранилище метаданных и данных о моделях машинного обучения в работе,
- средства оркестрации рабочих процессов,
- CI/CD инструментарий,
- средства мониторинга,
- и др. [6, с. 3–4].

Как итог, в сравнении с машинным обучением вне обращения к MLOps список рассматриваемых компонентов оказалось перспективно расширить. Их существование было обеспечено средствами ZenML, который представляет собой MLOps-фреймворк с открытым исходным кодом [17]. Степень проработки каждого компонента определялась его значением для достижения цели исследования.

Далее, принципы MLOps были привлечены в контексте реализации структурных компонентов решения. В частности, речь может идти о характерном для MLOps подходе к организации рабочих процессов в форме пайплайнов. Указанный подход был положен в основу обучения и использования нейронных сетей в составе проекта. Соответствующие пайплайны были соотнесены со смысловыми блоками решения задачи и разбиты на шаги с принятием во внимание последовательности их выполнения. В качестве примера на рис. 1 представлен пайплайн, объединяющий в себе

шаги по импорту данных для обучения нейронных сетей (рис. 1).

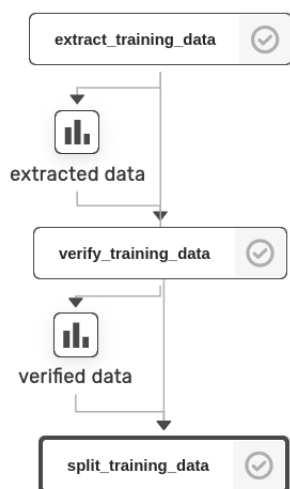


Рис. 1. Пайплайн на базе ZenML для импорта исходных данных в контексте решения задачи

Наконец, третий уровень применения парадигмы MLOps при решении задачи составило обращение к ней для проработки элементов в составе структурных компонентов проекта. Возможности использования контейнеризации позволили четче разграничивать и группировать эти элементы в соответствии с отличающимися их требованиями. Контроль версий для данных, пайплайнов и моделей нейронных сетей использовался для экономии времени при повторном прохождении ранее выполненных шагов. Информация о результатах запуска пайплайнов привлекалась с целью оптимизации решения. Например, на этапе импорта данных для обучения нейронных сетей проверялись альтернативные пути параллельной обработки видеофайлов при извлечении из них кадров. Тестирование проводилось локально на компьютере со следующими характеристиками: центральный процессор Intel Core i5-10300N 2.50GHz, RAM 8Gb, операционная система Ubuntu 22.04. Проверка сценариев с использованием многопоточности и многопроцессности продемонстрировала, что в актуальной реализации шага параллельная обработка видеофайлов приводит к сокращению временных затрат на извлечение данных в случае увеличения количества одновременно обрабатываемых файлов. При этом применение многопроцессности позволяет получить больший выигрыш по времени, чем применение многопоточности. Увеличение количества файлов в работе требует дополнительных ресурсов системы; один из способов сократить потребность в них представляет собой предварительное снижение разрешения видеофайлов. Примеры результатов, полученных в ходе обсуждаемой проверки, приведены на рис. 2.

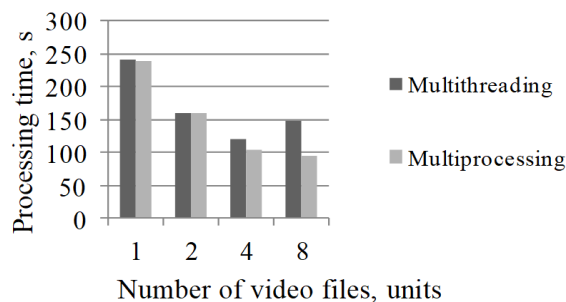


Рис. 2. Использование параллельной обработки видеофайлов на этапе импорта исходных данных

Подводя итог проделанной работе, можно констатировать, что привлечение парадигмы MLOps показало себя перспективным при организации процедуры обнаружения действий агента в видеоданных. Опора на обсуждаемую парадигму позволила принять во внимание особенности практического применения моделей машинного обучения в решении задачи, использовать актуальные подходы к выстраиванию структуры проекта и продуктивнее организовать работу над его компонентами.

#### IV. ЗАКЛЮЧЕНИЕ

Широкое внедрение машинного обучения в общественную жизнь создает новые перспективы и вызовы в русле применения соответствующих технологий. В числе активно обсуждаемых в настоящее время задач указанного спектра находится задача по обнаружению действий агентов в видеоданных. Она допускает решение в рамках различных парадигм машинного обучения, включая парадигму MLOps. При этом, обращение к последней обеспечивает доступ к дополнительным приемам и инструментам для решения задачи, а также органичную интеграцию решения в контекст практического применения машинного обучения. Работа в обозначенном ключе показывает, что при решении задачи автоматического обнаружения действий принципы MLOps могут использоваться на уровне общей структуры проекта, его компонентов, а также при работе над отдельными элементами в составе этих компонентов. С данной точки зрения особый интерес представляет оптимизация решений на основе MLOps, включая внутреннюю организацию шагов и пайплайнов в составе решений. Ввиду того, что парадигма MLOps является сравнительно недавним приобретением машинного обучения, указанная проблематика остается высокоактуальной и заслуживает специального рассмотрения.

#### СПИСОК ЛИТЕРАТУРЫ

- [1] Abdar M., Kollati M., Kuraparthi S., Pourpanah F., McDuff D., Ghavamzadeh M., Yan Sh., Mohamed A., Khosravi A., Cambria E., Porikli F. A Review of Deep Learning for Video Captioning [Electronic resource] // ArXiv. 2023. Pp. 1-42. URL: <https://arxiv.org/abs/2304.11431> (accessed: 11.03.2024).
- [2] Cob-Parro A. C., Losada-Gutiérrez C., Marrón-Romera M., Gardel-Vicente A., Bravo-Muñoz I. A New Framework for Deep Learning Video Based Human Action Recognition on the Edge // Expert Systems with Applications: An International Journal. 2024. Vol. 238. Iss. PE. Pp. 1-17.

- [3] Harris C., Finn K. R., Kieseler M.-L., Maechler M. R., Tse P. U. DeepAction: A MATLAB Toolbox for Automated Classification of Animal Behavior in Video [Electronic resource] // Scientific Reports. 2023. Vol. 13. Pp. 1-19. URL: <https://www.nature.com/articles/s41598-023-29574-0> (accessed: 02.03.2024).
- [4] Hu K., Shen Ch., Wang T., Xu K., Xia Q., Xia M., Cai Ch. Overview of Temporal Action Detection Based on Deep Learning [Electronic resource] // Artificial Intelligence Review. 2024. Vol. 57. Iss. 2. Pp. 1-77. URL: <https://link.springer.com/article/10.1007/s10462-023-10650-w> (accessed: 02.03.2024).
- [5] Ikram A., Tabassam U. MLOps: A Step Forward to Enterprise Machine Learning [Electronic resource] // ArXiv. 2023. Pp. 1-14. URL: <https://arxiv.org/abs/2305.19298> (accessed: 12.03.2024).
- [6] Kreuzberger D., Kühl N., Hirschl S. Machine Learning Operations (MLOps): Overview, Definition, and Architecture [Electronic resource] // ArXiv. 2022. Pp. 1-13. URL: <https://arxiv.org/abs/2205.02302> (accessed: 13.03.2024).
- [7] Mendenhall S. Building an Action Detection Scoring Pipeline for Digital Dailies [Electronic resource]. URL: <https://devblogs.microsoft.com/ise/building-an-action-detection-scoring-pipeline-for-digital-dailies/> (accessed: 15.03.2024).
- [8] Morshed M. G., Sultana T., Alam A., Lee Y.-K. Human Action Recognition: A Taxonomy-Based Survey, Updates, and Opportunities [Electronic resource] // Sensors. 2023. Vol. 23. No. 4:2182. Pp. 1-40. URL: <https://www.mdpi.com/1424-8220/23/4/2182> (accessed: 09.03.2024).
- [9] Pérez M., Toler-Franklin C. CNN-Based Action Recognition and Pose Estimation for Classifying Animal Behavior from Videos: A Survey [Electronic resource] // ArXiv. 2023. Pp. 1-29. URL: <https://arxiv.org/abs/2301.06187> (accessed: 11.03.2024).
- [10] Schindler F., Steinhage V., van Beeck Calkoen S. T. S., Heurich M. Action Detection for Wildlife Monitoring with Camera Traps Based on Segmentation with Filtering of Tracklets (SWIFT) and Mask-Guided Action Recognition (MAROON) [Electronic resource] // Applied Sciences. 2024. Vol. 54. Iss. 514. Pp. 1-17. URL: <https://www.mdpi.com/2076-3417/14/2/514> (accessed: 10.03.2024).
- [11] Steidl M., Felderer M., Ramler R. The Pipeline for the Continuous Development of Artificial Intelligence Models – Current State of Research and Practice [Electronic resource] // ArXiv. 2023. Pp. 1-85. URL: <https://arxiv.org/abs/2301.09001> (accessed: 07.03.2024).
- [12] Vahdani E., Tian Y. Deep Learning-Based Action Detection in Untrimmed Videos: A Survey [Electronic resource] // ArXiv. 2021. Pp. 1-24. URL: <https://arxiv.org/abs/2110.00111> (accessed: 12.03.2024).
- [13] Warnett S. J., Zdun U. On the Understandability of MLOps System Architectures [Electronic resource] // IEEE Transactions on Software Engineering. 2024. Pp. 1-25. URL: <https://ieeexplore.ieee.org/document/10440483> (accessed: 11.03.2024).
- [14] Wazir S., Kashyap G. S., Saxena P. MLOps: A Review [Electronic resource] // ArXiv. 2023. Pp. 1-9. URL: <https://arxiv.org/abs/2308.10908> (accessed: 07.03.2024).
- [15] Xing Zh., Feng Q., Chen H., Dai Q., Hu H., Xu H., Wu Z., Jiang Y.-G. A Survey on Video Diffusion Models [Electronic resource] // ArXiv. 2023. Pp. 1-21. URL: <https://arxiv.org/abs/2310.10647> (accessed: 10.03.2024).
- [16] Zeng Zh. A Review of Video Action Detection Based on Deep Learning [Electronic resource] // Frontiers in Computing and Intelligent Systems. 2023. Vol. 3. No. 2. Pp. 5-9. URL: <https://drpress.org/ojs/index.php/fcis/article/view/6908> (accessed: 11.03.2024).
- [17] ZenML Documentation. Introduction [Electronic resource]. URL: <https://docs.zenml.io/getting-started/introduction> (accessed: 15.03.2024).

# Personal Document Security Tool for Cloud Application

Bdoor Alaa Mahmood

College of Computer Science and Information Technology, University of Kerbala

Karbala 56001, Iraq

E-mail: bdoor.a@uokerbala.edu.iq

**Abstract**— The recent advancement in cloud storage has spurred the creation of more methods in developing the personal data security features in both applications and personal document storage. Personal document storage can be divided into different segments based on the types of document and the usage of the document in cloud storage. This research focuses on personal document storage used by individuals for their personal use only. The method involves developing a simple and easy-to-use security tool with secured features and efficient algorithm encryption and decryption with reliable speed for large files, such as Word files and pictures. In this research, more than one algorithm will be tested using graphical user interface, which has encryption and decryption through a private cloud computing server.

**Keywords**—component; formatting; style; styling; insert

## I. INTRODUCTION

Companies and users are both in a transition period as they are moving their operation to cloud system resources [1]. As more individuals migrate to the cloud usage, more personal data such as bank information, transaction reports, domain services, and even full imaged operating systems are increasingly targeted and vulnerable.

Cloud data are susceptible to attackers if continuous monitoring and maintenance not implemented. Attackers can maliciously access a cloud system, copy data, or even corrupt an entire storage system. To mitigate these attacks, researchers must explore and improve the cloud security model. A previous work explains the history of in-house developed tools and security teams working non-stop to scan their system and patch vulnerabilities[2]. Although several issues are addressed in this research, cloud computing has brought new challenges to technology users. One of the most significant issues is personal document security parked in the cloud. Many free cloud hosting facilities are provided by known and unknown sources. Most of these facilities do not provide robust security features to users in terms of protecting personal documents, such as identification cards, certificates, medical reports, and other significant documents.

However, in this study, we will build a modern security tool based on Elliptic Curve Cryptography (ECC) with Diffie Hellman (DF) to secure the channel between cloud clients and their server against disclosing any important data or changes in cloud storage structure [3].

## II. LITERATURE REVIEW

A common fallacy in the area of cloud security is the “fallacy of direct control.” According to this fallacy, an environment will be inherently more secure when the IT manager has direct control and physical contact with one of the major factors that traditionally hinders the security of cloud computing acceptance, including issues relating to availability, integrity, confidentiality, data access, data segregation, privacy, recovery, accountability, and multi-tenancy issues, although many of these issues may have more perception of vulnerability than reality. Two broad categories, namely, technological and economic, discuss why cloud is more secure than an on-premise environment [4]. A narrowly altered recreation of the concept of “network cloud” or “cloud” was presented in the 1990s to the networking industry [5]. The concept signified an abstraction layer originating from the delivery methods of data through heterogeneous public and semi-public networks that were principally packet switched. Cellular networks also sufficiently utilized the concept of “cloud” [6].

### A. Cloud computing era

Cloud computing facilities have grown in recent years because of the increase in the demand for cloud-based services [7], resulting in tremendous pressure on developers to provide sufficient hardware and software resources. Administrators must also ensure the efficient energy and hardware utilization of cloud computing facilities. Furthermore, administrators must address the security of cloud-based services, such as data and applications of on-demand services. Therefore, monitoring cloud computing resources and their applications is important. This study provides a thorough review of cloud computing tools that are used for monitoring performance of cloud infrastructure at consumer and provider ends. A server-monitoring tool can provide a visually appealing and intuitive interface to allow system administrators to view network data and results and identify problems or bottlenecks. Various features must be considered when selecting data visualization tools. Most tools can identify the particular host or node that is causing problems and include various graphics for evaluating customer metrics. Monitoring plays a crucial role in enhancing the cloud computing service quality. It can assist a company in adaptively scaling cloud computing resources, service developers in determining service defects, and end users in exploring usage patterns [8].

Nevertheless, these resources gather in several virtual machines, as shown in Fig. 1.

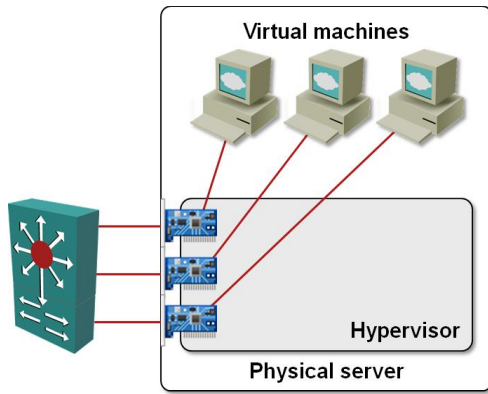


Fig. 1. Types of Vm servers studied by [9]

As a commodity mechanism, the virtual server represents the most foundational building block of cloud environments. Each virtual server can host numerous IT resources, cloud-based solutions, and various other cloud computing mechanisms. The instantiation of virtual servers from image files is a resource allocation process that can be completed rapidly and on demand [10].

Data storage security can be regarded as the most important concern and issue in cloud computing usage. A cloud environment allows users to remotely store their documents and files and relieve themselves from difficulties and problems of local storage and maintenance. Nevertheless, by using Cloud computing, users lose control over their documents. Present methods do not consider the dynamic nature of cloud, computation, and overhead communication [11].

### B. Cloud storage tools

The vast majority of the distance data storage nowadays totally depend on cloud virtualization services [12]. However, this phenomenon raises two key questions, particularly from the user perspective: What would happen if I were to place my data inside the cloud servers? Is this approach safe and secure? In 2014, the Apple iCloud service was attacked, and several passwords were revealed. This incident prompted infrastructure companies in the IT sector to review their cloud security tool. Table 1 illustrates the cloud security tools, and some of them are free of cost, whereas others are commercialized.

A short survey was conducted through a literature review on the encryption/decryption of information in cloud data security storage, as shown in Table 2.3, and a summary of the authentication approach discussed in this study is presented in Table 1.

TABLE I. RELATED LITERATURE REVIEW WORKS

AUTHOR(S)	WORK ANALYSIS
[13]	They address the challenging open problem using capability-based access control technique that ensures that only valid users will access the outsourced data. Their work also proposes a modified Diffie–Hellman key exchange protocol between cloud service provider and the user for secretly sharing a symmetric key for secure data access that alleviates the problem of key distribution and management at the cloud service provider.

AUTHOR(S)	WORK ANALYSIS
[14]	They support public adaptability without retrieving the data blocks themselves; they resort to a holomorphic authenticator technique using RSA Tag.
[15]	They specify elective delegation regarding the whole or partial message according to the reliability of the delegator toward a delegate using type-based re-encryption.
[16]	They present a secure outsourcing mechanism for training and evaluating large-scale logistic regression classifiers in the cloud. Their mechanism enables a customer to securely harness the cloud while maintaining the privacy of both the sensitive input and output of the computation. They also perform a thorough security analysis and conduct prototype experiments on Amazon EC2 to demonstrate the validity and practicality of their proposed design based on the Paillier cryptosystem.

Cloud computing has become a cutting-edge technology. However, recent server issues have shown the effect that this technology can have on clients and cloud servers. System administrators must be familiar with cloud server monitoring and network tools. Previous studies have examined the different components of cloud computing, such as its properties, technology, privacy, and security issues, but none of these studies have proposed suitable personal data tool for a cloud server. Therefore, studying security tool plays a significant role in enhancing the quality of security of personal data in cloud computing.

### III. METHODOLOGY

The proposed research methodology aims to achieve the goal of the study. The most popular method of experimentation in terms of the security tool for cloud computing is the simulation. One of the methodology approaches has inconclusive cloud security [17]; the authors merely used ECC to secure the sensitive data between the cloud server and its clients. In addition, the reasons for selecting MATLAB simulation for cloud security tool are explained. The implementation that is related to the simulation tool and the component of the performance metrics that is used in the analysis of the simulation results are discussed.

#### A. AElliptic curve Diffie–Hellman (ECDH)

The agreement on elliptic curve Diffie–Hellman (ECDH) key protocol can be between two users who agree on a common password. They can agree in the middle of a non-secure process, without having to share any private information in advance. This protocol is typically used in reaching an agreement on a shared secret password to derive encryption key materials. Similar standards such as advanced encryption standard (AES) use the encryption key to encrypt subsequent messages into the material [18–22]. Nonetheless, in Fig. 2 shown how the ECDH setting- which has adopted in this study as initial crypto process for the doc under graphic user interface (GUI).



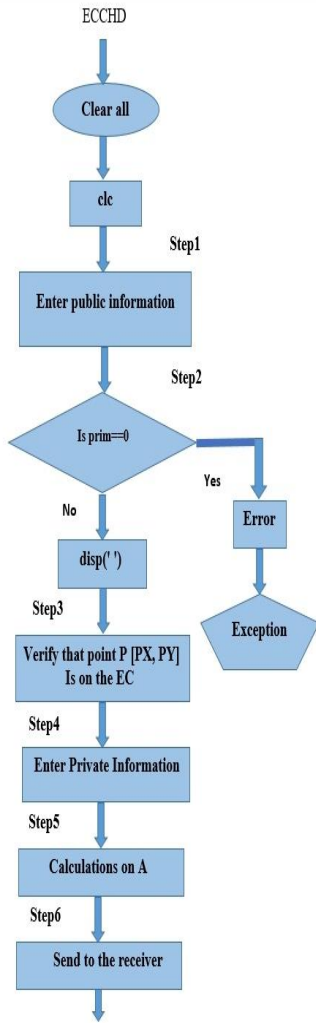


Fig. 2. ECDH setting

#### IV. SIMULATION ARCHITECTURE

To establish the platform, the following steps are required for designing a customized web server:

1. A computer with Windows 7 and MATLAB editor and compiler version 2015 or higher and
2. MATLAB Web service toolbox. Installing MATLAB compiler on the computer is preferred but is not compulsory. Multiple instances of web services can run simultaneously without considerable overhead.

The following items are required for operating a web service:

1. Apache Tomcat
2. A computer with Windows 7.

The first simulation features for MATLAB works start from Fig. 3, which shows the GUI process through MATLAB.

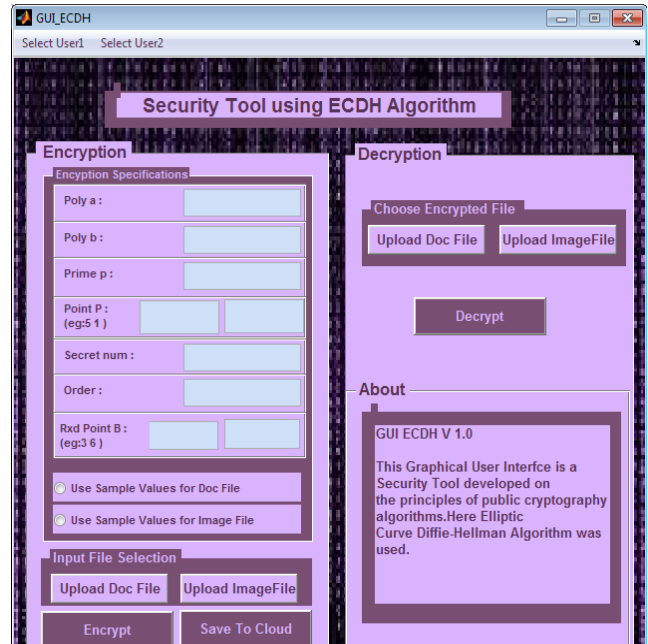


Fig. 3. GUI process through MATLAB

Step 3: First, we select one of the user(s) (user 1 or 2) to upload the text file or the image. Note that both users have the same steps to follow. Fig. 4, shows the selection user for the crypto GUI tool.

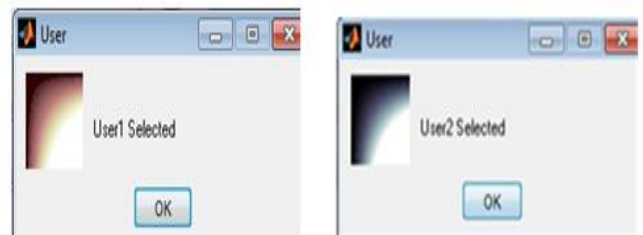


Fig. 4. User selection for the crypto GUI tool

Step 4. Selection ECDH the parameters that is, either doc. file sample values or image file simple values, depending on the file type. Doc. sample values will be selected for example.

Step 5: the select msgtosend.txt file to encrypt; then, press the encrypt button. After the encryption, the document text will be saved in tomcat directory. To decrypt the file, upload the msgtosave.txt file (encrypted file) and press the decrypt button to decrypt the message. Fig. 5 a,b,c illustrates the document encryption and decryption.



Fig. 5. The document encryption and decryption. a) Text before encryption; b) Text after encryption; c) Text after decryption

Step 6: To use the sample value for the image file, choose the uploaded image file; afterward, select the image and press the encrypt button to encrypt the image. The file will be saved also in tomcat directory. To decrypt the file, upload the encrypted image and press the decrypt button. Fig. 6 shows the original image (size 20 kb 256×256) as well as the encryption and decryption images.



Fig. 6. The encrypted and decrypted image file: a) Original image 20 kb 256x256; b) Encrypt image; c) Decrypt image

V. VERIFICATION & VALIDATION

Another important tool to evaluate the efficiency of algorithms is measuring the amount of time required to encrypt an image. In this investigation. In many real-time applications, it is important that the encryption and decryption algorithms are fast enough to meet the requirements. Designer should attempt to optimize a cryptosystem to make the execution time as lower as possible. We have calculated the execution time for text and image.

Fig. 7 show the result of five different size document files decryption and encryption time and it also explain the relationship between time and document file size.

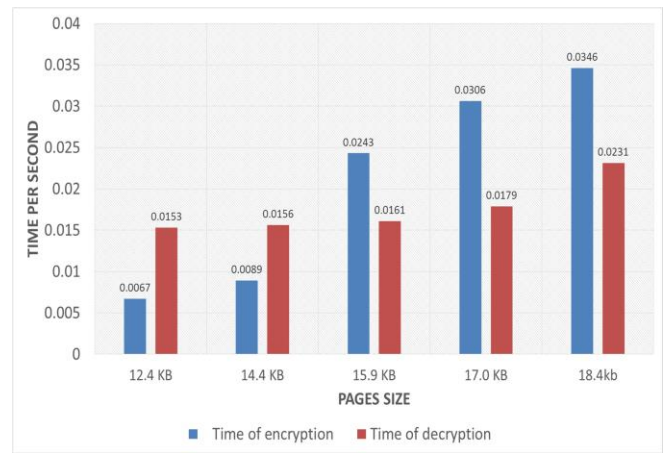


Fig. 7. Time-Page size relationship for the ECDH

Furthermore, Fig. 8 show the result of nine different size image files decryption and encryption time and it also explain the relationship between time and image file size.

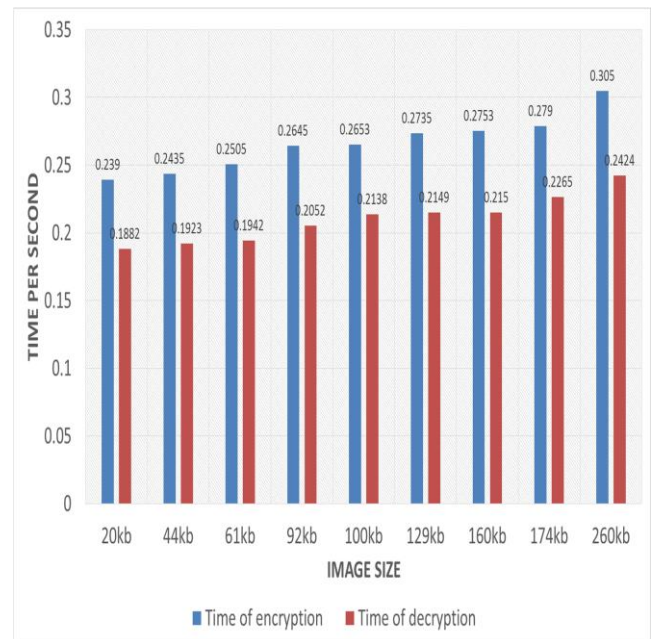


Fig. 8. Time-Image size relationship for the ECDH

Fig. 9 show the ECDH encryption time interval and decryption and it's explained clearly how dose effect the resolution on encryption and decryption time.

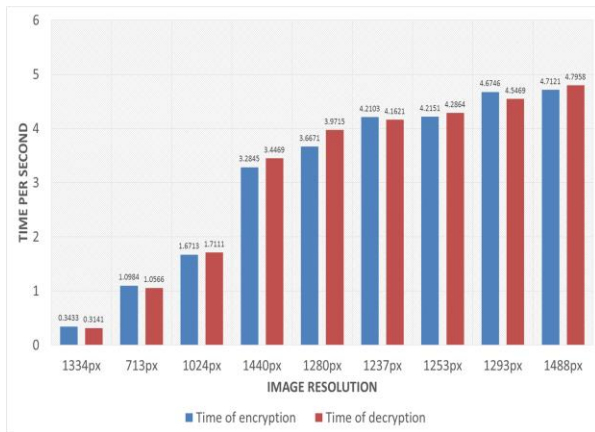


Fig. 9. Time-Image resolution relationship for the ECDH

Algorithm performance was very clear through result of text files that the time encryption and decryption had been increased logically with the text file size increment. It should be noted also that the text files have average encryption time of 0.0211 seconds, which is approximately 61 % of maximum encryption value per all measurements. Furthermore, two case selected for the image files, first one to find out the time-image size relationship. The result show that the time increased logically with the image size increment while the second case was to find out the time-image resolution relationship and the result show that the time had been increased as long as resolution rise up and it was logic result also. It should be noted that the image files have average encryption time of 0.2659 seconds which is approximately 87 % of maximum encryption value per all measurements. Apparently, images have better encryption efficiency in comparison to text documents. For example, the document format found to be 0.0243 seconds in encryption time and the optimum document size in 15.9 KB, while the best image format found to be 0.305 seconds in encryption time and the optimum image size in 260 KB. The developed personal data security tool has the basic needs for both text and image file encryption and decryption capacity.

## VI. CONCLUSION

Nowadays majority internet users moving towards cloud storage servers to have remote access to their documents but the safety of their personal documents was not given due attention in terms of safety. Most of the service providers both the open source and commercial are claiming for wide range of security availability in their services but they lack proof and investigation activities. Most of the time, the weakness will be identified when there are security breach incidences take place. Therefore, in this project, it has been shown how the strength of adopting the asymmetric key for personal document security and document privacy for cloud storage users. The best result has been achieved though the mechanism of the ECDH -which encrypt / decrypt text files, and image files for security purpose before storing and upon retrieval from the cloud storage. The good challenge here when we try to encrypt the image, we have faced the decrypt image quality and that has overcome by our specific code deployment in MATLAB. The conclusion is, selecting ECHD as the cloud storage security algorithm has provided an opportunity to design an efficient cloud storage personal security simulator and ensuring the quality of the document.

## REFERENCES

- [1] S. Marston, Z. Li, S. Bandyopadhyay, J. Zhang, and A. Ghalsasi, "Cloud computing—The business perspective," *Decision Support Systems*, vol. 51, no. 1, pp. 176-189, 2011.
- [2] J.-M. Chang, P.-C. Tsou, H.-C. Chao, and J.-L. Chen, "CBDS: a cooperative bait detection scheme to prevent malicious node for MANET based on hybrid defense architecture." pp. 1-5.
- [3] W. Jansen, and T. Grance, "Guidelines on security and privacy in public cloud computing," NIST special publication, vol. 800, pp. 144, 2011.
- [4] T. Karagiannis, "The New Face of the Internet," *Trust, Computing, and Society*, pp. 38, 2014.
- [5] P. Singh, V. P. Singh, and G. Pachauri, "Critical Analysis of Cloud Computing Using OpenStack," 2014.
- [6] A. Badanidiyuru, B. Mirzasoleiman, A. Karbasi, and A. Krause, "Streaming submodular maximization: Massive data summarization on the fly." pp. 671-680.
- [7] M. Armbrust, A. Fox, R. Griffith, A. D. Joseph, R. Katz, A. Konwinski, G. Lee, D. Patterson, A. Rabkin, and I. Stoica, "A view of cloud computing," *Communications of the ACM*, vol. 53, no. 4, pp. 50-58, 2010.
- [8] P. Mell, and T. Grance, "The NIST definition of cloud computing," 2011.
- [9] Q. Zhang, L. Cheng, and R. Boutaba, "Cloud computing: state-of-the-art and research challenges," *Journal of internet services and applications*, vol. 1, no. 1, pp. 7-18, 2010.
- [10] R. Buyya, C. S. Yeo, S. Venugopal, J. Broberg, and I. Brandic, "Cloud computing and emerging IT platforms: Vision, hype, and reality for delivering computing as the 5th utility," *Future Generation computer systems*, vol. 25, no. 6, pp. 599-616, 2009.
- [11] Q. Wang, C. Wang, J. Li, K. Ren, and W. Lou, "Enabling public verifiability and data dynamics for storage security in cloud computing," *Computer Security—ESORICS 2009*, pp. 355-370: Springer, 2009.
- [12] S. Kamara, and K. Lauter, "Cryptographic cloud storage," *Financial Cryptography and Data Security*, pp. 136-149: Springer, 2010.
- [13] S. Sanka, C. Hota, and M. Rajarajan, "Secure data access in cloud computing." pp. 1-6.
- [14] Q. Wang, C. Wang, K. Ren, W. Lou, and J. Li, "Enabling public auditability and data dynamics for storage security in cloud computing," *Parallel and Distributed Systems, IEEE Transactions on*, vol. 22, no. 5, pp. 847-859, 2011.
- [15] N. Park, "Secure Data Access Control Scheme Using Type-Based Re-encryption in Cloud Environment," *Semantic Methods for Knowledge Management and Communication, Studies in Computational Intelligence R. Katarzyniak, T.-F. Chiu, C.-F. Hong and N. Nguyen, eds.*, pp. 319-327: Springer Berlin Heidelberg, 2011.
- [16] X. D. Zhu, H. Li, and F. H. Li, "Privacy-preserving logistic regression outsourcing in cloud computing," *International Journal of Grid and Utility Computing*, vol. 4, no. 2, pp. 144-150, 2013.
- [17] S. Singh, and V. Kumar, "Secured user's authentication and private data storage- access scheme in cloud computing using Elliptic curve cryptography." pp. 791-795.
- [18] X. Yin, N. Thiranan, and H. Lee, "An Efficient and Secure Data Storage Scheme using ECC in Cloud Computing," *인터넷정보학회논문지*, vol. 15, no. 2, pp. 49-58, 2014.
- [19] Bader Alouffi , Muhammad Hasnain , Abdullah Alharbi3, Wael Alosaimi Hashem Alyami, And Muhammad Ayaz, "A Systematic Literature Review on Cloud Computing Security: Threats and Mitigation Strategies," *IEEE Access*, vol. 09, 2021.
- [20] D. Stalin David, Mamoona Anam, Chandrababha Kaliappan, S. Arun Mozhi Selvi, Dilip Kumar Sharma, Pankaj Dadheech and Sudhakar Sengan, "Cloud Security Service for Identifying Unauthorized User Behaviour," *Computers, Materials & Continua*, vol. 70, no. 02, pp. 2581-2600, 2022.
- [21] Feras M. Awaysheh, Mohammad N. Aladwan, Mamoun Alazab, Sadi Alawadi, José C. Cabaleiro, and Tomás F. Pena, "Security by Design for Big Data Frameworks Over Cloud Computing," *IEEE TRANSACTIONS ON ENGINEERING MANAGEMENT*, vol. 69, no. 06, pp. 3676-3693, 2022.
- [22] Arjun Reddy Kunduru, "Security Concerns and Solutions for Enterprise Cloud Computing Applications," *Asian Journal of Research in Computer Science*, vol. 15, no. 04, pp. 24-33, 2023.

# Bayesian Analysis of Zero-Inflated Count Data Using Beta-Lindley Distribution

Atheer Ismael Fadhel  
 University Mustansiriyah  
 Iraq - Baghdad  
 gqukcq1ns37r23@gmail.com

Hossein Jabbari Khamnei  
 University Tabriz  
 Iran - Tabriz  
 h\_jabbari@tabriz.ac.ir

**Abstract**— The process of analyzing data of large numbers assumes the presence of a unique and large set of challenges due to extra zeros and excessive dispersion processes. Therefore, data types that rely on the traditional counting principle attempt to capture the characteristics of this data appropriately and work on it bypassing these problems. Hence, the Bayesian principle was proposed, which seeks To model inflated zero count data based on the beta Lindley distribution, as this distribution is known for its flexibility and ability to deal with excess zeros and excess dispersion simultaneously, through the use of Markov chain Monte Carlo (MCMC) techniques for parameter estimation and model inference, as this paper seeks. The research aims to clarify the effectiveness of the proposed method by studying simulations and applying them to real-world data sets.

The results of the Beta-Lindley distribution provide a robust and flexible framework based on the principle of modeling zero-inflated count data. This approach will provide accurate parameter estimates and reliable model predictions and will therefore have multiple uses in a range of different activities, including epidemiology, some environmental studies, and even within the economic sciences, where data are common.

**Keywords**— *Beta lindley, zero inflated, Data, MCMC*

## I. INTRODUCTION

The situation in which the response variable is a non-negative integer and there is an ability to express it when we compare it to a census data set is often used frequently in data in the fields of environment, epidemiology, and economic sciences [1].

The census data has a relatively large value of zeros, higher than expected under standard census data models such as Poisson distributions or negative binomial distributions. Here we arrive at what is called the zero-inflation phenomenon, and its cause is due to heterogeneity or the presence of smaller subcategories susceptible to the event in question or to other factors. Others [1–2]

In order to arrive at a formula for dealing with zero-inflated count data, a wide range of researchers have worked with different modeling techniques, including zero-order Poisson (ZIP) and zero-order negative binomial (ZINB) models for this problem.

These models rely on several assumptions, and among these assumptions is that the observed counting data is a combination of a group of operations, and among these operations used is the binary operation that depends on generating additional zeros, and a counting operation that generates numbers other than zero [2].

In the midst of these ideas, the beta-Lindley distribution began to emerge as an alternative model to traditional count data models that solves this issue. The beta-Lindley distribution is known as a flexible probability distribution that can capture different shapes and patterns of dispersion, and this is what makes it suitable for modeling zero-sum data through the process incorporating the beta-Lindley distribution into the zero-sum distribution modeling framework and thus this flexibility allowed researchers to gain good ability in modeling processes [3].

The Bayesian model can provide an effective method for estimating the parameters of zero-count data models and thus the ability to provide a powerful analysis of different information through its integration with the beta-Lindley distribution. This is especially useful in cases of dealing with sparse or imbalanced data.

These modes of integrating the model with the distribution to arrive at analyzes provide Bayesian and Natural Method methodology for measuring uncertainty in parameter estimates and making probabilistic inferences about underlying processes. This approach, predictable under standard count data models, can therefore provide valuable insights and inform decision-making in a wide range of applications. Thus, it can be concluded that the goal of the research is to present an advanced Bayesian analytical approach to processing zero-inflated digital data through its integration with the beta-Lindley distribution to improve the process of modeling and statistical inference for ordinary and complex miscellaneous data that suffer from a significant increase in the number of zero values compared to traditional models of digital data. [4]

The research problem lies in the amplification of zero digital data and its increased dispersion and spread, which constitutes a major challenge for researchers in many fields, because it is not easy enough to deal with this phenomenon and creates hierarchical problems resulting from it in later stages, thus moving to inaccurate analyzes and this creates a definite need to reach a model. Composite hybrid by developing statistical models that are more flexible and able to understand the complex nature of these data[5].

Working on advanced statistical methodologies developed to analyze inflated numerical data is the goal of this thesis because it will give better opportunities for modeling and statistical inference and thus reach high accuracy with reliable conclusions and this will have a positive impact on decision-making processes [6]. Therefore, we need to discover how the analysis approach can be applied. Bayesian through beta Lindley distribution mechanisms within the zero-inflated digital data model more

effectively, and what are the advantages resulting from this method and the advantages that this approach gives compared to traditional zero-inflated digital data models and how to invest these results derived from it and the processes of interpreting and evaluating them through the beta Lindley distribution in applied contexts different. [7]

## II. LITERARY AND REFERENCE STUDIES

According to reference studies, it has been shown that data containing extra numbers followed by a number of zeros, known as zero-inflated data, is spread over a large set of data within several fields, and that the dominant feature of statistical models is that traditional statistical models often fail to adequately capture the basic structure of this data, and this It will give a set of biased estimates and imprecise inferences as the empirical Lindley distribution has emerged as a promising alternative for modeling null-inflated count data within a Bayesian framework [8].

Distribution scholars have described the Lindley beta distribution as a highly flexible distribution that has the potential to effectively eliminate the problems of overdispersion and trailing zeros that are often found in count data.

This is done by integrating the zero-inflation component and the beta-Lendley distribution. This new statistical method allows for a more accurate estimation of parameters, improving statistical inferences, and moving on to enable the Bayesian analysis framework to integrate all the previous information used and get rid of the obstacles of dispersion and excess zeros [9].

The components that go into forming the Bayesian model for all zero-number data inflated using the beta-Lendley distribution are the zero-order component and the beta-Lendley distribution component, which is built in order to design the distribution of non-zero numbers, which contains the following probability density function and the combined probability function with the elements, which are Product model then moving to Bayesian analysis of zero-inflated count data using beta [10].

Many researchers within this field have put forward a set of advantages and limitations and said that the Bayesian framework used for the beta-Lendley distribution approach gives it many advantages, the most important of which is that it is a versatile model that can accommodate a certain range of degrees of overdispersion and zero inflation. The framework is suitable for a wide range of applications [11–12]. Bayesian analysis also provides posterior distributions for model parameters, enabling probabilistic interpretations and quantification of uncertainty. The model can be used to make accurate predictions of future observations, taking into account the properties of zero inflation. Excessive dispersion of data [13].

However, they identified some limitations to this approach in terms of computational complexity. The Bayesian estimation process, especially MCMC sampling, can be computationally intensive, especially for large data sets. The choice of prior distributions for model parameters can affect the results, and care must be taken to ensure goodness-of-fit. Given the above information, the complex structure of the Lindley beta distribution and the default

framework may make the model less intuitive for some users, requiring a deeper understanding of basic statistical principles.

## III. MYTHODOLOGY:

The description of this methodology focuses specifically on relying on the Bayesian model in order to process zero-inflated numerical data using the beta-Lendley distribution[14]. It is divided into three main structures:

### 1. Through zero regression:

This component is used to model the excess probability of zero values and it assumes that the probability of obtaining a zero value follows the Bernoulli distribution with the parameter  $p$ :

$$p(y=0)=p+(1-p) * f(y=0 | \theta) \quad (1)$$

where  $f(y = 0 | \theta)$  is the probability of zero in the beta-Lindley distribution with parameters  $\theta$ .

### 2. Know the component of the beta-Lindley distribution:

In order to model the basic distribution of non-zero numerical data[15], the Beta-Lindley distribution formulas are defined according to the following:

$$(y | \alpha, \beta) = (1+y)^{\alpha-1} * (1+\beta)^{\beta-1} \left( \frac{\Gamma(\alpha)}{* (1+\alpha+\beta)^{\alpha+\beta}} \right) \quad (2)$$

where  $\alpha$  and  $\beta$  are the parameters of the beta-Lendley distribution. found that the weighting function for the full model is the product of the two components of the zero regression and the beta-Lindley distribution:

$$f(y | \alpha, \beta) = (1+y)^{(\alpha-1)*(1+\beta)(\beta-1)/(\Gamma(\alpha) * (1+\alpha+\beta)^{(\alpha+\beta)})} \quad (3)$$

where  $\alpha$  and  $\beta$  are the parameters of the beta-Lendley distribution. found that the weighting function for the full model is the product of the two components of the zero regression and the beta-Lindley distribution:

$$L(p, \alpha, \beta | y) = \prod [p+(1-p) * f(y | \alpha, \beta)]^I(y=0) * [(1-p) * f(y | \alpha, \beta)]^I(y>0) \quad (4)$$

where  $I(\cdot)$  is the indicator function which takes the value 1 if the condition is true, and 0 if it is false [16].

To access procedures for probability distributions by estimating the parameters  $p$ ,  $\alpha$ , and  $\beta$  related to Bayesian inference methods such as Markov Chain Monte Carlo (MCMC) until we reach direct statistical conclusions.

In the case of data on the number of insects within the soil environment according to specific samples [17], it was shown that it contains within its data empty inflated numerical characteristics with a high range of zero values.

Therefore, we can estimate the parameters  $p$ ,  $\alpha$ , and  $\beta$  according to the proposed Bayesian model and real data, for example:

When  $p = 0.3$ , this is an indication that 30% of the values are zero due to external factors, and then  $\alpha = 2$  and  $\beta = 3$ , this is an indication of the shape of the beta Lindley distribution for non-zero values [18] and that these estimates are used in order to make a set of predictions and conclusions. With more accurate statistical significance compared to traditional models for digital data, and to verify the quality of the proposed model through the use of evaluation criteria such as the coefficient of determination known as ( $R^2$ ), or through the use of the Akaike Information Criterion (AIC), or through appropriate statistical testing mechanisms, it is also possible Comparing the results with other models for zero-inflated numerical data [19]. As part of another process to study the number of orders that the online store receives on a daily basis as an economic example, it is possible that there is a large group of days without any orders (zero values), which leads to the inflation of zero numerical data, and therefore the proposed hypothetical model can be used using a beta distribution. Let us deal with this data more efficiently where the zero regression component is:  $p(y = 0) = p + (1-p) * f(y = 0 | \alpha, \beta)$ .

The beta – Lindley distribution is :

$$f(y | \alpha, \beta) = (1+y)^{\alpha-1} * (1+\beta)^{\beta-1} / (\Gamma(\alpha) * (1+\alpha+\beta)^{\alpha+\beta})$$

–Weighting function :

$$L(p, \alpha, \beta | y) = \prod [p + (1-p) * f(y | \alpha, \beta)]^I(y=0) * [(1-p) * f(y | \alpha, \beta)]^I(y>0) \tag{5}$$

For Bayesian analysis, the appropriate prior distributions for the parameters  $p$ ,  $\alpha$ , and  $\beta$  are determined based on the previously available information. It estimates probability distributions of parameters from observed data through MCMC methods [20] to be applied to the entire daily order data, where it is logical that this graphical information shows an average of 2 orders per da.. Today, with 60% of days without orders and through the proposed Bayesian model, we might get estimates like:  $p = 0.6$ ,  $\alpha = 1.5$ , and  $\beta = 2$ .

These estimates thus indicate that 60% of the days were without orders due to external factors, and that the underlying distribution of non-zero orders follows a beta-Lindley distribution with specified parameters. These conclusions can be explained by using these estimates to make predictions about the number of orders expected in the coming days, taking into account the null-inflated nature of the data. We can also test hypotheses about the influence of economic or marketing factors on the pattern of orders, and thus the Bayesian approach provides a strong framework for statistical and probabilistic conclusions about this data. Complex economic [22].

#### IV. RESULTS

If an e-commerce company controls and manages a group of e-stores and this company notices that there are a large number of days in which the stores do not achieve any sales (number of sales = 0). The company wanted to model

the number of daily sales for each store using a zero-inflated data model to determine the percentage of days in which sales are zero and determine the sales rate. It had data on the number of daily sales for 150 days for a specific store, and this data showed that there were 100 days in which no sales were achieved. Any sales, while on the other days sales were achieved according to the Poisson distribution. The estimate of the percentage of days in which no sales were achieved was ( $p$ ) while the estimate of the daily sales rate when they occurred was ( $\lambda$ ). From Fig. 1 it is clear that Estimated  $p$  (estimated percentage of days without sales) 0.6277465400982124. This percentage is 62.77%, which represents the days when there were no sales. This indicates that the store faces a large number of days in which no sales are achieved at all. Therefore, knowing this percentage is very important for the company because it shows the frequency of this phenomenon, which helps in making decisions to improve Store performance these days. while Estimated  $\lambda$  (daily sales rate when sales occur): 1.3252615642180827. This rate indicates the value of the average number of daily sales on days in which sales occur, which is about 1.33 sales per day. This number helps in understanding sales performance on days in which it is not zero, which gives a more accurate picture of the actual demand on active days [23].

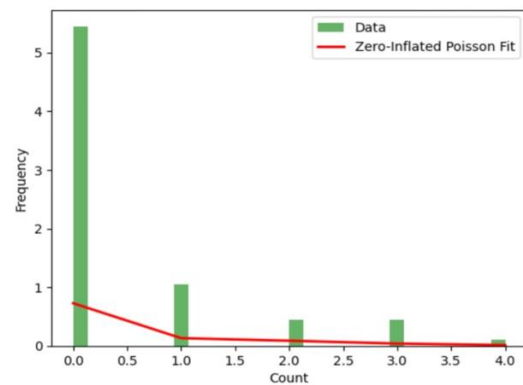


Fig. 1. Zero-Inflated Poisson Fit

According to Beta Lind, the results were according to Fig. 2. Estimated  $p$ : 0.5209459152924262. This percentage means that about 52.09% of the days there were no sales and this indicates that the store faces a large number of days in which there are no sales at all. And it was Estimated  $\alpha$ :

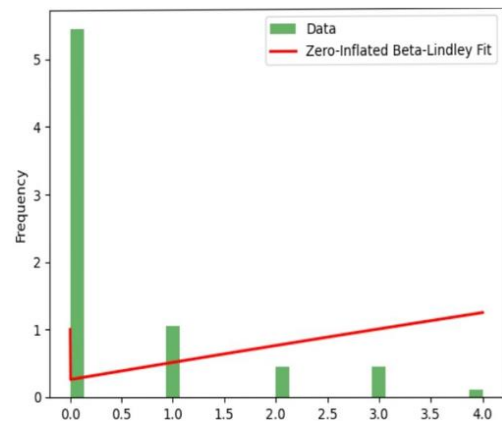


Fig. 2. Zero-Inflated Beta-Lindley Fit

2.4989200499437683 This represents the shape parameter of the Beta-Lindley distribution, showing how the underlying shape of the distribution is distributed. Estimated beta: 0.9585270877209962. This represents the average parameter in the Beta-Lindley distribution, and reflects how dispersed the data are around the expected value.

It turns out that the probability density function (PDF) of the Beta-Lindley distribution for some different values of the parameters is where Fig. 3. How does the probability density of the Beta-Lindley distribution change with the change in the values of the parameters `alpha` and `beta`

It has been shown that the beta-Lindley distribution is a flexible probability distribution that can effectively model zero-inflated count data, which is characterized by an excess of zeros and over-dispersion compared to standard count data models.

Beta-Lindley PDF file is provided by:

$$f(x; \theta, \theta) = (\theta + 1) * (\theta^2) * \frac{\binom{x}{lect}}{(1 + \theta * x)^{\theta + 2}} \quad (6)$$

where: x is the random variable (count data);  $\theta$  is the shape parameter of the beta-Lendley distribution;  $\alpha$  is the scale parameter of the beta-Lendley distribution.

Key Features and Benefits of Beta-Lindley PDF Flexibility in Modeling Zero Inflation and Overdispersion The Beta-Lindley distribution can accommodate both the excess of zeros (zero inflation) and overdispersion (higher-than-average variance) that are commonly observed in count data. This also allows for flexibility of the distribution. Capture complex data structures that exist in various fields, such as epidemiology, environment, and social sciences [24].

The shape parameter  $\theta$  controls the shape of the distribution, which affects the percentage of zeros and the overall dispersion of the data. In addition, the scale parameter  $\alpha$  affects the rate of decay of the probability mass function, which affects the spread of non-zero values.

Regarding the integration of Bayesian analysis, easily integrating the Lindley beta distribution into the Bayesian modeling framework, allowing the integration of prior information about the parameters, this Bayesian approach facilitated more robust parameter estimation, especially in the presence of limited data, and allows probabilistic interpretations of the model results.

This improves modeling and inference. Compared to traditional count data models, such as Poisson or negative binomial, the beta-Lindley distribution can provide more accurate parameter estimates and reliable model predictions for zero-inflated count data. These improved modeling and inference capabilities can lead to better informed decisions.

The cumulative distribution function (CDF) of the Beta-Lindley distribution for some different values of the parameters shows, according to Fig. 4, how the cumulative distribution of the Beta-Lindley distribution changes with the change in the values of the parameters `alpha` and `beta`.

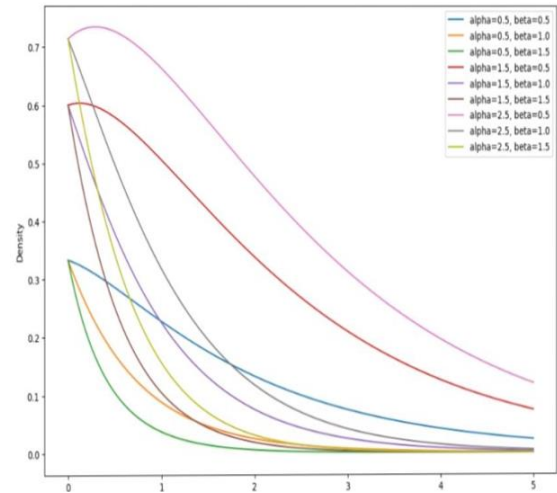


Fig. 3. Beta-Lindley PDF for Different Parameters

The comparison process gives a compatibility impression between the original distribution and the experimental data by using the zero-beta Lindley distribution approach to arrive at the zero-inflated model with the experimental data.

Discussing this result with PDF and CDF data and charts gives us the extent to which changing parameter values affects the shape of the distribution.

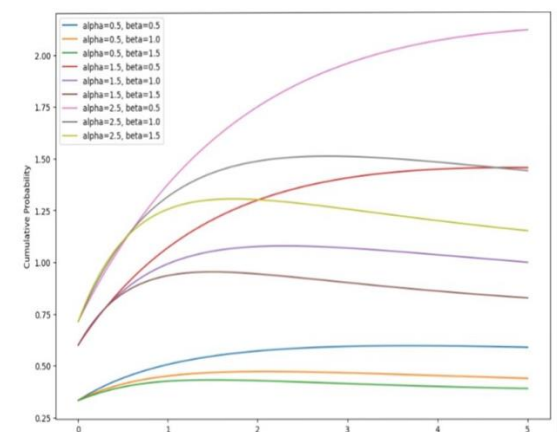


Fig. 4. Beta-Lindley CDF for Different Parameters

For beta-Lindley, we obtain the cumulative distribution function (CDF) through the following equation:

$$F(x; \theta, lect) = 1 - (1 + lect * x)^{-\theta - 1} \quad (7)$$

where: x Random variable (count data);  $\theta$  shape parameter beta Lendley distribution;  $\alpha$  is the scale parameter beta Lendley distribution.

The behavior of Beta-Lindley CDF with different parameter values is as follows:

To reach the effect of the shape parameter ( $\theta$ ) with a continuous increase within the shape parameter  $\theta$  the CDF becomes more “S-shaped” and the distribution becomes more dispersed and higher values of  $\theta$  lead to a slower increase in the CDF.

This will lead to a gradual rise in the probability of observing smaller values of the random variable  $x$ , where lower values of  $\theta$  indicate a rapid increase in CDF and are indicative of a higher probability of observing larger values of  $x$ .

Regarding the effect of the scale parameter, the scale parameter  $\lambda$  mainly affects the rate of change of the CDF and as larger values of  $\lambda$  lead to a faster increase in the CDF, which indicates a higher probability of observing smaller values of  $x$  as smaller values of  $\lambda$  lead to a slower increase in the CDF, which It indicates a higher probability of observing larger values of  $x$  The interaction between the shape parameter  $\theta$  and the scale parameter  $\lambda$  allows the Beta-Lindley CDF to capture a wide range of distributional shapes, making it suitable for modeling different types of zero-inflated count data [25].

Where the CDF always ranges from 0 to 1, which represents the cumulative probability distribution. For  $x = 0$ , the CDF is always less than 1, which indicates the probability of observing a value of zero. The CDF appears to be monotonically increasing, which means that as  $x$  increases, The CDF also increases and the shape and rate of increase of the CDF can be adjusted by manipulating the parameters  $\theta$  and  $\lambda$  to suit the specific properties of the zero-amplified count data being modeled [22–24].

Fit to the data Fig. 2 shows how well the zero-inflated distribution fits the real data, giving an idea of how accurately the model describes the data.

## V. CONCLUSIONS

This study gives a Bayesian approach to modeling zero-inflated count data using a flexible beta-Lindley distribution. This established mechanism effectively solves the challenges posed by excessive zeros and excessive dispersion that are constantly encountered in real-world count data from variable and diverse fields. The ability to accurately capture complex data structures with excessive zeros and overdispersion through the inflated zero component and the Lindley beta distribution component while incorporating prior information leads to more reliable parameter estimates especially in the presence of limited data and to measure uncertainty by estimating the posterior distributions of model data and parameters.

This gives more intuitive impressions of probability while arriving at a correct forecast condition taking into account future observations by taking into account the properties of zero inflation and over-dispersion of the data.

When we reach that stage, we reach the process of conclusively proving the ability of this model through simulations and applications on real data sets. The results show that our approach better models zero-inflated count data and thus outperforms traditional count data models. Therefore, it can be said that this research has applications.

This is very important and promising in various fields where multi-number data is common, as Bayesian analysis using the Lindley beta distribution provides a valuable and promising tool for researchers and practitioners to build

deeper insights while making more informed decisions based on complex number data.

## REFERENCES

- [1] Cowles, M.K., Carlin, B.P., 1996. Markov Chain Monte Carlo convergence diagnostics: a comparative review. *J. Amer. Statist. Assoc.* 91, 833–904.
- [2] Dahiya, R.C., Gross, A.J., 1973. Estimating the zero class from a truncated Poisson sample. *J. Amer. Statist. Assoc.* 68, 731–733.
- [3] Dempster, A.P., Laird, N.M., Rubin, D.B., 1977. Maximum likelihood from incomplete data via the EM algorithm (with discussion). *J. Roy. Statist. Soc. B* 39, 1–38.
- [4] Gilks, W.R., Wild, P., 1992. Adaptive rejection sampling for Gibbs sampling. *Appl. Statist.* 41, 337–348.
- [5] Gelfand, A.E., Smith, A.F.M., 1990. Sampling based approaches to calculating marginal densities. *J. Amer. Statist. Assoc.* 85, 398–409.
- [6] Gupta, P.L., Gupta, R.C., Tripathi, R.C., 1996. Analysis of zero-adjusted count data. *Comput. Statist. Data Anal.* 23, 207–218.
- [7] Barreto-Souza, W. a. (2010). The beta generalized exponential distribution. *Journal of statistical Computation and Simulation*, 80(2), 159–172.
- [8] Barreto-Souza, W. a. (2011). Some results for beta Frechet distribution. *Communications in Statistics—Theory and Methods*, 40(5), 798–811.
- [9] Barreto-Souza, W. a. (2011). The Weibull-geometric distribution. *Journal of Statistical computation and Simulation*, 81(5), 645–657.
- [10] Berger, J. O. (2013). *Statistical decision theory and Bayesian analysis*. Springer Science & Business Media.
- [11] Cordeiro, G. M. (2011). Closed-form expressions for moments of a class of beta generalized distributions. *Brazilian journal of probability and statistics*, 25(1), 14–33.
- [12] Cordeiro, G. M. (2013). The beta generalized Rayleigh distribution with applications to lifetime data. *Statistical papers*, 54(1), 133–161.
- [13] Merovci, F. a. (2014). The beta-Lindley distribution: properties and applications. *Journal of Applied Mathematics*, 2014, 10.
- [14] Nadarajah, S. a. (2004). The beta Frechet distribution. *Far east journal of theoretical statistics*, 14(1), 15–24.
- [15] Gurmu, S., 1997. Semiparametric estimation of hurdle regression models with an application to Medicaid Utilization. *J. Appl. Econometrics* 12, 225–242.
- [16] Hall, D.B., 2000. Zero-inflated Poisson and binomial regression with random effects: a case study. *Biometrics* 56, 1030–1039.
- [17] Heilbron, D.C., 1994. Zero-altered and other regression models for count data with added zeroes. *Biometrical J.* 36, 531–547.
- [18] Heilbron, D.C., Gibson, D.R., 1990. Shared needle use and health beliefs concerning AIDS: regression modeling Of zero-heavy count data. Poster session. Sixth International Conference on AIDS, San Francisco, CA.
- [19] Hinde, J., Demetrio, C., 1998. Overdispersion: models and estimation. *Comput. Statist. Data Anal.* 27, 151–170.
- [20] Lambert, D., 1992. Zero-inflated Poisson regression, with an application to defects in manufacturing. *Technometrics* 34, 1–14.
- [21] Li, C.S., Lu, J.C., Park, J., Kim, K.M., Brinkley, P.A., Peterson, J., 1999. A multivariate zero-inflated Poisson Distribution and its inference. *Technometrics* 41 (1), 29–38.
- [22] Lu, J.C., Bhattacharyya, G.K., 1991. Inference procedures for a bivariate exponential model of Gumbel based on Life test of system and components. *J. Statist. Plann. Inference* 27, 383–396.
- [23] Miaou, S.P., 1994. The relationship between truck accidents and geometric design of road sections. Poisson versus Negative binomial regressions. *Accident Anal. Prevention* 26, 471–482.
- [24] Ridout, M., Demetrio, C.G.B., Hinde, J., 1998. Models for count data with many zeros. *International Biometric Conference*, Cape Town.
- [25] Saei, A., McGilchrist, C.A., 1997. Random threshold models applied to zero class data. *Austral. J. Statist.* 39, 5–16.



# Моделирование дискретизации весов в нейронной сети на основе оптимизированных нанокompозитных мемристоров

А. Н. Мацукатова<sup>1,2</sup>, А. Д. Трофимов<sup>1,3</sup>, М. А. Рябова<sup>3</sup>,  
А. А. Несмелов<sup>1</sup>, А. В. Емельянов<sup>1,3</sup>

<sup>1</sup>Национальный исследовательский центр «Курчатовский институт»

<sup>2</sup>Московский государственный университет имени М. В. Ломоносова

<sup>3</sup>Московский физико-технический институт (НИУ МФТИ)

an.matcukatova@physics.msu.ru

**Аннотация.** Мемристоривные устройства являются многообещающими компонентами для нейроморфных вычислительных систем, в частности в качестве искусственных синапсов. Однако ненадежность и стохастичность мемристоривных характеристик становится основным ограничивающим фактором их промышленного использования в новейших вычислительных системах. В работе исследуются подходы к решению данной проблемы: совершенствование структуры мемристоров и подбор параметров нейронной сети для ее последующей эффективной аппаратной реализации. Были разработаны и всесторонне исследованы мемристоривные устройства на основе гибридного нанокompозитного слоя парилена-МоОх (PPX-Mo). Такие устройства продемонстрировали улучшенные характеристики в сравнении со своими аналогами на основе чистого парилена: более низкие напряжения переключения, меньшую дисперсию и большую пластичность. Для проверки возможности использования PPX-Mo мемристоров в роли синапсов была выбрана показательная задача классификации рукописных чисел MNIST. На ее примере удалось определить оптимальное число резистивных состояний мемристоров (то есть варьировалась разрядность весов нейронной сети – так называемая дискретизация или квантизация) для успешного решения выбранной задачи классификации. Полученные результаты могут быть полезны для разработки эффективных и надежных нейроморфных систем, основанных на частично ненадежных аналоговых элементах.

**Ключевые слова:** мемристор, резистивное переключение, нейроморфные вычислительные системы, аппаратные нейронные сети, дискретизация весов

## I. ОБЗОР ЛИТЕРАТУРЫ

Нейроморфные вычислительные системы (НВС), состоящие из нейронов (вычислительных элементов) и синапсов (элементов памяти), привлекают все большее внимание исследователей в связи с их низким энергопотреблением, биоподобностью и возможностью обойти ограничения стандартной архитектуры фон Неймана [1]. Их перспективность уже подтверждена первыми аппаратными реализациями НВС [2]. Однако

Работа выполнена при частичной финансовой поддержке РФФИ (проект № 20-79-10185) в части синтеза и экспериментального изучения мемристоривных структур. Мацукатова А.Н. является стипендиатом Фонда развития теоретической физики и математики «БАЗИС» (№ 19-2-6-57-1).

для производства каждого нейрона и синапса в таких НВС требуется большое количество транзисторов. Поэтому в последние годы активно исследуются так называемые мемристоры (сопротивления с эффектом памяти). Одного мемристора может быть достаточно для эмуляции работы как нейрона, так и синапса в НВС [1]. Типичный мемристор представляет собой сэндвич-структуру металл/активный слой/металл, которая способна изменять свое состояние в некотором резистивном окне от высокоомного состояния  $R_{off}$  до низкоомного  $R_{on}$  в зависимости от приложенного порогового напряжения и/или протекающего через нее электрического заряда [1]. После резистивного переключения (РП) мемристор может энергонезависимо сохранять установившееся сопротивление. При использовании мемристора в роли синапса его сопротивление играет роль синаптического веса. Большинство мемристоривных структур переключается посредством формирования/разрушения проводящих мостиков (филаментов) в диэлектрическом активном слое [3]. Такие структуры называются филаментарными. Одним из главных препятствий эффективного внедрения филаментарных мемристоров в НВС становится стохастичность их характеристик.

В данной работе исследуется возможность улучшения мемристоривных характеристик перспективных мемристоров на основе парилена (poly-*p*-xylylene, PPX). На данный момент эффект РП в структурах на основе PPX уже продемонстрирован [4, 5]. Такие мемристоры обладают многоуровневым РП, хорошо удерживают резистивные состояния, могут изменять сопротивление по биоподобным правилам. Однако мемристоры на основе PPX переключаются по филаментарному механизму, поэтому их РП достаточно стохастичны. Можно попробовать стабилизировать РП путем внедрения различных дефектов в активный слой, в котором происходит переключение [6]. В связи с вышесказанным, были разработаны и исследованы нанокompозитные мемристоры на основе PPX с внедренными частицами МоОх (PPX-Mo) [7]. Было изучено влияние внедренных дефектов на механизм РП и непосредственно на сами переключения. Глубокое понимание механизма РП при структурных изменениях мемристоривных структур важно для дальнейшей оптимизации структуры и улучшения мемристоривных

характеристик. Кроме этого, характеристики полученных структур были внедрены в модель НВС. Основной задачей стала проверка пригодности PПХ-Мо мемристоров для нейроморфных применений. На примере стандартной задачи классификации рукописных чисел MNIST было определено оптимальное число резистивных состояний PПХ-Мо мемристоров, необходимое для ее успешного решения.

## II. МЕТОДЫ

Мемристоры представляли собой сэндвич-структуру Cu/PПХ-МоO<sub>x</sub>/ITO (рис. 1а). Синтез нанокompозитного слоя ППХ был выполнен методом газовой полимеризации, подробнее процесс синтеза описан в [8] с тем отличием, что мишень для наполнителя в данном случае была из MoO<sub>3</sub>. Концентрация включений составила 50%, толщина активного нанокompозитного слоя – 640 нм. Затем образец отжигался в течение 2 часов при 200 °C на воздухе для стабилизации его электрофизических свойств и удаления примесей с низким молекулярным весом. Верхние медные контакты размерами 0.2 × 0.5 мм<sup>2</sup> и толщиной ~ 200 нм наносились магнетронным распылением меди на нанокompозитный слой PПХ-Мо через теньевую маску с помощью магнетрона Torr MagSput 4G2-RF/DC.

Структурные исследования проводились с использованием сканирующего/просвечивающего электронного микроскопа Osiris (Thermo Fisher Scientific, США), оснащенного кольцевым детектором темного поля (HAADF) (Fischione, США) и энергодисперсионным рентгеновским спектрометром (EDX) Bruker Super-X (Bruker, США) для анализа химического состава.

Мемристивные характеристики измерялись с помощью аналитической зондовой станции 3S SWIN EPS4. Импульсы напряжения подавались на верхний электрод при заземленном нижнем электроде от источника-измерителя Keithley 2636 В, запрограммированного в среде LabView. Были установлены токи ограничения  $I_{\max} = 1$  мА и  $I_{\min} = -50$  мА во избежание теплового разрушения образцов. Импедансная спектроскопия структур была выполнена с помощью прецизионного анализатора импеданса Wayne Kerr 6500 В в диапазоне частот  $f$  от 100 Гц до 13 МГц с амплитудой переменного сигнала 50 мВ.

## III. ОСНОВНЫЕ РЕЗУЛЬТАТЫ

Перед измерением мемристивных характеристик был проведен анализ структуры полученных образцов. В активном слое были обнаружены как одиночные включения молибдена, так и кластеры MoO<sub>x</sub> (рис. 1а).

На рис. 1б представлены типичные вольт-амперные характеристики (ВАХи) исследуемых структур. При приложении положительного напряжения к верхнему контакту мемристор из первоначального высокоомного состояния  $R_{\text{off}}$  переходит в низкоомное состояние  $R_{\text{on}}$ . Низкоомное состояние сохраняется при малых отрицательных напряжениях и переходит обратно в  $R_{\text{off}}$  при больших. На рис. 1в представлено 17 резистивных состояний, которые после первоначальной установки удалось удерживать без приложения дополнительного высокого напряжения. В данном эксперименте

сопротивление отслеживалось на напряжении чтения  $U_{\text{read}} = 0.1$  В, недостаточном для изменения состояния мемристора.

Можно предположить, что в PПХ-Мо мемристорах PП происходит с образованием/разрывом металлического мостика, как и в мемристорах на основе чистого PПХ [4, 5]. Данную гипотезу подтверждают результаты, полученные с помощью импедансной спектроскопии (рис. 1г). В проводящем состоянии годограф демонстрирует почти вертикальную линию, что свидетельствует о наличии металлического мостика между электродами. Возможная причина малой стохастичности PП и низких напряжений переключения PПХ-МоO<sub>x</sub> структур заключается в локализации электрического поля кластерами MoO<sub>x</sub>, что приводит к стабилизации процесса образования/разрушения филамента.

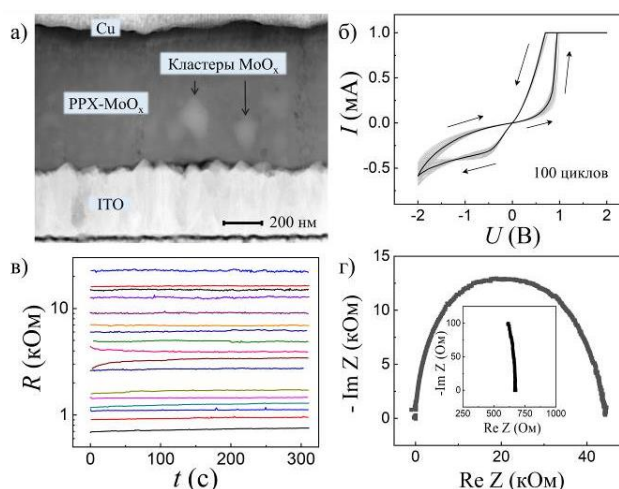


Рис. 1. а) Просвечивающая микроскопия мемристора в темном поле, демонстрирующая его сэндвич-структуру и кластеры MoO<sub>x</sub>; б) 100 ВАХ с выделенной черным цветом медианной кривой; в) демонстрация возможности энергонезависимого удержания установленных резистивных состояний; г) импедансная спектроскопия высокоомного и низкоомного (на вставке) состояния

Для проверки возможности применения исследуемых мемристоров в НВС была выбрана задача классификации рукописных цифр MNIST с полностью связанной двухслойной сетью 784×512×10 (рис. 2). Существует два общепринятых способа обучения мемристивных НВС: on-chip (in-situ, аппаратное обучение сети на чипе) и off-chip (ex-situ, программное определение точных весов для обученной сети и их последующий перенос на аппаратный чип, рис. 2) [9]. Было выбрано моделирование off-chip обучения. В таком типе обучения уменьшается влияние разброса мемристивных характеристик на точность работы сети, что особенно важно для филаментарных мемристоров. Кроме этого, при off-chip обучении можно единожды программно обучить сеть и многократно использовать полученные веса при моделировании сетей на мемристорах с различными состояниями, что ускоряет исследование. Результаты, полученные для off-chip обучения, помогут оценить успешность on-chip обучения: если точность классификации программной и мемристивной сетью будут приблизительно равны, то это укажет на возможность on-chip обучения без снижения точности.

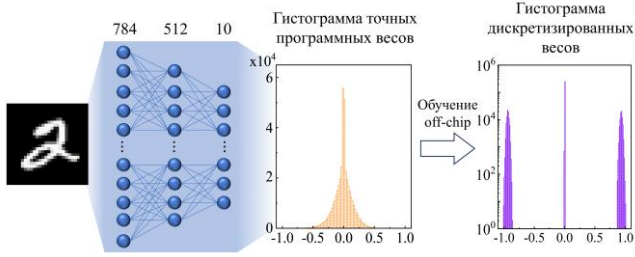


Рис. 2. Схематическое изображение сети и off-chip обучения.

Улучшенная стабильность структур привела к возможности обнаружения большого количества резистивных состояний. На рис. 3а продемонстрирован процесс установки 32 резистивных состояний РРХ-Мо мемристоров со средним коэффициентом вариации (coefficient of variation, CV) 0.4 %. Такое количество резистивных состояний указывает на отличную пластичность полученных мемристивных структур. Состояния устанавливались независимо друг от друга по алгоритму записи-чтения (write-verify), поэтому количество состояний можно было регулировать и исследовать влияние дискретизации (квантизации) весов на точность классификации смоделированной мемристивной сети. Дискретизированные резистивные состояния вводились в модель следующим образом. В-первых, сопротивления с рис. 3а были переведены в нормализованные проводимости (черные точки на рис. 3б). Так как проводимость мемристора всегда положительна, каждый синаптический вес представлялся разностью двух нормализованных проводимостей, выбранных из 32 возможных состояний (или меньшего количества для большей дискретизации). Таким образом, удавалось получить как отрицательные, так и положительные веса сети. Полученные при программном обучении точные веса приравнивались к ближайшим мемристивным синаптическим весам с учетом вариативности мемристивных характеристик (для каждого мемристора использовалось экспериментальное среднее значение  $CV = 0.4\%$ ).

Сеть обучалась программно 45 эпох, до достижения максимальной точности классификации  $\sim 98.3\%$  (рис. 3в). Обучение сети и последующая дискретизация весов со всеми возможными количествами мемристивных резистивных состояний (32, 16, 8, 4 на рис. 3б) повторялись 10 раз, полученные точности классификации представлены на рис. 3г. Заметное снижение точности классификации произошло только при использовании 4 резистивных состояний (средняя точность уменьшилась с  $\sim 98.3\%$  до  $\sim 98\%$ ). Таким образом, минимальное количество состояний, при котором не наблюдается снижение точности распознавания — 8. С течением времени возможна деградация любого мемристивного устройства. В таком случае точность установки состояний снижается (CV увеличивается), и меньшее количество состояний гарантирует отсутствие их пересечений. Можно утверждать, что 8 состояний является оптимальным количеством состояний для данной сети и задачи.

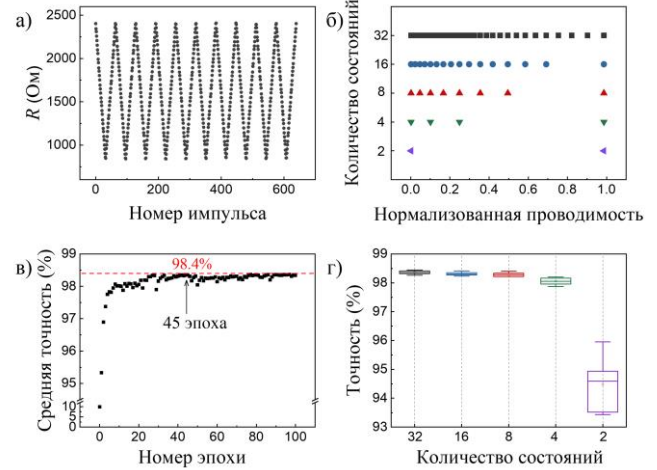


Рис. 3. а) Демонстрация 32 резистивных состояний, полученных с помощью алгоритма записи-чтения. б) Дискретизация нормализованных проводимостей, которые затем используются для расчета мемристивных сети весов. в) Подбор оптимального количества эпох обучения сети. г) Точность классификации для нейронной сети с дискретизированными весами, в случае 2 возможных состояний был выбран лучший результат (с оптимальным порогом бинаризации)

При использовании двух резистивных состояний (фиолетовые точки на рис. 3б) процесс дискретизации был изменен. Если в данном случае округлить все точные веса до ближайшего мемристивного синаптического веса (т.е. либо до 0, либо до  $\pm 1$ ), то абсолютное большинство весов будет равно 0, поскольку большинство точных программных весов находится в окрестности 0 (как видно из рис. 2). В результате точность классификации резко снизится. Поэтому в данном случае для процесса дискретизации весов было необходимо ввести порог бинаризации  $w_{thr}$ . Каждый вес сети  $w_i$  сравнивался с порогом и принимал значение 0 только в случае, если  $|w_i| < w_{thr}$ . Несмотря на возможность выбора оптимального порога  $w_{thr} = 0.1$  (рис. 4а), точность классификации с 2 состояниями все равно существенно уступает остальным (рис. 3г). Однако в данном случае крайние резистивные состояния могут быть заданы с намного меньшей точностью, чем изначально задавались 32 состояния. Уменьшение точности установки состояний не приведет к пересечению состояний, что позволит сохранить точность классификации даже при значительной деградации мемристивных структур (рис. 4б).

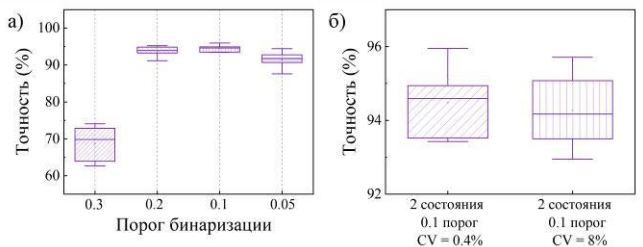


Рис. 4. Точность классификации для нейронной сети с 2 возможными резистивными состояниями а) с различными порогами бинаризации, б) с различным коэффициентом вариации резистивных состояний

## IV. ЗАКЛЮЧЕНИЕ

В данной работе исследованы новые мемристивные устройства на основе активного слоя PРХ-МоО<sub>x</sub>. Внедрение МоО<sub>x</sub> привело к низкой стохастичности мемристивных характеристик, низким напряжениям переключения, высокой пластичности. Было выдвинуто предположение о механизме переключения и влиянии МоО<sub>x</sub> на функционирование мемристоров. Полученные экспериментальные результаты могут способствовать разработке новых усовершенствованных мемристивных структур. Кроме этого, характеристики исследуемых мемристоров были внедрены в нейронную сеть, продемонстрирован потенциал использования PРХ-МоО<sub>x</sub> мемристоров в качестве синапсов в НВС. Выявлено оптимальное количество резистивных состояний (8 состояний) для задачи классификации MNIST и сети 784×512×10, при котором не наблюдается снижения программной точности классификации. Также продемонстрировано, что 2 состояния могут стать более оптимальным вариантом для стохастических мемристоров, подверженных деградации. Полученные результаты указывают на перспективность создания НВС на основе PРХ-МоО<sub>x</sub> как с off-chip, так и с on-chip обучением.

## БЛАГОДАРНОСТЬ

Измерения выполнены с использованием оборудования ресурсных центров НИЦ «Курчатовский институт».

## СПИСОК ЛИТЕРАТУРЫ

- [1] Wang J., Zhuge F. Memristive Synapses for Brain-Inspired Computing // *Adv. Mater. Technol.* 2019. Vol. 4, № 3. P. 1800544, doi: 10.1002/admt.201800544.
- [2] Akopyan F., Sawada J., Cassidy A., Alvarez-Icaza R., Arthur J., Merolla P., Imam N., Nakamura Y., Datta P., Nam G.-J., Taba B., Beakes M., Brezzo B., Kuang J.B., Manohar R., Risk W.P., Jackson B., Modha D.S. TrueNorth: Design and Tool Flow of a 65 mW 1 Million Neuron Programmable Neurosynaptic Chip // *IEEE Trans. Comput. Des. Integr. Circuits Syst.* 2015. Vol. 34, № 10. P. 1537, doi: 10.1109/TCAD.2015.2474396.
- [3] Ielmini D. Resistive switching memories based on metal oxides: mechanisms, reliability and scaling // *Semicond. Sci. Technol.* 2016. Vol. 31, № 6. P. 063002, doi: 10.1088/0268-1242/31/6/063002.
- [4] Minnekhanov A.A., Emelyanov A.V., Lapkin D.A., Nikiruy K.E., Shvetsov B.S., Nesmelov A.A., Rylkov V.V., Demin V.A., Erokhin V.V. Parylene Based Memristive Devices with Multilevel Resistive Switching for Neuromorphic Applications // *Sci. Rep.* 2019. Vol. 9, № 1. P. 10800, doi: 10.1038/s41598-019-47263-9.
- [5] Minnekhanov A.A., Shvetsov B.S., Martyshev M.M., Nikiruy K.E., Kukueva E.V., Presnyakov M.Y., Forsh P.A., Rylkov V.V., Erokhin V.V., Demin V.A., Emelyanov A.V. On the resistive switching mechanism of parylene-based memristive devices // *Org. Electron.* 2019. Vol. 74. P. 89, doi: 10.1016/j.orgel.2019.06.052.
- [6] Banerjee W., Liu Q., Hwang H. Engineering of defects in resistive random access memory devices // *J. Appl. Phys.* 2020. Vol. 127. № 5. P. 051101, doi: 10.1063/1.5136264.
- [7] Minnekhanov A., Matsukatova A., Trofimov A., Nesmelov A., Zavyalov S., Demin V., Emelyanov A. Reliable Memristive Synapses Based on Parylene-MoO<sub>x</sub> Nanocomposites for Neuromorphic Applications // *ACS Appl. Mater. Interfaces.* 2023. Vol. 15, № 47. P. 54996, doi: 10.1021/acsami.3c13956.
- [8] Matsukatova A.N., Vdovichenko A.Y., Patsaev T.D., Forsh P.A., Kashkarov P.K., Demin V.A., Emelyanov A.V. Scalable nanocomposite parylene-based memristors: Multifilamentary resistive switching and neuromorphic applications // *Nano Res.* 2023. Vol. 16, № 2. P. 3207, doi: 10.1007/s12274-022-5027-6.
- [9] Misba W.A., Lozano M., Querlioz D., Atulasimha J. Energy Efficient Learning With Low Resolution Stochastic Domain Wall Synapse for Deep Neural Networks // *IEEE Access.* 2022. Vol. 10. P. 84946, doi: 10.1109/ACCESS.2022.3196688.

# Отказоустойчивый нейросетевой датчик влажности материала

С. В. Артемова<sup>1</sup>,  
А. И. Ладынин<sup>2</sup>,  
А. Г. Шмелева<sup>3</sup>,  
Н. С. Ершов<sup>4</sup>

МИРЭА – Российский  
технологический университет

<sup>1</sup>sartemova@yandex.ru,

<sup>2</sup>andrey.ladyinin@hotmail.com,

<sup>3</sup>ann\_shmeleva@mail.ru,

<sup>4</sup>ershov@mirea.ru

М. А. Каменская

Тамбовский государственный  
технический университет

art\_mari@bk.ru

Д. Н. Лапаев

Нижегородский государственный  
технический университет  
им. Р.Е. Алексеева

dnlapaev@mail.ru

**Аннотация.** В статье рассматривается методика создания отказоустойчивых нейросетевых датчиков влажности пастообразных и сыпучих материалов, применяемых в информационно-управляющих системах. На вход нейросетевого датчика влажности подается информация от первичных измерительных преобразователей, контролирующих прохождение процесса сушки материалов в реальном режиме времени внутри сушильных установок различного типа. С выхода датчика получают значение относительной влажности материала. В работе представлена методика, основанная на авторской классификации ситуаций отказов первичных измерительных преобразователей. В зависимости от класса создаются модели оценки относительной влажности материалов путем обучения нейронных сетей с учетом отказов. Приводятся примеры создания нейросетевых датчиков влажности для сыпучих и пастообразных материалов. На основе поступающей от них информации принимается решение об управлении сушильной установкой с минимизацией потерь качества и производительности процесса сушки.

**Ключевые слова:** нейронные сети, оценка влажности, информационно-управляющая система, модель классификации

## I. ВВЕДЕНИЕ

На сегодняшний день, математическое и компьютерное моделирование является одним из наиболее актуальных методов обработки и анализа информации [1]. Одним из подходов к оперативной оценке влажности пастообразных и сыпучих материалов является использование отказоустойчивого нейросетевого датчика, позволяющего на основании информации о протекании процесса сушки оценивать влажность материала внутри сушильной установки. При этом на его вход подается информация от первичных измерительных преобразователей, контролирующих прохождение процесса сушки материалов в реальном режиме времени внутри сушильных установок различного типа. С выхода получают значение относительной влажности.

В основе алгоритма оценки относительной влажности материала лежит методика создания

отказоустойчивого нейросетевого датчика оценки относительной влажности, основанной на самодиагностическом контроле [2]. Применение этого алгоритма при функционировании датчика позволяет продолжать оценивать влажность материала в условиях отказов первичных измерительных преобразователей, входящих в его состав с точностью приемлемой для решения поставленной задачи.

Свойство самодиагностического контроля – это одно из самых важных свойств отказоустойчивого нейросетевого датчика влажности материала. Отказоустойчивый нейросетевой датчик влажности составляет основу функционирования информационно-измерительной системы оценки влажности материала в процессе его сушки. Применение отказоустойчивых нейросетевых датчиков может представлять перспективу развития в этой области.

## II. ПОСТАНОВКА ЗАДАЧИ

Рассмотрим методику создания отказоустойчивого нейросетевого датчика относительной влажности материала на примере двух сушильных установок вальцеленточного и барабанного типа. Разрабатываемый датчик представляет собой информационно-измерительную систему, информация в которую поступает от первичных измерительных преобразователей, устанавливаемых непосредственно на сушильной установке и предназначенных для мониторинга процесса сушки. При построении информационно-измерительной системы необходимо уделять особое внимание погрешностям составляющих ее первичных измерительных преобразователей и отказов их работоспособности, так как это оказывает существенное влияние на оценку относительной влажности материала.

Диагностический самоконтроль является одним из важных свойств отказоустойчивого нейросетевого датчика влажности. Он дает возможность сохранять характеристики в допустимых для потребителя пределах при возникновении отказов первичных измерительных преобразователей и его информационных каналов.

Совершенствование существующих интеллектуальных датчиков осуществляется путем доработки методов формального описания и анализа измерительных процедур за счет создания методик оптимальной обработки данных на основе более сложных статистических моделей – обученных нейронных сетей, используемых в качестве измерительной процедуры. Это становится возможным только при наличии большого массива реальных данных, характеризующих поведение исследуемого объекта с учетом его режимов работы, смен состояний функционирования, отказом компонентов информационно – измерительной системы составляющей основу разрабатываемого датчика.

### III. МЕТОДИКА СОЗДАНИЯ ОТКАЗОУСТОЙЧИВОГО НЕЙРОСЕТЕВОГО ДАТЧИКА ВЛАЖНОСТИ

При разработке методики создания отказоустойчивых нейросетевых датчиков влажности предусматривается определенная последовательность действий:

1. определяются возможные режимы работы сушильных установок;
2. формулируются множество ситуаций функционирования сушильных установок;
3. определяются нормальные и экстремальные ситуации функционирования при различных режимах работы;
4. определяются классы отказов первичных измерительных преобразователей с учетом нормальных и различных экстремальных ситуаций функционирования; вводятся следующие классы отказов: безотказная работа всех первичных измерительных преобразователей, мягкий отказ, средний отказ, жесткий отказ;
5. обучаются нейронные сети согласно классам отказов;
6. создается фреймовая модель функционирования отказоустойчивого нейросетевого датчика влажности.

Рассмотрим применение методики создания отказоустойчивого нейросетевого датчика влажности для барабанная кожухотрубная сушильная

Установка (БКСУ) фирмы VetterTec GmbH в которой производится сушка барды. Первичные измерительные преобразователи, осуществляющие мониторинг процесса сушки установлены на исследуемой сушильной установке. С целью повышения качества барды и процесса производства необходимо в реальном режиме времени оценивать ее относительную влажность с использованием разработанного отказоустойчивого нейросетевого датчика.

Основными параметрами процесса сушки при построении измерительных процедур отказоустойчивого нейросетевого датчика являлись: исходная влажности барды на входе БКСУ, % сухого вещества; температура барды на выходе БКСУ, °С; давление теплоносителя, атм.; температура воздуха в БКСУ, °С; мощность вытяжных вентиляторов, %; содержание кислорода в

воздухе внутри БКСУ, %; температура выпара, °С; нагрузка электродвигателя БКСУ, %; в модели оценки относительной влажности они соответственно обозначаются –  $x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8$  и являются входами нейронной сети.

Измерительная процедура оценки влажности материала реализована на базе искусственной нейронной сети. Нейронная сеть имеет архитектуру многослойного перцептрона. Она обучалась с учителем методом обратного распространения ошибки и аналитически записывается следующим образом:

$$\varphi = f_4 \left( \sum_{h=1}^{n_3} f_3 \left( \sum_{k=1}^{n_2} f_2 \left( \sum_{j=1}^{n_1} f_1 \left( \sum_{i=1}^n x_i W_{i,j}^{(1)} + b_j \right) W_{j,k}^{(2)} + b_k \right) W_{k,h}^{(3)} + b_h \right) W_h + b^{\text{вых}} \right) \quad (1)$$

где  $\varphi$  – оценка влажности материала (массовая доля сухого материала),

$x_i$  – значение  $i$ -той входной переменной (в нормируемом виде)

$$f_1, f_2, f_3 – \text{функции активации } ReLU\text{-функции } \varphi = f_4 \left( \sum_{h=1}^{n_3} f_3 \left( \sum_{k=1}^{n_2} f_2 \left( \sum_{j=1}^{n_1} f_1 \left( \sum_{i=1}^n x_i W_{i,j}^{(1)} + b_j \right) W_{j,k}^{(2)} + b_k \right) W_{k,h}^{(3)} + b_h \right) W_h + b^{\text{вых}} \right)$$

$$f_1 = \max(0, \sum_{i=1}^n x_i W_{i,j}^{(1)} + b_j),$$

$$f_2 = \max(0, \sum_{j=1}^{n_1} \max(0, \sum_{i=1}^n x_i W_{i,j}^{(1)} + b_j) W_{j,k}^{(2)} + b_k),$$

$$f_3 = \max \left( 0, \sum_{k=1}^{n_2} \max \left( 0, \sum_{j=1}^{n_1} \max \left( 0, \sum_{i=1}^n x_i W_{i,j}^{(1)} + b_j \right) W_{j,k}^{(2)} + b_k \right) W_{k,h}^{(3)} + b_h \right),$$

$$\varphi = \sum_{h=1}^{n_3} \max(0, \sum_{k=1}^{n_2} \max \left( 0, \sum_{j=1}^{n_1} \max \left( 0, \sum_{i=1}^n x_i W_{i,j}^{(1)} + b_j \right) W_{j,k}^{(2)} + b_k \right) W_{k,h}^{(3)} + b_h) W_h + b^{\text{вых}},$$

$$n – \text{число входных векторных переменных } (n=8), (i=1..8).$$

$$n_1 – \text{число нейронов, находящиеся в 1-ом скрытом слое } (n_1=36), (j=1..n_1).$$

$$n_2 – \text{число нейронов, находящиеся в 1-ом скрытом слое } (n_2 = 36), (k = 1 \dots n_2).$$

$$n_3 – \text{число нейронов, находящиеся в 1-ом скрытом слое } (n_3 = 36), (h = 1 \dots n_3).$$

$W^{(1)}_{i,j}$ ,  $W^{(2)}_{j,k}$ ,  $W^{(3)}_{k,h}$ ,  $W_k^{вых}$  – коэффициенты весов скрытых и выходного слоев. Относительная погрешность оценки влажности сыпучего материала не превышает 2 %.

Измерительная процедура, представленная нейронной сетью (1), используется при оценке влажности только в случае работоспособности всех первичных измерительных преобразователей. Иначе определяются классы отказов первичных измерительных преобразователей с учетом нормальных и различных экстремальных ситуаций функционирования.

#### IV. КЛАССИФИКАЦИЯ ОТКАЗОВ

После проведения анализа ситуаций вводятся следующие классы отказов: безотказная работа всех первичных измерительных преобразователей, мягкий отказ, средний отказ, жесткий отказ. Классы различаются уровнем значимости поступающей информации от первичных измерительных преобразователей участвующих в мониторинге процесса сушки.

Класс безотказная работа всех первичных измерительных преобразователей –  $G0$ .

Класс мягкие отказы –  $G1$ . Это ситуации отказов первичных измерительных преобразователей информация от которых не оказывает значимого влияния на прохождения процесса сушки к ней относятся первичные измерительные преобразователи, данные от которых являются шумом в измерительном канале (первичный измерительный преобразователь температуры барды на выходе барабанной сушильной установки). Класс средние отказы –  $G2$ . Это ситуации отказов первичных измерительных преобразователей: импульсного управления мощностью вытяжных вентиляторов, определения нагрузки электродвигателя сушильной установки. Класс жесткие отказы –  $G3$ .

ТАБЛИЦА I СООТВЕТСТВИЕ МОДЕЛЕЙ ОЦЕНКИ ВЛАЖНОСТИ КЛАССАМ И СИТУАЦИЯМ ОТКАЗОВ (БКСУ)

Класс отказов	БКСУ		
	Ситуации отказов датчиков	Отсутствующий параметр	Идентификатор модели
$G0$	Безотказная работа всех датчиков	-	$m_1$
$G1$	Температура барды на выходе	$x_2$	$m_2$
	Температура выпара	$x_7$	$m_3$
$G2$	Распределение высушенной и сырой барды на смесительном шнеке	$x_1$	$m_4$
	Импульсное управление мощностью вытяжных вентиляторов	$x_5$	$m_5$
	Нагрузка основного электродвигателя	$x_8$	$m_6$
$G3$	Давление острого пара	$x_3$	$m_7$
	Температура сушильного агента	$x_4$	$m_8$
	Содержание кислорода	$x_6$	$m_9$

ТАБЛИЦА II СООТВЕТСТВИЕ МОДЕЛЕЙ ОЦЕНКИ ВЛАЖНОСТИ КЛАССАМ И СИТУАЦИЯМ ОТКАЗОВ (СВЛ)

Класс отказов	СВЛ		
	Ситуации отказов	Отсутствующий параметр	Идентификатор модели
$G0$	Безотказная работа всех датчиков	-	$m_1, m_2$
$G1$	Температура и влажность в цехе	$x_2, x_3$	$m_3$
	Ворошитель	$x_5$	$m_4$
$G2$	Открытие шибсров	$x_{10}$	$m_5, m_6$
	Открытие окон	$x_9$	$m_7, m_8$
	Начальная влажность материала	$x_1$	$m_9$
$G3$	Скорость движения конвейера	$x_4$	$m_{10}$
	Включение/выключение вытяжного вентилятора	$x_6$	$m_{11}$
	Температура и влажность сушильного агента в камерах	$x_7, x_8$	$m_{12}, m_{13}$

На основании классов и ситуаций отказов строились модели оценки влажности материала. Модели для каждого из отсутствующих параметров основаны на соотношении (1), в которых варьируются параметры, но не топология нейронной сети.

Следующим этапом предлагаемой методики является разработка фреймовой модели. Фреймовая модель определяет логику функционирования отказоустойчивого нейросетевого датчика влажности и составляет основу его базы знаний. Фреймы знаний содержат слоты измерительных процедур для всех возможных ситуаций в соответствии с классами отказов первичных измерительных преобразователей, а также значения параметров обученных нейронных сетей.

В случае отказа первичного измерительного преобразователя из базы знаний автоматически выбирается измерительная процедура, соответствующая ситуации и по ней оценивается относительная влажность с допустимой погрешностью. В результате применения фреймовой модели нейросетевой датчик влажности является отказоустойчивой системой [3].

Разработанная методика инвариантна конструктивно разным сушильным установкам. Рассмотрим ее применение для создания отказоустойчивого нейросетевого датчика влажности пастообразного материала в вальце-ленточной сушильной установке. Основными параметрами процесса сушки при построении измерительных процедур отказоустойчивого нейросетевого датчика являются: начальная влажность материала; температуры и влажности окружающей среды; скорость движения ленты транспортера; наличие или отсутствие ворошителя; процент времени работы вытяжного вентилятора, средневзвешенные температуры и влажности сушильного агента в  $i$ -й камере; процент открытия воздухозаборного окна и сбросного шибера в модели оценки относительной влажности они соответственно обозначаются –  $x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9, x_{10}$  и являются входами нейронной сети. Наиболее нагруженными по влагоусъему оказываются первая и вторая секции вальце-ленточной сушильной установки, поэтому обучать и использовать нейронные сети будем только для этих зон.

Измерительная процедура оценки влажности материала реализована на базе искусственной нейронной сети. Нейронная сеть имеет архитектуру многослойного перцептрона. Она обучалась с учителем методом обратного распространения ошибки и аналитически записывается следующим образом:

$$\bar{\varphi}_i = \left[ 1 + \exp \left( -\beta_{\text{вых}}^{(i)} \left[ \sum_{j=1}^{N_i} w_j^{(i \text{ вых})} \left( 1 + \exp \left( -\beta_j^{(i)} \left[ w_{1,j}^{(i)} \bar{x}_1 + w_{2,j}^{(i)} \bar{x}_2 + w_{3,j}^{(i)} \bar{x}_3 + w_{4,j}^{(i)} \bar{x}_4 + w_{5,j}^{(i)} \bar{x}_5 + w_{6,j}^{(i)} \bar{x}_6 + w_{7,j}^{(i)} \bar{x}_7 + w_{8,j}^{(i)} \bar{x}_8 + w_{9,j}^{(i)} \bar{x}_9 + w_{10,j}^{(i)} \bar{x}_{10} \right] - \theta_j^{(i)} \right) \right] - \theta_j^{(i)} \right) \right]^{-1}, \quad (2)$$

где для сигмоидальных активационных функций нейронов обозначены через  $\beta_j^{(n)}$ ,  $\beta_j^{(n+1)}$  – углы наклона для скрытого слоя, а  $\beta_{\text{вых}}^{(n)}$ ,  $\beta_{\text{вых}}^{(n+1)}$  – для выходного слоя нейронной сети;  $\theta_j^{(n)}$ ,  $\theta_j^{(n+1)}$  – сдвиги для скрытого слоя и  $\theta_{\text{вых}}^{(n)}$ ,  $\theta_{\text{вых}}^{(n+1)}$  – для выходного слоя сети;  $w_{i,j}^{(n)}$ ,  $w_{i,j}^{(n+1)}$  и  $w_j^{(n \text{ вых})}$ ,  $w_j^{(n+1 \text{ вых})}$  – весовые коэффициенты связей для скрытого и выходного слоя сети;  $N_i$  – число нейронов в скрытом слое сети при расчете влажности материала в  $i$ -й камере сушильной установки; нормированные значения данных влажности материала  $\bar{\varphi}_i$ , начальной влажности материала  $\bar{x}_1$ , температуры и влажности окружающей среды  $\bar{x}_2$ ,  $\bar{x}_3$ ; скорости движения пластинчатого конвейера  $\bar{x}_4$ , наличия или отсутствия ворошителя  $\bar{x}_5$ , процента времени работы вытяжного вентилятора  $\bar{x}_6$ , температуры и влажности сушильного агента  $\bar{x}_7$ ,  $\bar{x}_8$ , процента открытия воздухозаборного окна и сбросного шибера  $\bar{x}_9$ ,  $\bar{x}_{10}$  [4]. Относительная погрешность оценки влажности пастообразного материала не превышает 2 %.

Измерительная процедура вида (2) используется только при работоспособности всех используемых в модели первичных измерительных преобразователей. Поэтому важной задачей является проведение анализа возможных режимов работы вальце-ленточной сушильной установки с целью выявления экстремальных ситуаций, определения множества ситуаций функционирования с отказом первичных измерительных преобразователей и определения класса к которому принадлежит отказ.

В результате анализа можно распределить первичные измерительные преобразователи по классам следующим образом. К мягким – относятся датчики передающие информацию по информационно-измерительным каналам, отказ которых может восприниматься как помехи. Это датчики: наличия ворошителя предназначенного для равномерного распределения материала на перфорированной ленте продуктового транспортера; стоящие вне сушильной установки первичные измерительные преобразователи температуры и влажности сушильного агента.

Наибольшее влияние на качественные и количественные показатели процесса сушки оказывают скорость движения ленты продуктового транспортера, включение и выключение общего вытяжного вентилятора, температура и влажность сушильного агента в камерах сушильной установки. Отказ этих первичных измерительных преобразователей приводит к существенным ошибкам измерения относительной влажности пастообразного материала и относится к жестким отказам.

## V. ВЫВОДЫ

Отказоустойчивый нейросетевой датчик влажности красного пастообразного красителя внедрен на ОАО «Пигмент», г. Тамбова, отказоустойчивый нейросетевой датчик влажности барды внедрен на ОАО «Амбер «Талвис». Применение этих датчиков позволило увеличить выпуск качественной продукции без снижения производительности процесса сушки. Таким образом, достигается поставленная техническая задача оперативная оценка в реальном времени влажности движущегося пастообразного материала в процессе его сушки для управления этим процессом с целью обеспечения качества высушиваемого материала.

## СПИСОК ЛИТЕРАТУРЫ

- [1] Shatina A.V., Borets A.S. A mathematical model of the gravitational potential of the planet taking into account tidal deformations. *Russian Technological Journal*. 2024;12(2):77–89. <https://doi.org/10.32362/2500-316X-2024-12-2-77-89>
- [2] Chernyshov V.N. (2016). Methods and information-measuring systems of non-destructive control of thermophysical properties of materials and products / Chernyshov V.N., Chernyshova T.I.// Scientific edition. Monograph / St. Petersburg, 2016
- [3] M. Kamenskaya, T. Chernyshova, D. Zemskoy, A. Ladynin, S. Artemova and P. Karelin, "Metrological Level Evaluation Method for Electric Power Complexes Information and Measuring Systems," *2024 Conference of Young Researchers in Electrical and Electronic Engineering (EICon)*, Saint Petersburg, Russian Federation, 2024, pp. 367-369, doi: 10.1109/EICon61730.2024.10468385.
- [4] S.V. Artemova, M.A. Kamenskaia, E.S. Mityakov, C.Vu Chi, A.I. Ladynin and D.N. Lapaev, "Neural Networks in Material Humidity Sensors," *2022 III International Conference on Neural Networks and Neurotechnologies (NeuroNT)*, Saint Petersburg, Russian Federation, 2022, pp. 20-23, doi: 10.1109/NeuroNT5429.2022.9805539.



# Коллективные классификаторы повторного ВХОДА

А. Ю. Дорогов

Санкт-Петербургский государственный электротехнический университет  
«ЛЭТИ» им. В.И. Ульянова (Ленина)

vaksa2006@yandex.ru

**Аннотация.** В статье предложен новый метод обучения частных классификаторов, а также способ агрегации их прогнозов в составе комитета. Обучение основано на гипотезе итерационного повторного входа биологических нейронных сетей и использует для своей реализации метод главных компонент. Для частных классификаторов задаются области компетенции в совокупности покрывающих обучающее множество примеров. Показано, что процесс итерационного обучения за несколько шагов сходится, обеспечивая точность распознавания 100 % в области компетенции частного классификатора. Агрегация прогнозов реализуется по принципу максимальной проекции образа на собственные подпространства классов частных классификаторов. Приведены примеры использования комитета компетентных классификаторов для набора данных MNIST. Предложена модель непрерывного обучения комитета классификаторов пригодная для построения самообучающихся систем распознавания.

**Ключевые слова:** *частный классификатор; комитет; коллективное распознавание; область компетенции; повторный вход; непрерывное обучение; нейронная сеть*

## I. ВВЕДЕНИЕ

Метод коллективного решения задачи классификации заключается в объединении моделей одиночных классификаторов в комитет с некоторым общим принципом агрегации прогнозов частных классификаторов. При формировании комитета необходимо оптимизировать два критерия – качество обучения частных классификаторов и оптимальность их объединения [1, 2]. Простейшим методом агрегации частных прогнозов может быть простое голосование, которое приводит к увеличению точности коллективного решения за счет снижения степени неопределенности при сложении частных прогнозов. Общее название методов построения коллективных классификаторов выражается термином «boosting» – усиление [3]. Наиболее известными являются методы bagging и adaboost.

Метод bagging предложенный Лео Брейманом в 1994 году [4], предполагает обучение частных классификаторов на случайных выборках, формируемых из обучающего множества. Агрегация классификаторов выполняется простым голосованием.

Метод adaboost предложенный Йоавом Фройндом и Робертом Шапире [5] в 1995 году предполагает обучение частных классификаторов на одной выборке с пошаговой адаптацией по ошибкам обучения как самих

классификаторов так и комитета классификаторов. Агрегация частных прогнозов выполняется взвешенным голосованием. Оптимальные веса для классификаторов рассчитываются на каждом шаге алгоритма. Одновременно видоизменяется обучающая выборка, где для каждого примера также выполняется расчет веса, который используется в процедуре обучения, например, для повторения в выборке примеров на которых обнаружены ошибки.

В 70-х и 80-х годах советские ученые Л.А. Растринин и Р.Х. Эренштейн в ряде работ разработали метод коллективного распознавания на основе компетентных классификаторов [6, 7, 8]. Идея таких методов базируется на том, что каждый базовый классификатор может работать хорошо в некоторой области пространства признаков (эта область интерпретируется как область компетентности классификатора), превосходя в этой области остальные классификаторы по точности и достоверности решений. Классификация образов выполняется в два этапа: первый этап связан с определением классификатора, область компетенции которого покрывает распознаваемый образ и на втором этапе найденный классификатор используется для отнесения образа к некоторому классу. Авторы предложили несколько способов определения области компетенции классификаторов и доказали применимость используемого подхода на примерах.

Лауреат нобелевской премии американский нейробиолог Д.М. Эдельман в 1977 году выдвинул гипотезу [9, 10] о том, что ресинтезируемая (повторно-вводимая) передача сигналов служит общим механизмом для накопления знаний в коре головного мозга. Повторный вход по Эдельману – это динамический процесс постоянной пространственно-временной корреляции, происходящий между функционально разделёнными нейронными областями, который опосредуется повторной передачей сигналов через массивно параллельные взаимные нервные волокна. С точки зрения теории динамических систем повторный вход подобен итерационному процессу, сходящемуся к некоторому устойчивому аттрактору, обобщающему накопленные знания.

## II. КЛАССИФИКАТОР ПОВТОРНОГО ВХОДА

Рассмотрим классификатор, основанный на оценке длины проекции образа на подпространства классов. Подпространства класса порождаются линейной комбинации векторов обучающей выборки, принадлежащих одному классу. Проекцию вектора

образа легко определить, если в подпространстве класса задать ортогональный базис. Сумма квадратов коэффициентов разложения вектора образа по этому базису будет определять квадрат проекции образа на подпространство класса. Образы нормируются по энергии к единице. Классификатор относит образ к классу, для которого проекция на подпространство класса оказывается максимальной.

Среди возможных ортогональных базисов, можно подобрать оптимальный базис, который обеспечивает наиболее быструю сходимость коэффициентов разложения. Известно, что таким свойством обладает ортогональный базис, полученный методом главных компонент [11] (англ. principal component analysis, PCA) из набора векторов класса. Эмпирическая или выборочная ковариационная матрица класса определяется выражением:

$$C = \frac{1}{m-1} \sum_{i=1}^m (x_i - \bar{x})^T (x_i - \bar{x}).$$

где  $m$  – число примеров в классе,  $x_i$  – вектор-строка представителя класса,  $\bar{x}$  – вектор-строка среднее значение примеров в классе. Ковариационная матрица является симметричной и положительно определённой. Для такой матрицы можно найти собственные ортогональные вектора  $v$  такие что

$$Cv = \lambda v.$$

Скаляр  $\lambda$  называется собственным числом, соответствующим собственному вектору  $v$ . Собственные числа положительные, и чем большее значение они имеют, тем большей значимостью обладает собственный вектор. Для построения модели классификатора отбирают наиболее значимые вектора.

При распознавании вектор образа центрируется по среднему значению примеров в классе, и раскладывается по выбранным собственным векторам класса. Далее по коэффициентам разложения вычисляется длина проекции образа на подпространство класса. Образу присваивается метка класса, в котором длина проекции максимальна.

Вспользуемся теперь принципом повторного входа для обучения PCA-классификатора. В обучающей выборке выберем представительный фрагмент, который далее будем называть областью компетенции классификатора. В области компетенции произвольно выделим базовое множество примеров, которое будем использовать для построения PCA-модели классификатора. На каждом шаге эволюции по результатам классификации выполняется контроль ошибок модели на области компетенции. Примеры, на которых произошли ошибки, добавляются к базовому множеству с дифференциацией по классам и модель PCA-классификатора строится заново. Процесс многократно повторяется до полного устранения ошибок на области компетенции. В процессе эволюции размер базового множества постепенно возрастает. Темп сходимости к нулевой ошибке зависит от размера области компетенции, начального базового множеств и

размерности модели (числа удерживаемых собственных векторов).

Покажем эффективность метода обучения классификатора на примере набора данных MNIST [12]. Набор данных содержит двумерные образы рукописных цифр от 0 до 9 в виде пиксельных изображений размером 28×28. Объем обучающей выборки равен 60000 образов и 10000 изображений содержит тестовая выборка. На рис. 1 показана выборка с десятью представителями для каждого класса. При проведении экспериментов все примеры выборок нормировались по энергии к единичному уровню. Для набора данных максимальной объем области компетенции, на котором эволюционный процесс ещё сходится, составляет примерно 20000 примеров. Наилучшие результаты достигаются на малых выборках.

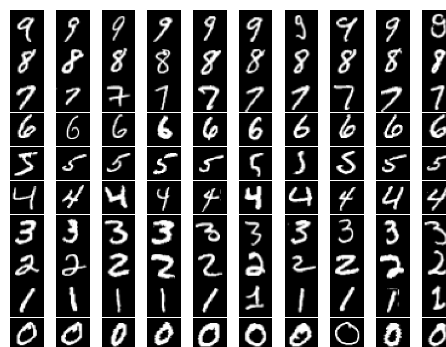


Рис. 1. Выборка из набора рукописных цифр

На рис. 2 показана зависимость точности классификатор примеров с областью компетенции из 5000 примеров и начальной базовой выборки 1000 примеров.

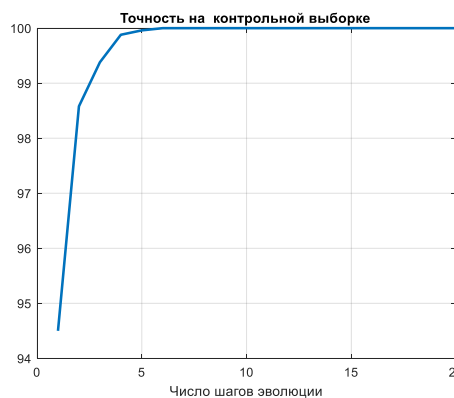


Рис. 2. Зависимость точности классификатора на контрольной выборке от числа шагов эволюции

Как видно из рисунка итерационный процесс сходится примерно за 6 шагов. Размер моделей в этом эксперименте составлял 30 собственных векторов. В процессе эволюции базовое множество возросло с 1000 до 1385 примеров. Обученный классификатор проверялся на тестовой выборке размером 10000 примеров. Тестовая выборка не пересекается с областью компетенции.

Матрица ошибок. Точность классификатора=94.5995

1	961		2		2	1	5	2	5	2	961	19	
2		1123	4	2		1	1			3	1	1123	12
3	8	4	969	14	4		4	11	14	4	969	63	
4	1	1	14	939		28	1	7	14	5	939	71	
5		2	3		951		4	8		14	951	31	
6	3	2		29	3	825	5	5	17	3	825	67	
7	8	2	1		4	15	924			3	924	33	
8		10	16	1	2	2		974	1	22	974	54	
9	6	6	8	35	5	14	6	8	871	15	871	103	
10	4	5	3	9	31	3		23	9	922	922	87	
	1	2	3	4	5	6	7	8	9	10			

Фактический класс

Предсказанный класс

Рис. 3. Таблица ошибок РСА-классификатора повторного входа на тестовой выборке

На рис. 3 показана таблица ошибок классификатора. Достигнутая точность на тестовой выборке составила 94.6 %.

### III. КОМИТЕТ КОМПЕТЕНТНЫХ КЛАССИФИКАТОРОВ

Для построения комитета классификаторов разделим всю обучающую выборку на представительные области компетенции и на каждой области обучим РСА-классификатор с повторным входом. Построенные классификаторы будем называть компетентными. Комитет классификаторов использует принцип максимума при формировании агрегированного результата. Фактически за результат комитета будет отвечать классификатор, с максимальным значением проекции образа на подпространство класса.

Матрица ошибок. Точность классификатора=99.0167

1	5913	2				3	2		1	2	5913	10
2		6737							3	2	6737	5
3	9	12	5901	10	1		1	9	12	3	5901	57
4	3	1	13	6049		22	2	3	33	5	6049	82
5	2	9	1	1	5782			8		39	5782	60
6	5	1	2	31		5337	18		20	7	5337	84
7	9	4	1		1	4	5897			2	5897	21
8	1	13	8	2	6			6203	1	31	6203	62
9	7	32	8	20	6	14	8	3	5731	21	5731	119
10	10	3	3	11	24	2	1	29	7	9859	9859	90
	1	2	3	4	5	6	7	8	9	10		

Фактический класс

Предсказанный класс

Рис. 4. Таблица ошибок комитета эволюционных классификаторов для обучающей выборки

На рис. 4 показана матрица ошибок для полного обучающего множества, состоящего из 60000 примеров. Точность классификации составляет чуть более 99 процентов при размере комитета равным 12. На рис. 5 показана матрица ошибок этого же классификатора на полной тестовой выборке, состоящей из 10000 примеров. Точность классификации составляет примерно 97 процентов.

Матрица ошибок. Точность классификатора=96.9997

1	972		2				2	1	3		972	8
2		1130	2	1				2			1130	5
3	8	6	995	1	2			2	6	10	995	37
4	3		7	967		12			6	11	967	43
5	1	2	3		962			3	1	2	962	20
6	2		1	13		855	7	2	10	2	855	37
7	5	3			2	4	940			3	940	17
8		11	8	1	2	1		989			989	39
9	7	3	2	11	3	5	3	5	929	6	929	45
10	6	6	2	6	11	2		11	5	960	960	49
	1	2	3	4	5	6	7	8	9	10		

Фактический класс

Предсказанный класс

Рис. 5. Таблица ошибок комитета эволюционных классификаторов для тестовой выборки

Для эксперимента использовались классификаторы повторного входа с областью компетенции 5000 примеров и начальным размером базового множества 1000 примеров. На текущий момент мировое достижение [13] по точности классификации тестового набора данных MNIST составляет 99.87 %.

### IV. КОЛЛЕКТИВНЫЙ КЛАССИФИКАТОР НЕПРЕРЫВНОГО ОБУЧЕНИЯ

Природа не разделяет поток данных на обучающую и тестовую выборки, биологический мозг обучается непрерывным потоком поступающих данных. С этой точки зрения имеет смысл отказаться от общепринятого деления выборки на обучающую и тестовую. Принцип разделения классификаторов по областям компетенции, с агрегированием по максимуму, позволяет неограниченно наращивать размер комитета, покрывая всю выборку и обучающую, и тестовую, а также все последующие поступившие данные. На рис. 6 приведена матрица ошибок для объединённой выборки коллективного классификатора для объединённой выборки с числом примеров 70000.

Матрица ошибок. Точность классификатора=99.2657

1	8890	1	2			2	4	1	2	1	8890	13	
2		7867	1						5	3	1	7867	10
3	7	8	8929	12	1			1	6	24	2	8929	61
4			15	7061		15	1	8	32	5	7061	80	
5	1	3	6		6769			4	4	1	36	6769	55
6	5			28		5245	15			18	2	5245	68
7	13				1	7	6848			6		6848	27
8		5	10		3			7240	3	32		7240	53
9	7	25	3	12	3	8	5	6	6738	17		6738	86
10	4	3	1	5	15	1		24	8	8897		8897	61
	1	2	3	4	5	6	7	8	9	10			

Фактический класс

Предсказанный класс

Рис. 6. Таблица ошибок непрерывного классификатора на объединённой выборке

В комитете использовались эволюционные классификаторы с областью компетенции 3000 примеров и начальным базовым множеством 1000 примеров. Число компетентных классификаторов для объединённой выборки из 70000 примеров составило 23. Точность коллективного классификатора составила 99.27 %. Модель непрерывного обучения может быть использована для построения самообучающихся систем распознавания реального времени, за счет добавления новых классификаторов в процессе функционирования системы.

## V. ЭВОЛЮЦИОНИРУЮЩИЙ КОМИТЕТ КЛАССИФИКАТОРОВ

Эволюционный принцип повторного входа можно распространить также на весь комитет классификаторов. С этой целью в модели каждого компетентного классификатора запоминаются номера примеров базового множества на момент завершения обучения (т.е. достижения 100 % точности на области компетенции). Алгоритм эволюции для комитета, на каждом шаге фиксирует номера примеров в объединённой выборке, на которых произошли ошибки. Эти примеры разделяются по областям компетенции частных классификаторов и добавляются в списки примеров базовых множеств. Далее заново выполняется обучение частных классификаторов с обновлёнными базовыми множествами до достижения 100 % точности на областях компетенции, и вновь контролируются ошибки комитета классификаторов, далее итерации комитета повторяются.

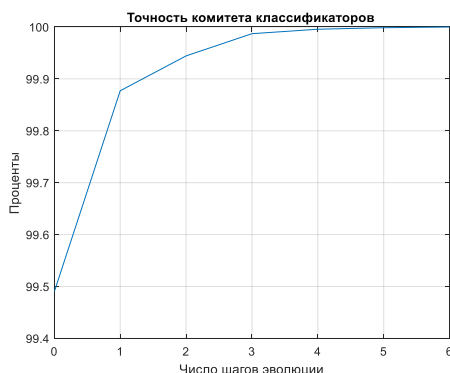


Рис. 7. Зависимость точности комитета классификаторов от числа шагов эволюции

Эволюционный процесс сходится к 100% процентной точности. На рис. 7 показана зависимость точности комитета классификаторов от числа шагов эволюции.

Предложенная модель непрерывного эволюционного обучения комитета классификаторов обеспечивает 100 процентную точность распознавания объединённой выборки.

## VI. ЗАКЛЮЧЕНИЕ

Рассмотренный метод построения коллективных классификаторов отличается от методов bagging и adaboost [3, 4, 5] использованием классификаторов с заданными областями компетенции и способом получения агрегативного решения комитета.

В отличие от метода коллективного распознавания Л.А. Растригина и Р.Х. Эренштейна [6, 7, 8], области компетенции классификаторов не локализуются расчётным путём, а назначаются исходя из условия достаточной представительности классов в базовых множествах и областях компетенции. Допустимо, также частичное пересечение областей компетенции. Благодаря принципу повторного входа частные классификаторы идеально обучаются в пределах областей их компетенций. Это позволяет сделать процедуру классификации одноступенчатой, без предварительного определения области компетенции.

РСА-классификатор при вычислении векторных проекций, предварительно осуществляет разложение вектора-образа по усечённому ортогональному базису. Эти операции являются линейными и соответствуют скалярным произведениям векторов. Для их выполнения можно использовать самоподобные пирамидальные нейронные сети [14], что обеспечит распараллеливание алгоритмов классификации. Регулярный лес пирамидальных сетей [15] позволяет в параллельном варианте реализовать полный комитет РСА-классификаторов.

Для построения ортогонального базиса подпространства класса можно использовать также сингулярное разложение матрицы [16] составленной из векторов обучающей выборки. Эти классификаторы также обучаются по методу повторного входа, причём скорость обучения оказывается несколько выше, чем у РСА-классификаторов. Точность комитета классификаторов данного типа примерно соответствует точности комитета с РСА-классификаторами.

## СПИСОК ЛИТЕРАТУРЫ

- [1] Городецкий В.И., Серебряков С.В. Методы и алгоритмы коллективного распознавания: обзор // Труды СПИИРАН. Вып. 3, т. 1. СПб.: Наука, 2006.
- [2] Терехов С.А. Гениальные комитеты умных машин. Научная сессия МИФИ 2007. IX Всероссийская научно-техническая конференция «НЕЙРОИНФОРМАТИКА-2007»: Лекции по нейронинформатике. Часть 2. М.: МИФИ, 2007. С. 11-43.
- [3] Eric Bauer, Ron Kohavi. An empirical comparison of voting classification algorithms: Bagging, boosting, and variants // Machine Learning. – 36 (1999) pp. 105-142. URL: <http://citeseer.ist.psu.edu/bauer99empirical.html>
- [4] Breiman L. Bagging Predictors. Department of Statistics University of California Berkeley, California. Technical Report No. 421. 1994.
- [5] Yoav Freund, Robert E. Schapire. A decision-theoretic generalization of online learning and an application to boosting // Second European Conference on Computational Learning Theory. – 1995. URL: <http://citeseer.ist.psu.edu/89601.html>
- [6] Растринин Л.А., Эренштейн Р.Х. Обучение коллектива решающих правил // Адаптивные системы. 1974. Вып. 4. 1974. С. 8-20.
- [7] Растринин Л.А., Эренштейн Р.Х. Принятие решений коллективом решающих правил в задачах распознавания образов // Автоматика и телемеханика. 1975. №. 9. С. 133-144.
- [8] Растринин Л.А., Эренштейн Р.Х. Метод коллективного распознавания. Москва: Энергоиздат, 1981. (Серия «Библиотека по автоматике» Выпуск 0615.)
- [9] Edelman G.M. (1978). "Group selection and phasic reentrant signaling: a theory of higher brain function," in *The Mindful Brain: Cortical Organization and the Group-Selective Theory of Higher Brain Function*, eds G.
- [10] Дж. Эделман, В. Маунткасл. Разумный мозг. Мир, 1981.
- [11] Лагутин М.В. Наглядная математическая статистика. Учебное пособие. М.: БИНОМ. Лаборатория знаний. 2007. 472 с.
- [12] <http://yann.lecun.com/exdb/mnist/>
- [13] Adam Byerly, Tatiana Kalganova, Ian Dear. No Routing Needed Between Capsules. arXiv:2001.09136.
- [14] Дорогов А.Ю. Глубокие пирамидальные нейронные сети быстрого обучения для корреляционной обработки сигналов и изображений // IV Международная конференция по нейронным сетям и нейротехнологиям (NeuroNT'2023). СПб.: СПбГЭТУ «ЛЭТИ». 2023. с. 49-52.
- [15] Дорогов А.Ю. Быстрые нейронные сети глубокого обучения // III Межд. науч. конф. по проблемам управления в технических системах (CTS'2019): сб. докладов. Санкт-Петербург. 30 октября - 1 ноября 2019 г. СПб.: СПбГЭТУ «ЛЭТИ». С. 275-280.
- [16] Воеводин В.В., Кузнецов Ю.А. Матрицы и вычисления. М.: Наука. 1984. 320 с.

# Применение сверточных нейронных сетей для классификации ЭЭГ/ВП сигналов единичных проб объединённой выборки данных испытуемых в ответ на предъявление математических примеров

М. О. Петров

*Санкт-Петербургский  
государственный  
электротехнический университет  
«ЛЭТИ» им. В.И. Ульянова (Ленина)*

petrovmo94@yandex.ru

Ж. В. Нагорнова

*Институт эволюционной  
физиологии и биохимии  
имени И. М. Сеченова РАН*

nagornova\_zh@mail.ru

Н. В. Андреева

*Санкт-Петербургский  
государственный  
электротехнический университет  
«ЛЭТИ» им. В.И. Ульянова (Ленина)*

nvandr@gmail.com

Н. В. Шемякина

*Институт эволюционной  
физиологии и биохимии  
имени И. М. Сеченова РАН*

shemyakina\_n@mail.ru

**Аннотация.** Использование сверточных нейронных сетей (CNN) является одним из эффективных подходов для классификации электроэнцефалографических (ЭЭГ) данных. В работе проводится сравнительный анализ применения CNN для классификации 3 классов данных – сырого сигнала вызванных потенциалов (ВП) мозга при предъявлении математических примеров, правильных и неправильных ответов к ним. Наборы единичных проб всех испытуемых (20 человек) формировали общую выборку данных для классификации, в которой в каждый класс (правильный/неправильный ответ, математический пример) включалось ~ 100 проб от каждого испытуемого. Для обучения и тестирования использовали перекрывающиеся 80 % и 20 % выборки, соответственно. В работе тестируется архитектура ResNet18 и точность классификации сырого сигнала ЭЭГ/ВП для разных временных окон длительностью 200 – 500 мс (с шагом 100 мс) от начала предъявления стимула (математического примера, правильного или неправильного ответа). Лучшая точность классификации 3-х указанных классов составила  $72 \pm 3\%$ . В большей степени степени ошибочно разделяется сырой сигнал ЭЭГ/ВП на правильный и неправильный ответы, ошибка различения зрительных проб ВП с предъявлением математических примеров – примерно в два раза ниже. Классификация сырого ЭЭГ/ВП сигнала и его характеристик в ответ на предъявление правильного/ошибочного ответа может использоваться в организации обратной связи при использовании интерфейсов мозг-компьютер (ИМК).

**Ключевые слова:** ЭЭГ, вызванные потенциалы, сверточные нейронные сети, ResNet18, интерфейс мозг-компьютер

## I. ВВЕДЕНИЕ

Задача распознавания по физиологическим данным состояний мозга человека является актуальной в контуре разработки интерфейсов мозг-компьютер. Электроэнцефалография – неинвазивный метод измерения биоэлектрической активности мозга человека, который позволяет получить информацию о его функционировании в режиме реального времени. В последние годы использование нейросетей в анализе и интерпретации данных ЭЭГ стало достаточно востребованным. Нейронные сети (нейросети) широко применяются в задачах классификации данных ЭЭГ, в том числе благодаря своей способности распознавать сложные закономерности и создавать эффективные модели для классификации. Статьи, посвященные использованию нейросетей в задачах классификации ЭЭГ, обсуждают различные методы классификации ЭЭГ-паттернов, например, в статье [1] авторы сравнивают несколько подходов, такие как общая пространственная модель (CSP) [2–4], Риманова геометрия и глубокие нейронные сети для задачи классификации ЭЭГ-паттернов воображаемых движений. Точность таких подходов для задачи двухклассовой классификации составила ~ 90 % для римановой геометрии, ~ 77 % для CSP и ~ 81 % для глубоких нейросетей. В статье [5] авторы показали, что нейросети типа ResNet показывают лучший результат для классификации «непрерывных», длительных состояний творческой деятельности, в то время как для более быстрых процессов, таких как парадигма вызванных потенциалов, лучше использовать классификацию «временных рядов» с пространственной фильтрацией.

Одной из ключевых тем в статьях об использовании нейросетей в классификации ЭЭГ является выбор архитектуры нейронной сети. Различные типы нейронных сетей, такие как сверточные нейронные сети [6, 7], рекуррентные нейронные сети [8–10] и комбинированные модели [11, 12], применяются для анализа данных ЭЭГ. Каждый тип сети имеет свои преимущества и недостатки в зависимости от задачи классификации. Также в статьях рассматриваются методы предобработки данных для эффективного использования нейросетей в классификации ЭЭГ. Эти методы включают в себя фильтрацию шумов, нормализацию данных, разбиение данных на сегменты и многие другие. Правильная предобработка данных играет важную роль в улучшении качества классификации [1, 5].

В статье [13] авторы исследуют применение глубокого обучения для автоматической обработки ЭЭГ-данных. Они разрабатывают и обучают сверточную нейронную сеть для классификации ЭЭГ по двум основным категориям: нормальная активность и эпилепсия. Авторы показывают, что одномерная сверточная нейронная сеть способна классифицировать данные с точностью до 90% при распознавании нормальных и аномальных сигналов ЭЭГ, по сравнению с традиционными методами машинного обучения, такими как метод опорных векторов и логистическая регрессия.

В работе [14] использование сверточных нейронных сетей также позволило различать данные пациентов с эпилепсией и здоровых добровольцев с точностью 85 %. Статья [15] представляет собой еще одно исследование, где авторы применяют глубокое обучение для классификации данных ЭЭГ. В данной работе рассматриваются различные алгоритмы глубокого обучения для многоклассовой классификации временных рядов различных электрических активностей в мозге. Сравнительное исследование, простой рекуррентной нейронной сети (RNN), сети долгой краткосрочной памяти (LSTM) и модели управляемых рекуррентных блоков (GRU), проводится для сигналов ЭЭГ, полученных от людей с различными патологическими и физиологическими состояниями мозга. Авторы исследуют время работы алгоритмов и их точность в зависимости от количества слоев в нейронной сети. Основные выводы, которые делают авторы время и сложность увеличиваются, по мере роста используемых слоев, при этом точность переставала существенно увеличиваться после 3 слоя. Наилучшая точность для RNN составила 52 %, для LSTM 70 %, а для GRU 73 %. В работе используется очень популярный набор данных для обнаружения эпилептических припадков [16].

В статье [17] рассматривается алгоритм классификации моторных воображений на основе электроэнцефалограммы с использованием гибридной нейронной сети с пространственно-временной сверткой и механизмом многоголовочного внимания. Авторы показывают, что обеспечение производительности системы требует высокого отношения сигнал-шум в записанных ЭЭГ-сигналах. Методы глубокого обучения, такие как сверточные нейронные сети, сети глубокого доверия и рекуррентные нейронные сети, являются

основными алгоритмами классификации сигналов ЭЭГ. Эти методы позволяют улучшить точность классификации сигналов и обеспечить более эффективное управление устройствами с помощью мозговых сигналов.

Целью данной работы является анализ особенностей обучения сверточной нейронной сети типа ResNet, применяемой для классификации данных ЭЭГ/ВП, полученных при выполнении 20 испытуемыми-подростками когнитивной задачи дискриминации предъявленных правильных и неправильных ответов, при решении математических примеров с переходом через десяток. Рассматривались 3 класса данных сигналов вызванных потенциалов мозга при предъявлении испытуемым математических примеров, правильных и неправильных ответов к ним.

## II. МЕТОДЫ И ПОДХОДЫ

Для решения задачи классификации использовалась сверточная сеть ResNet18 (Residual Network). Эта архитектура сверточной нейронной сети, разработанная в 2015 году для решения задач компьютерного зрения [18]. Она была создана для классификации изображений и представляет собой относительно небольшую и быструю модель с 18 слоями, что делает ее идеальным выбором для обучения на устройствах с ограниченными ресурсами.

Основная идея ResNet заключается в использовании так называемых остаточных связей, которые позволяют обучить более глубокие модели без ухудшения производительности. Это достигается за счет того, что часть входных данных передается напрямую через определенные слои, минуя промежуточные операции.

ResNet18 состоит из нескольких блоков, каждый из которых содержит два сверточных слоя, слой нормализации и функцию активации ReLU. Блоки соединены между собой с помощью функции residual connection, которая позволяет передавать информацию между слоями без изменений. В конце модели используется полносвязный слой для классификации изображений.

Одним из преимуществ ResNet18 является ее эффективность: она требует меньше вычислительных ресурсов и времени для обучения, чем другие модели с аналогичной точностью. Кроме того, ResNet18 имеет высокую точность классификации и может использоваться для решения различных задач, таких как распознавание объектов, сегментация изображений и определение ключевых точек.

Для решения поставленной задачи использовалась конфигурация сети ResNet18 с оптимизированными гиперпараметрами, что позволило исключить необходимость использования валидационной выборки при решении задачи классификации ЭЭГ-сигналов.

В работе использовались данные, представляющие собой запись ЭЭГ от 20 испытуемых. Регистрация ЭЭГ проводилась с помощью 24-х канального цифрового электроэнцефалографа SmartBCI™ (ООО «Мицар», Санкт-Петербург) посредством программного пакета WinEEG (Пономарев В.А., Кропотов Ю.Д., №

государственной регистрации 2001610516 от 08.05.2001) от 19 хлорсеребряных электродов, расположенных по международной системе 10–20. Референтный электрод располагался на мочках обеих ушей, заземляющий электрод – в переднецентральном отведении на поверхности головы. ЭЭГ регистрировалась в полосе 0.53–30 Гц. Сопротивление электродов не превышало 5 кОм, частота дискретизации сигналов составила 250 Гц.

Данные ЭЭГ были организованы в пробы, длительность каждой пробы составляла 3 секунды. С учетом частоты дискретизации (250 Гц), каждая проба представляла собой массив данных, размером 750 x 19, включающий 750 дискретных значений сигнала (3 сек x 250 точек) от 19 электродов.

В ходе исследования на 300 мс записи ЭЭГ испытуемым предъявлялся математический пример в течение 400 мс, и на 1400 мс – ответ (правильный или неправильный).

Для решения задачи классификации данные разбивались на три класса с вызванным ответом на предъявление различных зрительных стимулов: правильного ответа (первый класс), неправильного (второй класс), математического примера (третий класс).

От каждого испытуемого было получено 300 образцов данных, 100 образцов данных соответствовали реакции испытуемого на предъявление правильного ответа; 100 образцов – неправильного ответа и 100 образцов – на предъявление математического примера.

Формирование выборок данных рассматриваемых классов осуществлялось с использованием двух параметров. В качестве первого параметра  $T$  выступало время с момента предъявления правильного или неправильного ответов испытуемым для первого и второго классов, соответственно, и с момента начала пробы (предъявления математического примера) для третьего класса. Второй параметр  $\Delta t$  определял промежуток времени (окно), на протяжении которого выбирались данные ЭЭГ сигналов пробы, начиная с момента времени  $T$ .

Таким образом, каждому классу соответствовал набор из 100 выборок данных (из проб) для 20 испытуемых, содержащих записи с 19 электродов, начиная с момента времени  $T$  и длительностью, определяемой окном  $\Delta t$ .

При предварительном анализе были выкинуты из рассмотрения пробы, содержащие явные артефакты ЭЭГ-сигнала, что привело к уменьшению выборок для каждого класса. Итоговое количество выборок данных составило 5820 (1940 x 3 класса). В качестве примера на рис. 1 приведены частичные выборки (для 8 из 19 электродов) для трех рассматриваемых классов.

Для обучения ResNet18 был сформирован обучающий набор данных, которых содержал ~75 % от общего количества выборок (4373 выборки). Оставшиеся 25 % набора данных использовались для тестирования результатов обучения (1447 выборок). Наборы данных для обучения и тестирования не пересекались между собой, разделение внутри отдельных классов было случайным.

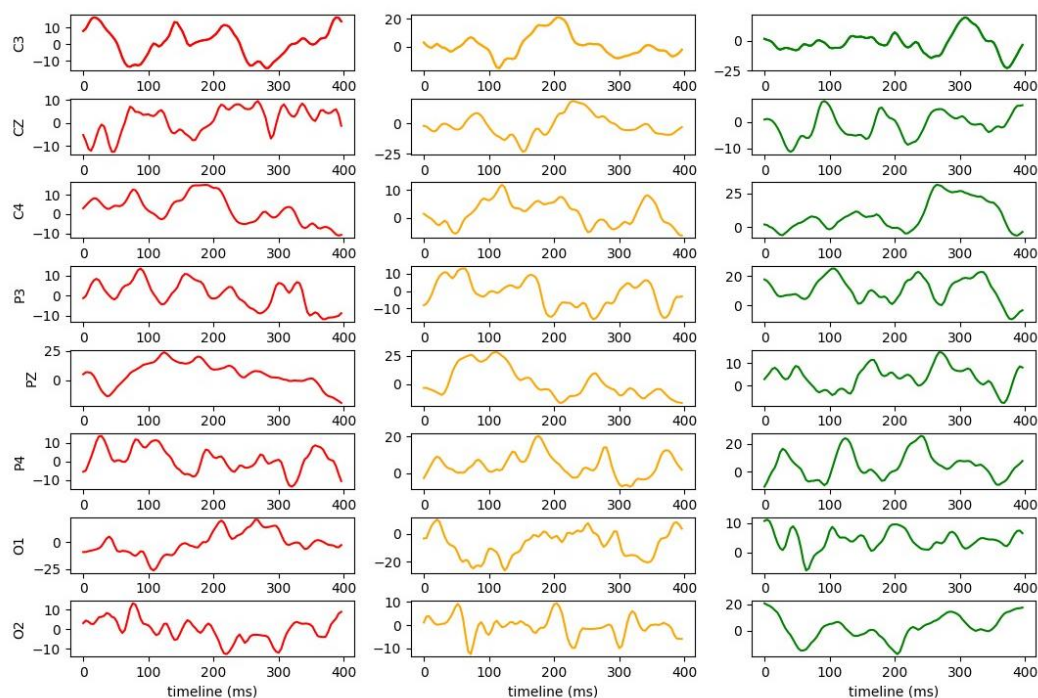


Рис. 1. Частичные выборки ЭЭГ-сигналов (для 8 из 19 электродов) для трех классов, содержащих реакцию на предъявление: (красный) – правильного ответа,  $T = 1400$  мс,  $\Delta t = 400$  мс; (оранжевый) – неправильного ответа  $T = 1400$  мс,  $\Delta t = 400$  мс; и (зеленый) – математического примера  $T = 300$  мс,  $\Delta t = 400$  мс

Обучение нейросети происходило в ходе 30 эпох, одна эпоха соответствовала одному предъявлению полного набора тренировочных данных нейронной сети. В процессе обучения с учителем, точность решения задачи классификации определялась после каждой эпохи обучения, как отношение количества правильных ответов к общему количеству ответов и выражалась в процентах. Ошибка сети характеризовалась с использованием функции потерь, определяющей меру расхождения между ответом нейросети (отнесением выборки к определенному классу данных) и правильным ответом.

В ходе работы было проведено исследование влияния параметров формирования выборок ( $T$  и  $\Delta t$ ) на точность решения задачи классификации ResNet18 ЭЭГ-данных. Для первого и второго классов рассматривались два возможных значения параметра  $T=1400$  мс и  $T=1600$  мс, для третьего класса данный параметр не изменялся и составлял 300 мс. Для выбранных значений параметра  $T$ , изменялось временное окно  $\Delta t$  выборки данных, с 200 до 500 мс, с шагом 100 мс.

### III. РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

В процессе обучения ResNet18 на сформированных выборках ЭЭГ-данных для трех рассматриваемых классов, при оптимизированных гиперпараметрах, было установлено, что влияние исходного распределения конфигурации весов сети и случайного формирования набора обучающих данных на точность распознавания обучающей нейросети находится в пределах 3–5 %. В таблице 1 представлены результаты, полученные при обучении ResNet18 на выборках, сформированных при разных параметрах  $T$  и  $\Delta t$ . Анализируя полученные данные, можно сделать вывод о существенном влиянии параметра  $T$ , определяющего время с момента предъявления стимула, на точность распознавания. Очевидно, что сдвиг параметра  $T$  в большую сторону повышает точность распознавания. Данное обстоятельство вероятно объясняется тем, что вызванный ответ на предъявление зрительного стимула содержит большей частью в ЭЭГ-данных, начиная с 1600 мс от момента предъявления ответа, что позволяет уменьшить вклад ЭЭГ-сигнала, не связанного с реакцией стимул.

Изучение влияния размера временного окна выборки сигнала для всех трех классов позволяет сделать заключение о существовании оптимального значения параметра  $\Delta t$ . При фиксированном значении  $T$ , с увеличением окна выборки наблюдается незначительное повешение точности распознавания, однако, только для неоптимального  $T=1400$  мс. Сдвиг  $T$  в сторону больших значений, достоверно свидетельствует о существовании оптимального значения окна выборки данных, равного 400 мс. Более того, в диапазоне значений параметра  $\Delta t$ , ограниченном 300–400 мс, наблюдается наиболее выраженная динамика к повешению точности распознавания класса обученной нейросетью.

ТАБЛИЦА 1. Влияния ПАРАМЕТРОВ ФОРМИРОВАНИЯ ВЫБОРКИ ЭЭГ-ДААННЫХ НА ТОЧНОСТЬ RESNET18

$T$ , мс	$\Delta t$ , мс	Точность, %	Обозначение на рисунках 2 и 3
1400	200	$52 \pm 5$ %	-
	300	$54 \pm 3$ %	-
	400	$58 \pm 3$ %	A
	<b>500</b>	<b><math>60 \pm 3</math> %</b>	B
1600	200	$63 \pm 5$ %	C
	300	$65 \pm 5$ %	D
	<b>400</b>	<b><math>72 \pm 3</math> %</b>	E
	500	$70 \pm 5$ %	-

Анализ процесса обучения ResNet18 позволяет сделать вывод о том, что, начиная с 15 эпохи обучения, функция потерь достигает своего минимального значения (рис. 2). То обстоятельство, что в качестве минимального значения выступает 0, свидетельствует об избыточности используемой конфигурации нейросети и, косвенно, о возможном наличии проблемы переобучения.

Зависимость точности ResNet18 от номера эпохи обучения (сплошные линии на рис. 3) для одной из нейронных сетей, обученных при разных стартовых распределениях весовых коэффициентов для разных правил формирования выборки данных (A, B, C, D и E), коррелирует с динамикой функции потерь в процессе обучения. Точность нейросети на тестовых данных (пунктирные линии на рис. 3) незначительно повышается, начиная с 10-й эпохи, выходя на плато после 20 эпохи.

Матрицы корреляции для разных параметров формирования выборок ЭЭГ-данных для трех рассматриваемых классов, в соответствии с табл. 1, приведены на рис. 4. Так, согласно полученным результатам, обученная нейросеть существенно с большей точностью распознает третий класс, соответствующий предъявлению математического примера, в то время как точность распознавания первого и второго классов практически в два раза хуже.

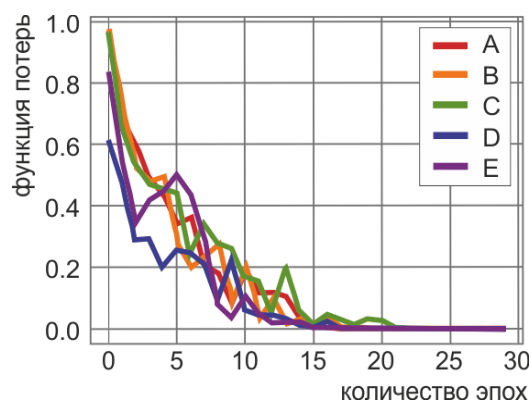


Рис. 2. Зависимость функции потерь от количества эпох обучения ResNet18 на различных выборках ЭЭГ-данных, правила формирования которых для принятых обозначений, приводятся в табл. 1



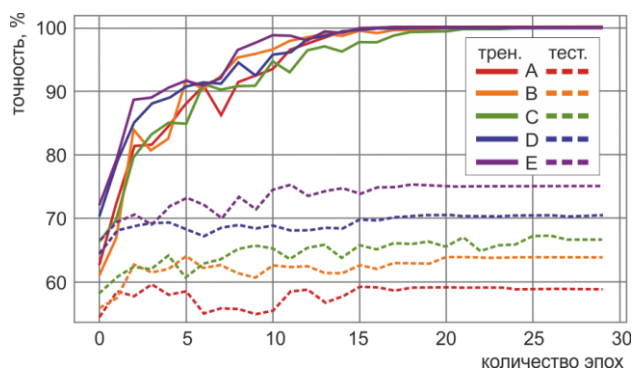


Рис. 3. Зависимость точности ResNet18 от номера эпохи для обучающего (трени.) и тестового (тест.) набора ЭЭГ-данных. Правила формирования выборок данных для наборов приводятся в соответствии с данными табл. 1.

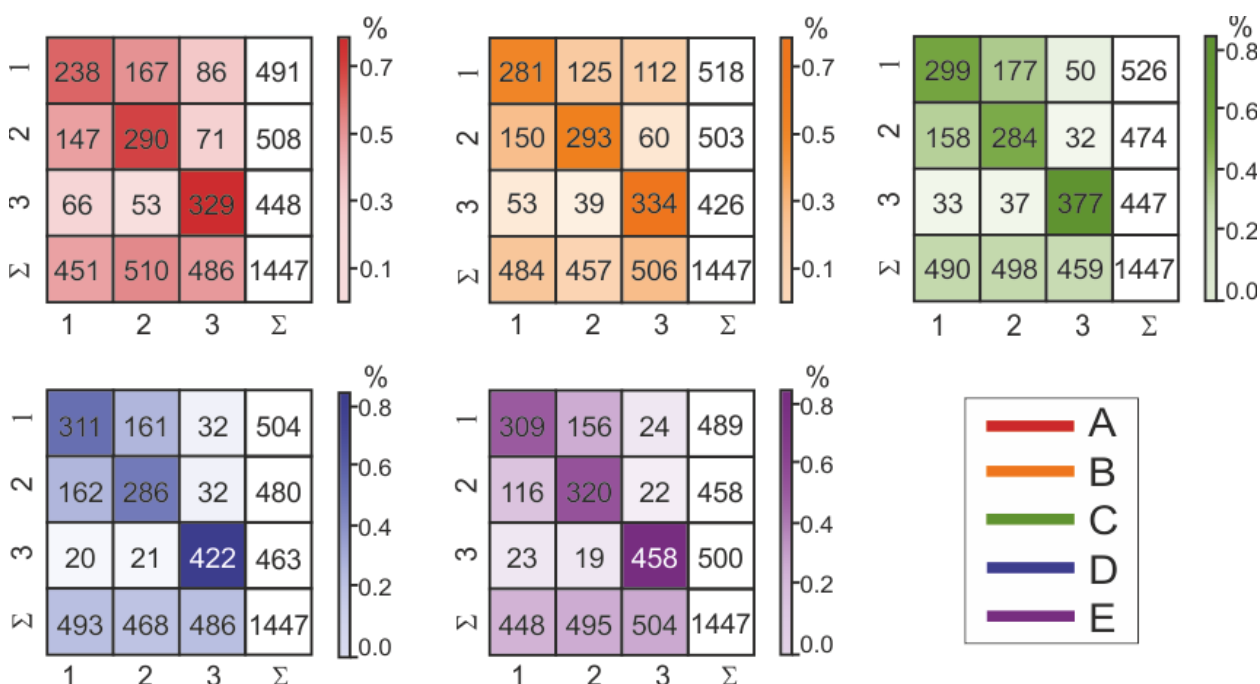


Рис. 4. Матрицы корреляции классов, обученной ResNet18 для выборок ЭЭГ-данных, сформированных в соответствии с правилами табл. 1. Значения на главной диагонали показывают количество правильные ответы сети для рассматриваемых правил формирования выборок

Рассматриваемые ЭЭГ-данные были проверены на кластеризуемость с использованием алгоритма GaussianMixture (GM). GM – это класс алгоритмов в машинном обучении и статистике. Модель учитывает структуру ковариации данных и вероятность того, что каждая точка будет получена из каждого гауссовского распределения. Проще говоря, это способ аппроксимировать данные, которые лучше всего соответствуют нескольким гауссовским распределениям (нормальным распределениям). GM часто используется в кластеризации, где цель состоит в том, чтобы разделить данные на несколько групп или кластеров. Этот алгоритм особенно полезен, когда распределение данных сложное и включает несколько отдельных компонентов. GM обучается путем максимизации функции правдоподобия, которая представляет вероятность того, что наблюдения были сгенерированы из смеси распределений. В результате были получены схожие ответы, а именно хорошо выделялся 3 класс, а первые два «путаются» между собой.

Следует отметить, что при использовании архитектуры EEgNet для решения рассматриваемой задачи классификации, максимальная точность не превышала ~ 65 %.

Таким образом, в ходе проведенного исследования установлено существенное влияние параметров формирования выборок данных на точность ResNet18 в решении задачи классификации ЭЭГ/ВП-сигналов, полученных при выполнении испытуемыми когнитивной задачи дискриминации предъявленных правильных и неправильных ответов, при решении математических примеров. В большей степени ошибочно разделяется сырой сигнал ЭЭГ/ВП на правильный и неправильный ответы, ошибка различения зрительных проб ВП с предъявлением математических примеров – примерно в два раза ниже. Определяющее влияние на точность нейросети при формировании выборки данных оказывает время с момента предъявления правильного или неправильного ответов

испытуемым. Оптимальная величина окна выборки находится в диапазоне 300–400 мс. Максимальная полученная точность классификации ЭЭГ/ВП-сигналов с использованием ResNet18 при оптимально сформированных выборках данных достигает 75 %.

#### СПИСОК ЛИТЕРАТУРЫ

- [1] N. Kapralov, Z. Nagornova, and N. Shemyakina, "Classification Methods for EEG Patterns of Imaginary Movements," *Informatics and Automation*, 2021, vol. 20, no. 1, pp. 94–132. (in Russian)
- [2] K. Fukunaga. *Introduction to Statistical Pattern Recognition*, 2nd ed. New York: Academic Press, 1990. 592 p.
- [3] H. Ramoser, J. Müller-Gerking, and G. Pfurtscheller, "Optimal spatial filtering of single trial EEG during imagined hand movement," *IEEE Trans. Rehab. Eng.*, 2000, vol. 8, no. 4, pp. 441–446.
- [4] J. Müller-Gerking, G. Pfurtscheller, and H. Flyvbjerg, "Designing optimal spatial filters for single-trial EEG classification in a movement task," *Clin. Neurophysiol.*, 1999, vol. 110, no. 5, pp. 787–798.
- [5] N.V. Shemyakina, G.S. Velicoborets, and Z.V. Nagornova, "Machine-learning applications for differentiation across states/stages of creative thinking based on time-series and time-frequency features of electroencephalography signals and event-related potentials," *Genes & Cells*, 2023, vol. 18, no. 4, pp. 421–432.
- [6] L. Alzubaidi, J. Zhang, A.J. Humaidi, A. Al-Dujaili, Y. Duan et al., "Review of deep learning: concepts, CNN architectures, challenges, applications, future directions," *J. Big Data*, 2021, vol. 8, no. 1, 53.
- [7] S. Selitsky, "Hybrid convolutional-multilayer perceptron artificial neural network for person recognition by high gamma EEG features," *Medical News of North Caucasus*, 2022, vol. 17, no. 2, pp. 192–196.
- [8] W. Yin, K. Kann, M. Yu, and H. Schütze, "Comparative study of CNN and RNN for natural language processing," arXiv preprint arXiv:1702.01923, 2017.
- [9] T.Yu. Menshikh, "Simulation modelling of pseudorandom sequences for digital communication channels," *Modern Problems of Mathematics and Computer Science: Proceedings of the X Republican Scientific Conference of Young Scientists and Students*, 2017, pp. 28–29.
- [10] I.A. Satam, "EEG signal ANFIS classification for motor imagery for different joints of the same limb," *Vojnotehnički glasnik/Military Technical Courier*, 2024, vol. 72, no. 1, pp. 330–350.
- [11] V. Fadel. *Neural network algorithms for recognizing emotions based on EEG signals*. Master's Thesis. Tomsk, 2023. (in Russian)
- [12] S.A. Gulyaev, L.M. Khanukhova, and A.A. Garmash, "Features of EEG microstate analysis in post-stroke aphasia," *Extreme Medicine*, 2023, vol. 25, no. 3, pp. 71–79. (in Russian)
- [13] Ö. Yıldırım, U.B. Baloglu, and U.R. Acharya, "A deep convolutional neural network model for automated identification of abnormal EEG signals," *Neural Comput. Appl.*, 2020, vol. 32, no. 20, pp. 15857–15868.
- [14] Y. Gao, B. Gao, Q. Chen, J. Liu, and Y. Zhang, "Deep convolutional neural network-based epileptic electroencephalogram (EEG) signal classification," *Front. Neurol.*, 2020, vol. 11, 375.
- [15] K.K. Dutta, "Multi-class time series classification of EEG signals with recurrent neural networks," 2019 9th International Conference on Cloud Computing, Data Science & Engineering (Confluence), 2019, pp. 337–341.
- [16] Rieke C Mormann F-David P Elger CE Andrzejak RG, Lehnertz K. Indications of nonlinear deterministic and finite dimensional structures in time series of brain electrical activity: Dependence on recording region and brain state. *PHYSICAL REVIEW E*, 64:061907–1–061907–8, 2001.
- [17] X. Shi, B. Li, W. Wang, Y. Qin, H. Wang, and X. Wang, "Classification algorithm for electroencephalogram-based motor imagery using hybrid neural network with spatio-temporal convolution and multi-head attention mechanism," *Neuroscience*, 2023, vol. 527, pp. 64–73.
- [18] R.V. Singh, "ImageNet Winning CNN Architectures—A Review," Rajat Vikram Singh—Institute of Software Research at Carnegie Mellon University, 2015.

# Прогнозирование неисправностей с использованием данных искусственного интеллекта и технологии перовскитовых материалов

Д. Ю. Терехов, А. Р. Попова, Е. В. Патока,  
В. Д. Бурлака, М. С. Козлов

Санкт-Петербургский государственный электротехнический университет  
«ЛЭТИ» им. В.И. Ульянова (Ленина)

ditereh@gmail.com, renoridoru@gmail.com, yegor-patoka@mail.ru,  
byrlaka60@gmail.com, mex.koz01@gmail.com

**Аннотация.** Данная статья посвящена тщательному анализу исследовательской модели, направленной на обнаружение неисправностей в системах уличного освещения, с использованием синтетического набора данных, описывающего работу уличных фонарей. Предлагается углубленное исследование применения искусственного интеллекта для прогнозирования различных типов неисправностей и тщательная оценка эффективности разработанной модели. Кроме того, в статье рассматриваются перспективы применения перовскитовых материалов в контексте оптимизации датчиков и других компонентов систем уличного освещения. Особое внимание уделено возможностям перовскитов по повышению точности и скорости обнаружения неисправностей. Цель исследования - предложить инновационный подход к обнаружению и управлению неисправностями в системах уличного освещения. Рассматривается применение передовых технологий и материалов для повышения эффективности обслуживания и снижения затрат на ремонт и содержание таких систем. Работа направлена на формирование практических рекомендаций и предложений для отрасли с целью создания более устойчивых и эффективных систем уличного освещения в будущем.

**Ключевые слова:** прогнозирование неисправностей, уличное освещение, искусственный интеллект, перовскитовая технология, набор данных, анализ данных, датчики, моделирование, энергоэффективность, неисправности уличного освещения

## I. ВВЕДЕНИЕ

Уличное освещение – важный элемент городской инфраструктуры, обеспечивающий безопасность и комфорт в общественных местах. Однако существующие методы обслуживания, основанные на традиционных подходах к мониторингу и обнаружению неисправностей, часто не обеспечивают необходимой эффективности, что приводит к задержкам в ремонте и снижению качества обслуживания.

Проблема неэффективности уличного освещения часто поднимается в научной литературе. Например,

исследование [1], проведенное в 2019 году, показало, что во многих городах используются устаревшие технологии и неэффективные методы управления освещением, что приводит к ненужному потреблению энергии и высоким эксплуатационным расходам.

На данный момент проблема так и не решена. А именно, эффективность и стабильность работы освещения колеблется в зависимости от благоустройства города, до его фактического бюджета и наличия специалистов по сфере монтажа, и ремонта систем освещения. В реальных случаях ремонт освещения производится по факту его неисправности, будь то отдельный элемент освещения всей системы или вся система, которая включает несколько элементов освещения, отключение которых может представлять опасность для местного населения, а время простоя только усугубляет эту ситуацию.

Однако мониторинг каждого отдельного элемента освещения до сих пор являются нерентабельными. Дополнительные элементы, включающие в себя установку фотодатчиков, требуют ощутимых дорогостоящих конструкций, без которых такая техническая система будет неустойчива к погоде, вибрациям или иным повреждающим факторам. К тому же сама конструкция будет перекрывать сам элемент освещения что понизит его эффективность. Именно поэтому предлагается использовать современный материал такой, как перовскит, который может быть легко нанесён в рассеивающую панель, что позволит проводить мониторинг всей системы освещения, касаясь отдельных частей или его каскадов.

## II. ПРОГНОЗИРОВАНИЕ НЕИСПРАВНОСТЕЙ УЛИЧНОГО ОСВЕЩЕНИЯ С ПОМОЩЬЮ ИИ

Традиционные методы обнаружения неисправностей, такие как визуальное наблюдение и мониторинг энергии, имеют ограничения по точности и своевременности обнаружения проблем. Например, визуальное обнаружение может быть ограничено недоступными для осмотра зонами или плохими условиями видимости, а мониторинг энергопотребления может не дать полной картины общего состояния системы.

Важно отметить, что такие системы также не подходят для динамического мониторинга, когда изменчивость, разнообразие классов ламп, различные уровни энергопотребления, влияние погодных условий и другие факторы могут существенно меняться с течением времени. Это подчеркивает необходимость разработки более гибких и адаптивных методов обнаружения неисправностей, способных быстро реагировать на динамические изменения в системе.

Для этих целей предлагается использовать технологии искусственного интеллекта для создания более эффективной модели прогнозирования неисправностей. В данном случае мы применим метод логистической регрессии [3], который используется в машинном обучении, особенно в задачах многоклассовой классификации, где целью является предсказание, к какому из четырех классов относится объект. В нашем случае:

1. Новая идеально исправная лампа.
2. Лампа со сроком службы без отказов.
3. Лампа с высокой вероятностью отказов
4. Неисправная лампа

Логарифмическая регрессия после получения данных о состоянии освещения, погодных условиях, работе системы, номере лампы, временной метке записи данных о потребляемой мощности (Вт), силе света, колебаниях тока, температуре окружающей среды, погодных условиях, факторах окружающей среды, способствующих колебаниям электрического тока, состоянии лампы. Логистическая регрессия использует функцию, известную как сигмоида, для преобразования непрерывных входных данных в вероятности принадлежности к классу. Сигмоида – это функция, которая берет любое вещественное число и преобразует его в значение между 0 и 1 [3].

Данная модель позволит представить не просто принадлежность данного состояния к классу, а вероятность принадлежности к классу от 0 до 1. Это позволит не просто детектировать состояние одного конкретного элемента системы освещения, но и представить его в доступном и подробном числовом значении, что позволяет предупредить поломку задолго до её фактического выхода из строя. Это особенно актуально в провинциальных местностях, где ремонт элементов освещения, в основном, плановый, а ремонт происходит непосредственно после поломки, что может привести к неблагоприятным условиям и даже опасностям.

В случае логистической регрессии входное значение обычно представляет собой линейную комбинацию признаков объекта и параметров модели. Параметры модели определяются в процессе обучения модели путем минимизации функции потерь, например, функции потерь кросс-энтропии. После того как модель логистической регрессии обучена и предсказала вероятность принадлежности объекта к классу 1, эта вероятность сравнивается с заранее установленными порогами классификации.

Кросс-энтропия является подходящим выбором функции потерь для многоклассовой классификации в логистической регрессии по нескольким причинам.

Во-первых, она отражает меру расхождения между предсказанными и фактическими распределениями классов. Это особенно важно для многоклассовой классификации, где необходимо учитывать вероятности принадлежности к каждому классу. Кросс-энтропия наказывает модель за отклонения в вероятностях принадлежности к классу, что способствует точности и уверенности в прогнозах. Кроме того, перекрестная энтропия демонстрирует хорошие численные свойства, обеспечивая быструю сходимость при обучении модели. Ее гладкий характер улучшает процесс оптимизации и обеспечивает эффективное обучение даже на больших объемах данных. Кроме того, кросс-энтропия обладает свойством масштабной инвариантности, что делает ее устойчивой к масштабированию признаков. Это означает, что результаты классификации остаются неизменными даже при изменении масштаба входных данных, что является важным преимуществом в реальных приложениях.

Формула функции кросс-энтропии описана в (1).

$$L(y, \hat{y}) = -\frac{1}{N} \sum_{i=1}^N \sum_{j=1}^C y_{ij} \log(\hat{y}_{ij}) \quad (1)$$

где:  $N$  – количество обучающих примеров,  $C$  – количество классов,  $y_{ij}$  – бинарная переменная, которая равна 1, если пример  $i$  принадлежит классу  $j$ , и 0 в противном случае,  $\hat{y}_{ij}$  – предсказанная вероятность того, что пример  $i$  принадлежит классу  $j$ .

Если предсказанная вероятность принадлежности к классу «лампа хорошая» близка к 0,25, но не значительно меньше, это означает, что лампа, скорее всего, относится к этому классу. Однако если вероятность значительно меньше 0,25, то лампа, скорее всего, не принадлежит к этому классу.

Для других классов все аналогично: если вероятность близка к порогу, но не значительно меньше, то лампа, скорее всего, принадлежит к этому классу.

Если вероятность принадлежности к классу «лампа исправна» близка к 0,25, но не значительно меньше, лампа классифицируется как «исправная».

Если вероятность принадлежности к классу «лампа старая, но в хорошем рабочем состоянии» близка к 0,5, но не значительно меньше, лампа классифицируется как «старая, но в хорошем рабочем состоянии».

Если вероятность принадлежности к классу «лампа с высокой вероятностью поломки» близка к 0,75, но не значительно меньше, лампа классифицируется как «с высокой вероятностью поломки».

Если вероятность принадлежности к классу «лампа неисправна» близка к 1, лампа классифицируется как «неисправная».

### III. РЕЗУЛЬТАТЫ

Для объективной оценки эффективности модели, учитывающей как ее способность обнаруживать

правильно определенные классы, так и способность отличать их от ложных определений, необходимо использовать сложные метрики, такие как ROC-кривая.

ROC-кривая (Receiver Operating Characteristic) – это график, показывающий эффективность работы классификатора. Она измеряет способность классификатора правильно различать классы, показывая, как часто классификатор правильно определяет положительные случаи (True Positive Rate, TPR) по сравнению с частотой ложных срабатываний (False Positive Rate, FPR). ROC-кривая используется для оценки качества модели классификации, особенно в задачах, где важны как чувствительность (TPR), так и специфичность (True Negative Rate, TNR).

Одним из ключевых аспектов ROC-кривой является ее площадь под кривой (AUC, Area Under the Curve). AUC – это показатель качества модели классификации, который может варьироваться от 0 до 1. Модель с AUC, равным 1, является идеальной, в то время как модель с AUC, равным 0,5, не полностью правильно классифицирует входные данные. AUC также может использоваться для сравнения производительности различных моделей классификации. В контексте данной статьи результаты классификации интерпретируются на ROC-кривой, показанной на рис. 1.

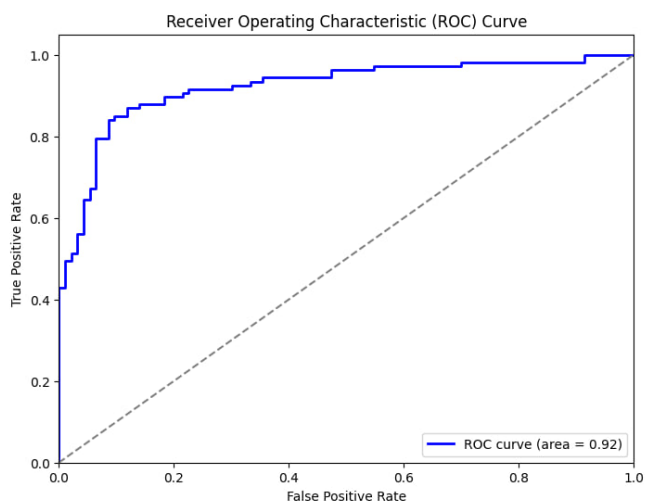


Рис. 1. График ROC-кривой

Если площадь под ROC-кривой равна 0,92, это говорит о высокой эффективности модели. Значение AUC, близкое к 1, означает, что модель обладает высокой способностью различать классы, демонстрируя хорошую разделяемость. В данном случае значение AUC, равное 0,92, указывает на то, что модель классификации в большинстве случаев правильно распознает объекты разных классов, что делает ее надежным инструментом для прогнозирования.

#### IV. МОДЕРНИЗАЦИЯ УЛИЧНОГО ОСВЕЩЕНИЯ С ПОМОЩЬЮ ПЕРОВСКИТНЫХ СОЛНЕЧНЫХ ПАНЕЛЕЙ

Кремниевые панели широко используются в современном уличном освещении, но существует альтернатива – перовскитовые панели. Это новая, но быстро развивающаяся технология с высокой эффективностью преобразования энергии и

возможностью работы в различных условиях освещенности. По сравнению с традиционными кремниевыми панелями перовскитовые панели обладают рядом преимуществ, таких как более высокая эффективность преобразования энергии, возможность работы при более низких уровнях освещенности и более широком угле падения света [4].

Использование перовскитовых материалов в качестве основы для колпака лампы является очень перспективным решением. Перовскитовые материалы обладают уникальными оптическими и электронными свойствами, которые делают их идеальным выбором для применения в фотодетекторах. В закрытом колпаке лампы, где нет прямого контакта с человеком, использование перовскитовых материалов становится особенно ценным, поскольку позволяет обеспечить высокую производительность детектора без необходимости регулярного обслуживания. Кроме того, их способность быстро реагировать по сравнению с кремниевыми фотодетекторами и простота изготовления удовлетворяют требованиям к динамичности системы.

#### V. ИНТЕГРАЦИЯ ТЕХНОЛОГИЙ

Поскольку перовскитовые панели в настоящее время ненадежны [5]. На основе данных, полученных от перовскитовых панелей и других источников, можно разработать алгоритмы для автоматического обнаружения потенциальных неисправностей или ухудшения производительности. Например, аномальные показания датчиков могут указывать на потенциальные проблемы с панелями или системой хранения энергии. Эти алгоритмы могут быть интегрированы в систему мониторинга, что позволит операторам системы быстро реагировать на проблемы и направлять технических специалистов для проведения диагностики и ремонта.

На основе накопленных данных о работе перовскитовых панелей и окружающей среды можно разработать модели предиктивного обслуживания. Эти модели могут предсказывать вероятность возникновения неисправностей или снижения производительности на основе текущих данных и исторической информации. Такие прогнозные модели можно использовать для планирования технического обслуживания и оптимизации затрат на ремонт и замену.

Интеграция перовскитовых солнечных панелей в системы прогнозирования неисправностей и автоматизации процесса технического обслуживания может принести значительные экономические и экологические выгоды. Повышенная оперативность реагирования на неисправности и оптимизация затрат на обслуживание могут снизить эксплуатационные расходы и повысить эффективность системы уличного освещения. Кроме того, повышение энергоэффективности за счет использования солнечной энергии позволяет снизить выбросы углекислого газа и нагрузку на электросети.

#### VI. ЗАКЛЮЧЕНИЕ

Дальнейшее развитие темы данной работы может включать несколько направлений, расширяющих исследования и применение обсуждаемых в ней технологий. Вот несколько предложений.

Применение машинного обучения для оптимизации работы уличных фонарей: Исследования могут включать разработку и тестирование алгоритмов машинного обучения, способных автоматически регулировать интенсивность света уличных фонарей в зависимости от времени суток, погодных условий и уровня освещенности для минимизации энергопотребления и улучшения качества освещения.

Интеграция с интеллектуальными сетями (Smart Grids): Исследование может расширяться до изучения возможности интеграции уличного освещения с интеллектуальными сетями для более эффективного распределения и управления энергией. Это может включать разработку протоколов для обмена данными между уличными фонарями и сетью, а также алгоритмов для оптимизации энергопотребления.

Использование перовскитовых материалов в других компонентах уличного освещения: Помимо использования перовскитовых солнечных панелей, в рамках исследований можно рассмотреть возможность применения перовскитовых материалов в других компонентах уличного освещения, таких как светодиоды или датчики. Это может включать разработку новых типов светодиодных ламп, в которых используются перовскитовые материалы для повышения энергоэффективности и долговечности.

Разработка мобильных приложений для мониторинга и управления уличным освещением: Исследование может включать разработку мобильных приложений, позволяющих пользователям отслеживать состояние уличного освещения в режиме реального времени, получать уведомления о потенциальных проблемах и взаимодействовать с системой управления освещением для решения возникающих вопросов.

Проведение тематических исследований на реальных объектах: Исследование может включать в себя проведение тематических исследований на реальных

объектах, где уже внедрены технологии искусственного интеллекта и перовскита для управления уличным освещением. Это поможет оценить практическую эффективность этих технологий и выявить возможные проблемы и ограничения.

Исследование воздействия на окружающую среду: Исследование может быть расширено за счет изучения воздействия перовскитовых материалов и технологий искусственного интеллекта на окружающую среду, включая экологические последствия производства и использования этих материалов, а также влияние на энергодобавку и климат.

Разработка стандартов и нормативно-правовой базы: Исследования могут включать разработку стандартов и нормативной базы для использования перовскитовых материалов и технологий искусственного интеллекта в уличном освещении с целью обеспечения их безопасного и эффективного применения.

Эти направления могут значительно расширить представленные в статье исследования и привести к разработке новых технологий и подходов к управлению уличным освещением, способных повысить его качество, эффективность и устойчивость.

#### СПИСОК ЛИТЕРАТУРЫ

- [1] "Energy Efficiency Analysis of Street Lighting in Smart Cities." IEEE Xplore, IEEE, 2019, [https://ieeexplore.ieee.org/document/8959824].
- [2] NREL. Best Research-Cell Efficiencies: Emerging Photovoltaics. NREL, accessed 18.07.2023 [Electronic resource]. 2023. URL: https://www.nrel.gov/pv/assets/pdfs/cell-pv-eff-emergingpv.pdf.
- [3] LaValley M.P. Logistic regression // *Circulation*. 2008. Т. 117. №. 18. С. 2395-2399.
- [4] Conings B. et al. Intrinsic thermal instability of methylammonium lead trihalide perovskite // *Adv. Energy Mater.* Wiley Online Library, 2015. Vol. 5, № 15. P. 1500477.
- [5] "Recent Advances in Perovskite Solar Cells: Efficiency, Stability, and Scaling Up." *ACS Energy Lett.*, 2020, 5, 3, 743–760. DOI 10.1021/acsenerylett.9b02720.

# Оптимизация генеративного ридера в составе RAG системы

А. П. Соколов<sup>1</sup>, П. А. Замелин<sup>2</sup>, Ю. Д. Камелина<sup>3</sup>, П. И. Пластова<sup>3</sup>

<sup>1</sup>Московский государственный университет имени М.В. Ломоносова

<sup>2</sup>Национальный исследовательский университет «Высшая Школа Экономики»

<sup>3</sup>Нижегородский государственный университет им. Н.И. Лобачевского

andrey.sokolov@yadro.com

**Аннотация.** В статье предлагается простая методология для оптимизации подсистемы генеративного ридера, являющегося одним из ключевых компонентов RAG систем. Описанная методология может быть использована в качестве предварительной оптимизации, которая позволяет быстро построить вопросно-ответную систему (QA-система) приемлемого качества. Основным принципом этой методологии является использование генеративной модели GPT4 от OpenAI в качестве эталонного генеративного ридера. Наша методология описывает процедуру создания синтетического датасета и определяет основные метрики для оптимизации. Использование синтетических данных позволяет точно контролировать, что изменения, вносимые в QA-систему, приближают ее ответы к эталонным. На основе разработанной методики мы провели ряд простых экспериментов по оптимизации для быстрого поиска лучшей конфигурации подсистемы ридера. Эти эксперименты позволили существенно улучшить качество ответов QA-системы. В результате мы добились относительного прироста средней семантической близости ответов нашего ридера к эталонным GPT4 почти 50 % по шкале BLEURT и 26 % по шкале SAS. Наша методология была проверена для построения русскоязычной вопросно-ответной системы, но она также может быть применена для построения вопросно-ответной системы на любом другом языке без принципиальных изменений.

**Ключевые слова:** RAG, генерация с ассистированием от ретривера, вопросно-ответные системы, LLM, семантическая близость

## I. ВВЕДЕНИЕ

Вопросно-ответные системы – быстро развивающаяся область в компьютерных науках и обработке естественного языка. Разработка таких систем ставит множество проблем. Одной из главных проблем является задача ответа на вопрос в открытом домене (ODQA) [1]. В этом сценарии пользователь задает вопрос на естественном языке, а системы должна дать верный и фактический ответ, т.е. ответ, основанный на фактах из каких-то источников. В сценарии ODQA предполагается, что система может иметь доступ к некоторой большой базе знаний. Часто для оценки качества работы таких систем в качестве такой базы знаний выступает Википедия [2].

Существует множество архитектур ODQA систем [3]. Обычно выделяют два ключевых компонента: ретривер и ридер. Подсистема ретривера отвечает за поиск релевантных документов в базе знаний в ответ на вопрос

пользователя. Подсистема ридера должна на основе вопроса пользователя и найденных релевантных документов дать верный ответ, подкрепленный фактами из документов. Существуют архитектуры, в которых отсутствует один из этих компонентов. Например, в системах, включающих в себя только генеративный ридер (generator-only системы), разработчики стремятся агрегировать всю информацию из базы знаний в ридере, основанном на генеративной языковой модели. Примерами таких систем являются модели T5 [4], GPT-3 [5], GPT-4 [6] и другие.

Одной из многообещающих архитектур ODQA систем является архитектура RAG (Retrieval Augmented Generation) [7]. Эта архитектура предполагает тесную интеграцию непараметрического ретривера и параметрической генеративной модели в качестве ридера. Как результат, RAG-система с «маленьким» ридером с всего 400 млн. параметров превосходит generator-only систему с 11 млрд параметров на 10 % по шкале EM (точное совпадение) на датасете NQ [9]. Вскоре после появления архитектуры RAG появилась архитектура FiD (Fusion-In-Decoder architecture) [8]. Данную архитектуру можно рассматривать как вариант архитектуры RAG. В ней интеграция подсистем ретривера и ридера становится заметно менее тесной. Более того, каждый компонент обучается независимо на специфичной для него задаче. Такое отдаление компонентов позволило добиться преимуществ RAG-подобной FiD системы над полностью параметрической генеративной моделью еще больше. Так, вопросно-ответная система архитектуры FiD, имеющая в своем составе ридер всего с 770 млн. параметров превзошла модель GPT3 со 175 млрд параметров на датасете NQ более, чем на 20 % по шкале EM. Эти выдающиеся результаты привлекли пристальное внимание к архитектуре RAG со стороны сообщества исследователей искусственного интеллекта.

В данной статье мы описываем простую методологию, которая может быть использована в качестве предварительной оптимизации подсистемы генеративного ридера в составе RAG-системы. Сначала мы постулируем ридер на основе GPT4 от OpenAI в качестве эталонного. Далее, на основании широко-известного датасета NQ при помощи машинного перевода и эталонного ридера мы построили датасет ruNQ\_GPT4. Этот датасет пригоден как для выбора оптимизационных метрик, так и для самой оптимизации конфигурации подсистемы ридера. Затем мы выбираем

основные метрики семантической близости, которые наилучшим образом коррелируют с решениями эталонного ридера. После этого мы провели ряд экспериментов, варьируя основные конфигурационные параметры подсистемы ридера, такие как базовая генеративная модель, ее параметры и запрос (промпт). Как результат мы смогли существенно повысить среднюю близость ответов подсистемы ридера к ответам эталонного ридера на базе GPT4.

## II. МЕТОДИКА ОПТИМИЗАЦИИ

### A. Основные принципы

Данная методика нацелена на предварительную оптимизацию подсистемы генеративного ридера в составе RAG QA-системы. Поэтому она не включает в себя дообучения языковой модели ридера. Рассматривается только выбор базовой модели и ее основных конфигурационных параметров. Оценка качества ответов генеративного ридера осуществляется на тестовом датасете, который состоит из набора примеров, содержащих вопрос, релевантный ему документ и ответ, данных эталонным ридером.

В качестве эталонного ридера мы рассматриваем модель GPT4 от OpenAI с некоторым фиксированным промптом для ответа на вопрос с учетом релевантного документа. Также мы фиксируем основные параметры эталонного ридера в некоторые стандартные значения.

В своей работе мы рассматриваем задачу построения вопросно-ответных систем, поддерживающих русский язык. В связи с тем, что количество русскоязычных датасетов, пригодных для отработки вопросно-ответного сценария (QA), сильно ограничено, мы решили сгенерировать специальный QA-датасет на базе широкоизвестного датасета NQ. Вопросы и документы из датасета NQ были переведены на русский язык при помощи машинного перевода. Для этого мы использовали Yandex Translate API [10]. Этот «переводчик» продемонстрировал высокое качество перевода с английского на русский язык в наших экспериментах. Вместо использования ответов из оригинального датасета NQ, которые были получены при помощи разметчиков-людей, мы сгенерировали двойные эталонные ответы при помощи GPT4 ридера.

Мы рассмотрели несколько лексических и семантических мер близости для сравнения ответов с эталоном. Чтобы выбрать из них основную метрику, при помощи которой далее будет осуществляться оптимизация конфигурации ридера, мы снова воспользовались моделью GPT4 в качестве источника информации о семантической близости парных эталонных ответов. Метрика, которая продемонстрировала наибольший уровень корреляции с эталонными оценками GPT4, выбирается как основная.

После генерации тестового датасета и выбора основной оптимизационной метрики мы провели набор экспериментов, варьируя параметры конфигурации подсистемы ридера, чтобы сделать его ответы в семантическом смысле более похожими на ответы эталонного ридера.

### B. Датасет ruNQ\_GPT4

Датасет ruNQ\_GPT4 это вопросно-ответный датасет с четверной разметкой, содержащий 1 тыс. вопросов. Для каждого вопроса в датасете присутствуют два ответа двух эталонных разметчиков и две оценки семантической близости данных ответов. Оценки семантической близости даны двумя другими эталонными разметчиками. В качестве эталонных разметчиков во всех случаях выступает модель GPT4. Для генерации ответов и оценки семантической близости используются различные запросы (промпты). Для генерации парных ответов (как в случае ответов, так и в случае оценок семантической близости) выполняется два последовательных вызова API генеративной модели. При этом промпт и параметры исполнения зафиксированы.

Каждая запись в датасете содержит следующие поля: вопрос ( $q$ ) на естественном языке; документ ( $d$ ), который содержит ответ на вопрос  $q$ ; ответ разметчика 1 ( $a_1$ ); ответ разметчика 2 ( $a_2$ ); оценка семантической близости ответов  $a_1$  и  $a_2$ , данная разметчиком 3 ( $s_1$ ); оценка семантической близости ответов  $a_1$  и  $a_2$ , данная разметчиком 4 ( $s_2$ ).

Оценки семантической близости даны в соответствии со следующей шкалой: «0» – ответы имеют полностью различное значение (семантику); «1» – ответы имеют похожее значение, но есть ряд важных различий; «2» – ответы имеют полностью одинаковое значение.

Также отметим, что вопросы и документы были переведены из датасета NQ при помощи Yandex Translate API. Ответы  $a_1$  и  $a_2$  и оценки семантической близости  $s_1$  и  $s_2$  были сгенерированы при помощи GPT4 API при помощи двух разных запросов.

### C. Анализ метрик

Для оценки близости между сгенерированными и эталонными ответами применяются различные метрики. Наиболее часто применяются лексические и семантические метрики. Лексические метрики основаны на подсчете количества униграмм (слов), которые присутствуют в обоих сравниваемых предложениях. Метрики семантической близости нацелены на оценку того, насколько близким является значение (семантика) сравниваемых предложений. Мы рассмотрели лексические метрики близости F1-score [11] и METEOR [12], а также следующие метрики семантической близости: BERTScore [13], BLEURT [14] и SAS [15].

Табл. I показывает величину коэффициентов корреляции Пирсона между значениями метрик семантической близости и их оценками от эталонного разметчика ( $s_1$  и  $s_2$ ) на датасете ruNQ\_GPT4. Как мы видим, популярные метрики лексической и семантической близости хорошо согласуются с оценками GPT4. Лексические метрики показывают более низкую корреляцию с эталоном, чем метрики семантической близости. Также видно, что двойные оценки семантической близости от эталонного разметчика очень хорошо согласуются друг другом (см. высокое значение корреляции между  $s_1$  и  $s_2$ ). Максимальное значение корреляции достигается для метрик SAS и BLEURT. В связи с этим далее мы будем



рассматривать эти две метрики как основные для следующих оптимизаций подсистемы ридера.

ТАБЛИЦА I. КОРРЕЛЯЦИЯ МЕТРИК СЕМАНТИЧЕСКОЙ БЛИЗОСТИ С ОЦЕНКАМИ GPT4

Метрика	$s_1$	$s_2$
SAS	<b>0.816</b>	<b>0.817</b>
BLEURT	0.785	0.791
BERTScore (F1)	0.769	0.772
F1	0.767	0.773
BERTScore (точность)	0.736	0.743
BERTScore (полнота)	0.713	0.711
METEOR	0.713	0.715
$s_1$	1	0.972
$s_2$	0.972	1

Возникает вопрос, насколько близки ответы ридера на базе модели GPT4 между собой. В связи с тем, что датасет ruNQ\_GPT4 содержит двойные эталонные ответы мы можем оценить эту близость. Гистограммы на рис. 1 показывают распределение значений семантической близости SAS и BLEURT между парами эталонных ответов  $a_1$  и  $a_2$ .

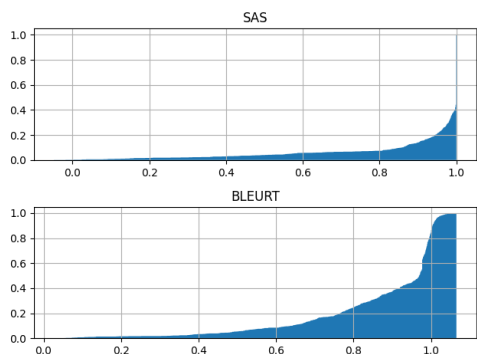


Рис. 1. Распределение семантических близостей между парами эталонных ответов GPT4

Видно, что несмотря на тот факт, что ответы даны одной и той же генеративной моделью с одним и тем же промптом, существуют примеры пар ответов, которые существенно отличаются друг от друга в семантическом смысле. В связи с этим для оценки схожести двух ридеров между собой далее мы будем полагаться на величину средней близости по SAS и BLEURT на тестовом датасете ruNQ\_GPT4.

### III. ЭКСПЕРИМЕНТЫ ПО ОПТИМИЗАЦИИ

В данном разделе представлены результаты экспериментов, которые были проведены нами в рамках оптимизации конфигурационных параметров подсистемы ридера.

#### A. Как промпт влияет на качество ответов ридера?

Сначала мы исследовали, как промпт (запрос) влияет на ответы ридера. Сначала ридеру с одним запросом (промптом) последовательно подавались все вопросы из датасета ruNQ\_GPT4 и оценивалась их средняя близость с эталонными ответами. Далее, запрос менялся на следующий и процедура повторялась. В данных экспериментах в качестве эталонного рассматривался ответ  $a_1$ . Табл. II содержит результаты экспериментов по вариации запроса (промпта).

ТАБЛИЦА II. ЭФФЕКТИВНОСТЬ ТЕХНИК ПОСТРОЕНИЯ ЗАПРОСА?

Техника построения запроса	SAS (средний)	BLEURT (средний)
«Улучшитель» запроса GPT (GPT Prompt Enhancer) + «Несколько примеров» (Few-Shot Prompting)	<b>0.695</b>	<b>0.534</b>
«Улучшитель» запроса GPT «Несколько примеров»	0.689	0.527
Упрощенная техника «Цель рассуждений»	0.677	0.514
Эмоциональный запрос	0.592	0.407
Изохронный подкуп	0.582	0.398
Упрощенная техника «Дерево рассуждений»	0.555	0.369
Акцент CapsLock	0.543	0.348
Простой «подкуп»	0.515	0.316
	0.499	0.306

Видно, что техники построения запросов — мощный инструмент, способный существенно улучшить качество ответов генеративного ридера. Наилучший результат показала комбинация техники «Улучшитель запроса GPT» (GPT Prompt Enhancer) и техника «Несколько примеров» (Few-Shot Prompting).

#### B. Как параметры модели влияют на качество ответов?

Чтобы оценить, как параметры исполнения (температура, top\_k и top\_p) [16] влияют на качество генерируемых ответов, мы провели серию экспериментов на датасете ruNQ\_GPT4 для следующей сетки данных параметров: top\_p=[0.1, 0.45, 0.8], top\_k=[10, 45, 80], температура=[0.2, 0.45, 0.7]. Эксперименты показали, что варьирование этих параметров оказывает незначительное влияние на среднюю семантическую близость ответов оптимизируемого ридера с эталонными. Так, средние значения SAS и BLEURT на датасете ruNQ\_GPT4 варьировались менее, чем на 2%.

#### C. Как базовая модель влияет на качество ответов?

Далее мы проанализировали, как вариация базовой модели влияет на семантическую близость ответов к эталонными. Для объективности сравнения мы рассматривали только модели класса 7B (7 млрд. параметров). Также мы рассматривали только публично-доступные модели, которые показали высокие результаты в русскоязычном бенчмарке SuperGLUE [17]. Список рассмотренных моделей и достигнутые величины средней семантической близости к эталонным ответам приведены в Таблице III. Как видно, выбор базовой модели существенно влияет на качество ответов. Таким образом, процедура дообучения базовой модели является принципиально-важным способом оптимизации вопросно-ответных систем, построенных на базе генеративных моделей.

ТАБЛИЦА III. ВЛИЯНИЕ БАЗОВОЙ МОДЕЛИ НА ОТВЕТЫ

Модель с HuggingFace [18]	SAS (средний)	BLEURT (средний)
IlyaGusev/saiga2_7b_gguf	0.817	0.303
TheBloke/saiga_mistral_7b-GPTQ	0.790	0.385
IlyaGusev/saiga_mistral_7b_lora	0.779	0.359
IlyaGusev/saiga_mistral_7b_gguf	0.778	0.356
TheBloke/Llama-2-7B-GGUF	0.748	0.397
TheBloke/Llama-2-7B-GGUF	0.748	0.396
TheBloke/Llama-2-7B-GPTQ	0.728	0.363

IV. ДИСКУССИЯ

В предыдущих разделах мы использовали ответы ридера на базе модели GPT4 в эталонных. Построение запросов и вариация базовой модели позволили сделать ответы вопросно-ответной системы намного более похожими на ответы GPT4, в семантическом смысле. Возникает вопрос, как изменилась смысловая «похожесть» ответов оптимизируемого ридера на ответы, данные разметчиками людьми в датасете NQ? Таблица IV показывает величину относительного изменения средних величин метрик BLEURT и SAS на датасетах ruNQ\_GPT4 (ответы получены при помощи модели GPT4) и ruNQ (ответы получены путем машинного перевода ответов разметчиков людей из датасета NQ), которая возникла в результате оптимизации конфигурации подсистемы ридера.

ТАБЛИЦА IV. ИЗМЕНЕНИЕ СЕМАНТИЧЕСКОЙ БЛИЗОСТИ ОТВЕТОВ РИДЕРА В РЕЗУЛЬТАТЕ ОПТИМИЗАЦИИ ПО ОТНОШЕНИЮ К ОТВЕТАМ GPT4 И ОТВЕТАМ ЧЕЛОВЕКА

Датасет	Относительное изменение среднего BLEURT, %	Относительное изменение среднего SAS, %
ruNQ_GPT4	49.6	26.6
ruNQ (ответы человека)	-28.7	-24.1

Мы видим, что после оптимизации конфигурации ридера ее ответы стали намного ближе к ответам GPT4, но в то же время значительно менее похожими на ответы разметчиков людей. Неясно чьи ответы лучше: человека или робота. Ясно то, что результат оптимизации генеративного ридера целиком и полностью зависит от того, что будет взято в качестве эталона и, следовательно, в какую сторону будет идти процесс оптимизации.

V. ЗАКЛЮЧЕНИЕ

Мы представили простую методику для быстрой предварительной оптимизации генеративного ридера в составе вопросно-ответной системы, построенной по архитектуре RAG. Показано, что такие простые методы, как вариация базовой модели и техники запроса (prompt engineering), позволяют быстро поднять качество ответов системы до приемлемого уровня. Дальнейшее улучшение качества ответа может быть достигнуто иными способами, в том числе, за счет дообучения базовой модели.

СПИСОК ЛИТЕРАТУРЫ

[1] Ellen M. Voorhees and Dawn M. Tice. 2000. The TREC-8 question answering track. In Proceedings of the Second International Conference on Language Resources and Evaluation (LREC' 00), Athens, Greece. European Language Resources Association (ELRA).

[2] Wikipedia. 2004. Wikipedia. PediaPress.

[3] Qin Zhang, Shangsi Chen, Dongkuan Xu, Qingqing Cao, Xiaojun Chen, Trevor Cohn, and Meng Fang. 2023. A Survey for Efficient Open Domain Question Answering. In Proceedings of the 61st

Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pages 14447–14465, Toronto, Canada. Association for Computational Linguistics.

[4] Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, Peter J Liu, et al. 2020. Exploring the limits of transfer learning with a unified text-to-text transformer. *J. Mach. Learn. Res.*, 21(140):1–67.

[5] Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel Ziegler, Jeffrey Wu, Clemens Winter, Chris Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, and Dario Amodei. 2020. Language models are few-shot learners. In *Advances in Neural Information Processing Systems*, volume 33, pages 1877–1901. Curran Associates, Inc.

[6] Benj Edwards. "OpenAI's GPT-4 exhibits "human-level performance" on professional benchmarks". *Ars Technica*. Archived from the original on March 14, 2023. Retrieved March 15, 2023.

[7] Patrick Lewis, Yuxiang Wu, Linqing Liu, Pasquale Minervini, Heinrich Kuttler, Aleksandra Piktus, Pontus Stenetorp, and Sebastian Riedel. 2021. PAQ: 65 Million Probably-Asked Questions and What You Can Do With Them. *Transactions of the Association for Computational Linguistics*, 9:1098–1115.

[8] Gautier Izacard and Edouard Grave. 2021. Leveraging passage retrieval with generative models for open domain question answering. In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume*, pages 874–880, Online. Association for Computational Linguistics.

[9] Tom Kwiatkowski, Jennimaria Palomaki, Olivia Redfield, Michael Collins, Ankur Parikh, Chris Alberti, Danielle Epstein, Illia Polosukhin, Jacob Devlin, Kenton Lee, et al. 2019. Natural questions: a benchmark for question answering research. *Transactions of the Association for Computational Linguistics*, 7:453–466.

[10] Yandex Translate API - <https://yandex.com/dev/translate/>

[11] Danqi Chen, Adam Fisch, Jason Weston, and Antoine Bordes. 2017. Reading Wikipedia to answer open-domain questions. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1870–1879, Vancouver, Canada. Association for Computational Linguistics.

[12] Banerjee, S. and Lavie, A. "METEOR: An Automatic Metric for MT Evaluation with Improved Correlation with Human Judgments" in *Proceedings of Workshop on Intrinsic and Extrinsic Evaluation Measures for MT and/or Summarization at the 43rd Annual Meeting of the Association of Computational Linguistics (ACL-2005)*, Ann Arbor, Michigan, June 2005

[13] Tianyi Zhang, Varsha Kishore, Felix Wu, Kilian Q. Weinberger, Yoav Artzi. 2020. BERTScore: Evaluating Text Generation with BERT. *International Conference on Learning Representations*.

[14] Thibault Sellam, Dipanjan Das, and Ankur Parikh. 2020. BLEURT: Learning Robust Metrics for Text Generation. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 7881–7892, Online. Association for Computational Linguistics.

[15] Julian Risch, Timo Möller, Julian Gutsch, and Malte Pietsch. 2021. Semantic Answer Similarity for Evaluating Question Answering Models. In *Proceedings of the 3rd Workshop on Machine Reading for Question Answering*, pages 149–157, Punta Cana, Dominican Republic. Association for Computational Linguistics.

[16] Transformers Library, Text generation parameters, [https://huggingface.co/docs/transformers/en/main\\_classes/text\\_generation](https://huggingface.co/docs/transformers/en/main_classes/text_generation)

[17] Russian SuperGLUE - <https://russiansuperglue.com/>

[18] Hugging Face - <https://huggingface.co/>

# Аппаратная реализация ART-2 классификатора

О. И. Буренева<sup>1</sup>, Я. Б. Ибрагим<sup>1</sup>, М. С. Прасад<sup>2</sup>, Ш. Верма<sup>2</sup>

<sup>1</sup>Санкт-Петербургский государственный электротехнический университет «ЛЭТИ» им. В.И. Ульянова (Ленина)

<sup>2</sup>Университет Эмити. Институт космической науки и технологий Эмити  
OIBureneva@etu.ru

**Аннотация.** В статье представлено исследование классификаторов на основе адаптивной резонансной теории. Рассмотрена модель последовательного поиска резонирующего нейрона и предложены варианты реализации элементов этой модели. Показаны результаты имплементации модели в ПЛИС IntelFPGA.

**Ключевые слова:** адаптивная резонансная теория, распознавание полутоновых шаблонов, FPGA реализация ART архитектур, аппаратный классификатор

## I. ВВЕДЕНИЕ

Теория адаптивного резонанса (ART, Adaptive Resonance Theory) – это биологический термин, обозначающий способность человеческого мозга удерживать ранее усвоенную информацию, сохраняя при этом способность к приему новой. Теория основана на предположении, что мозг имеет некоторые знания, называемые категориями распознавания, которые активируются, когда поступающая сенсорная информация соответствует им, то есть входит в резонанс. Это позволяет организму эффективно интегрировать новую информацию в существующие знания [1]. Параметр бдительности в ART, введенный в [2], показывает возможности отслеживать и регулировать уровень активации категорий распознавания, что обеспечивает адаптацию к новой информации, сохраняя стабильность и предотвращая путаницу или чрезмерное обобщение.

ART-классификатор представляет собой вычислительную модель, основанную на принципах адаптивной резонансной теории, для выполнения задач распознавания образов и классификации. Она объединяет сенсорные данные, поступающие снизу вверх, с ожиданиями сверху вниз для создания резонансных паттернов нейронной активности, что позволяет осуществлять динамическое и адаптивное обучение.

ART решают дилемму «пластичность-стабильность», поэтому ART-сети приобрели популярность в инженерных приложениях благодаря своим преимуществам, таким как быстрое, стабильное и инкрементное обучение, а также способность работать с ограниченной памятью и простыми алгоритмами [3]. Термин «быстрое обучение» в данном контексте означает способность нейронов сети быстро сходиться к

Работа выполнена при поддержке Минобрнауки России в рамках соглашения № 075-15-2022-291 от 15.04.2022 о предоставлении гранта в форме субсидий из федерального бюджета на осуществление государственной поддержки создания и развития научного центра мирового уровня «Павловский центр «Интегративная физиология – медицине, высокотехнологичному здравоохранению и технологиям стрессоустойчивости».

конечным значениям весов при каждом предъявлении входного образца. Эти свойства, наряду с другими делают ART-сети привлекательными для использования в различных приложениях [4].

## II. КЛАССИФИКАТОРЫ НА ОСНОВЕ ART

Широкий класс моделей ART-классификаторов строится в соответствии с базовой архитектурой. Ее модификация позволяет получать разные варианты. Базовая архитектура ART-модели, ориентированная на аппаратную реализацию, состоит из следующих элементов.

1. Входной слой (F0): начальный слой модели, через который исходные данные поступают в сеть ART, при аппаратной реализации этот слой определяется как входные порты и выполняет следующие функции:

- предварительная обработка: масштабирование, нормализация или извлечение признаков;
- представление паттерна: каждый нейрон в этом слое соответствует признаку во входных данных, шаблон активации в этом слое представляет собой вектор входных данных.

2. Конкурентный слой (F1) реализует основной процесс обучения, выполняя представление и сопоставление категорий. В аппаратной реализации эти процессы требуют нескольких итераций, пока процесс не станет стабильным и не будет достигнуто условие резонанса. Ключевые элементы слоя F1:

- нейроны (кластеры): каждый нейрон в слое F1 представляет собой прототип или центр кластера, активация нейронов представляет степень соответствия между входным образцом и хранящимися в памяти прототипами;
- конкурентная динамика: нейроны в этом слое конкурируют за то, чтобы стать активными, тот нейрон, чей прототип ближе всего к входному образцу, обычно побеждает в этом соревновании, то есть применяется механизм «победитель получает все» (WTA, Winner Take All);
- резонансное состояние: победивший нейрон инициирует состояние резонанса, если совпадение между входным паттерном и прототипом нейрона превышает определенный порог (параметр бдительности), обычно победителем считается нейрон, который поднимается выше порога. Если условие резонанса выполняется, входной сигнал считается достаточно похожим на прототип, то происходит дообучение.

3. Параметр бдительности ( $\rho$ ): является важнейшим компонентом, который контролирует детализацию категорий, формируемых слоем, отвечающим за сравнение. Он определяет порог соответствия между входным образцом и хранящимися в памяти прототипами, в чем и заключается стабильность АРТ-классификатора:

- высокая бдительность: параметр бдительности требует очень близкого соответствия между входным образцом и прототипом, что приводит к формированию специфических категорий;
- низкая бдительность: параметр бдительности позволяет формировать более широкие, обобщенные категории;
- адаптивная настройка: в зависимости от решаемых задач, некоторые АРТ позволяют адаптивно регулировать параметр бдительности, чтобы динамически балансировать между стабильностью и пластичностью.

4. Механизм сброса и обучающий слой: механизм сброса – уникальная особенность АРТ-классификатора, которая обеспечивает стабильное обучение: типичный поток обработки входных паттернов выглядит следующим образом:

- обнаружение несоответствия: если условие резонанса не выполняется (т.е. совпадение между входным паттерном и прототипом ниже порога бдительности), подается сигнал сброса;
- обновление прототипа: при перезагрузке текущий прототип либо корректируется, чтобы лучше соответствовать входному шаблону, либо создается новый прототип, если соответствие слишком плохое, что не позволяет классификатору изменять существующие категории непропорционально.

Таким образом, после обновления прототипов происходит обучение. Обучение в АРТ включает в себя обновление весов связей между слоями F0 и F1: Обучение можно разделить на 2 части в зависимости от скорости, с которой они происходят:

- быстрое обучение: реализуется с использованием правил обучения Хебба, где веса быстро обновляются, чтобы отразить новые паттерны [5].
- медленное обучение: включает в себя постепенную корректировку весов для стабилизации процесса обучения и предотвращения резких изменений из-за шума или выбросов.

5. Слой нормализации и механизмы управления усилением: в некоторых моделях АРТ может присутствовать слой нормализации, чтобы обеспечить нормализацию входных векторов перед обработкой конкурентным слоем. Обычно это включает вычисление функции нормы для каждого из входов. Это помогает поддерживать масштаб входных признаков и обеспечивает последовательное обучение.

АРТ-сети часто включают механизмы контроля усиления, чтобы регулировать активность нейронов в зависимости от локальности или глобальности контроля усиления:

- глобальный контроль усиления: обеспечивает сохранение общей активности нейронов слоя F1 в пределах определенного диапазона, предотвращая активацию в режиме «бегства»;
- локальный контроль усиления: поддерживает относительные уровни активности отдельных нейронов, обеспечивая сбалансированность процесса соревнования.

### III. РЕАЛИЗАЦИЯ КЛАССИФИКАТОРА НА ОСНОВЕ АРТ-2

АРТ-2 была впервые представлена в [6] как расширение АРТ-1 для улучшения ограничений этой сети и применения к реальным ценным данным. Использование программируемых логических интегральных микросхем для реализации АРТ классификаторов показано в [7], реализация архитектуры АРТ-2, показанная на рис. 1, была в значительной степени вдохновлена работой [8].

#### A. Базовые операции классификатора

Алгоритм классической реализации АРТ-2 можно разбить на 9 этапов,  $i$  на всех этапах находится в диапазоне от 1 до  $n$ . Архитектура, реализующая алгоритм, показана на рис. 1. Работа классификатора предполагает выполнение следующих этапов.

1. Вход и начальная комбинация: нейроны входного слоя  $L_0$ , состоящего из  $n$  нейронов, принимают потоки входных двоичных чисел и комбинируют их с сигналами нейронов выходного слоя  $L_4$  в соответствии с формулой:

$$L1_i = L0_i + a L4_i.$$

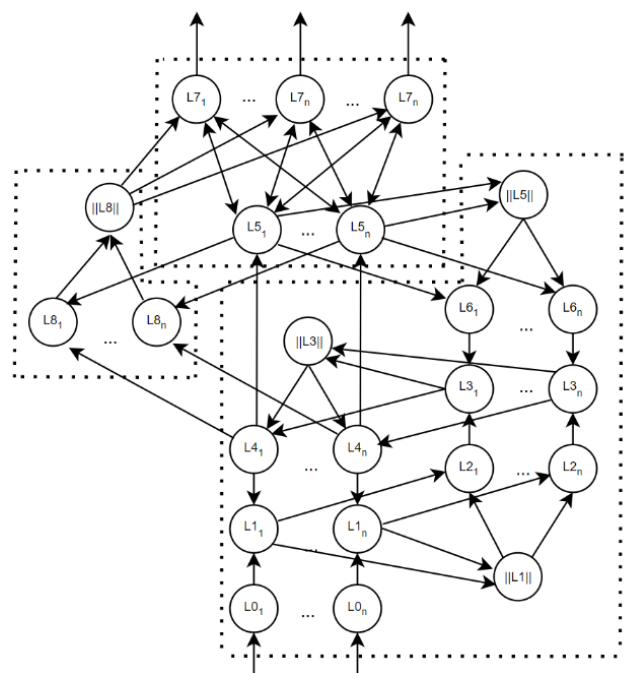


Рис. 1. Архитектура АРТ-2 классификатора

2. Нормализация выходов комбинированного слоя: выходной нейрон из комбинированного слоя  $L1_i$  отправляется в слой  $L2$  и делится на меру своей длины, в данном случае на сумму всех  $i$ -ых нейронов, которые поступают в этот порт для обработки:

$$\|L1\| = \sqrt{\sum_{i=1}^n L1_i^2}, \quad L2_i = \frac{L1_i}{\epsilon + \|L1\|}$$

где  $\epsilon$  – небольшая константа, предотвращающая деление на 0 или любых других аномалий.

3. Пороговая обработка и подавление шумов: слой  $L3$  обеспечивает вычисление порога для каждого входа, тем самым подавляя шум. Для расчета порога используется следующая функция:

$$L3_i = \text{threshold}(L2_i) + b \text{threshold}(Q_i).$$

где  $\text{threshold} (*)$  математически определяется следующим уравнением:

$$\text{threshold}(x) = \begin{cases} x & \text{при } x > \theta \\ 0 & \text{иначе} \end{cases}$$

где  $\theta$  пороговое значение для шума.

4. Нормализация выходов слоя  $L3$ : выходной нейрон из слоя подавления шума  $L3_i$  передается в слой  $L4$  и делится на сумму всех  $i$ -ых нейронов, которые хранятся в этой сущности для обработки:

$$\|L3\| = \sqrt{\sum_{i=1}^n L3_i^2}, \quad L4_i = \frac{L3_i}{\epsilon + \|L3\|}$$

5. Стабилизация с помощью итераций: сигналы  $L5_i$  получаются из  $L4_i$  и нормируются:

$$\|L5\| = \sqrt{\sum_{i=1}^n L5_i^2}, \quad L6_i = \frac{L5_i}{\epsilon + \|L5\|}$$

Слой отвечает за повторение процесса обучения в течение определенного времени, пока нейрон из слоя может быть заторможен, и значения перестают меняться или меняются минимально.

6. Активация слоя распознавания: выходы слоя  $L5$  поступают в слой распознавания  $L7$ :

$$L7_k = \sum_{i=1}^n L9_{ik} L5_i.$$

где  $k$  изменяется в диапазоне от 1 до количества уже распознанных шаблонов в процессе работы.

Коррекция слоя  $L5$  выполняется следующим образом:

$$L5_i = L4_i + d L9_{ikmax}.$$

где  $L9_{ik}$  представляет собой вес от нейронов стабилизационного слоя  $L5$  к нейронам  $L7$ .

7. Расчет управляющего слоя: сигналы управляющего слоя рассчитываются следующим образом:

$$L8_i = \frac{L4_i + c L5_i}{\epsilon + \|L4\| + \|L5\|}, \quad \|L8\| = \sqrt{\sum_{i=1}^n L8_i^2}$$

8. Тест на бдительность: если нормальная функция выхода управляющего слоя  $L8$  превышает параметр бдительности  $\rho$ , то принимается решение, что элемент в слое  $L7$  с индексом  $j_{max}$  является «победителем» и переходит в следующий слой, в связи с тем, что в принятой модели используется идеология «победитель получает все». В противном случае самый последний нейрон удаляется из конкурентного слоя. Это означает, что на данном этапе классификатор начал обнаруживать новый класс объектов.

9. Создание нового нейрона: проверяются элементы, поступающие на этот слой, если они пусты, то это подтверждает, что действительно был обнаружен новый объект или класс объектов, и это отражается путем увеличения значения  $k$ . В случае, если входные данные не пусты, нейрон возвращается в слой конкуренции для дальнейших итераций.

### V. Реализация классификатора

С точки зрения оптимизации рассмотренной архитектуры можно выбрать различные пути вычисления сигналов на разных слоях, особенно это касается сложных и долгих арифметических действий (деление, извлечение корня квадратного, многократное суммирование). При этом многое зависит и от характеристик конкретного кристалла FPGA, выбранного для имплементации: наличие встроенных арифметических блоков, возможности использования логических ячеек в арифметическом режиме, наличие блочной памяти и др. При аппаратной реализации и оптимизации этой модели архитектуры выбран подход, основанный на предложенном алгоритме: будут реализованы расчеты, предложенные в теоретическом описании.

Разработанная аппаратная реализация ART-2 представлена языковым описанием на Verilog HDL. Реализация представляет собой структурное описание с использованием модулей слоев, модулей нейронов, модулей базовых арифметических операций, таких как пользовательские реализации сумматора, умножителя, делителя, извлекателя квадратного корня и других. Рассматривались различные реализации модулей, и на основании сравнения их характеристик выбирались предпочтительные решения.

Для приближения к «аналоговому» характеру классификатора можно использовать имитацию непрерывности путем изменения разрядности цифровых данных, соответствующим образом подстраиваясь под диапазоны входных сигналов. Базовые элементы классификатора описаны как параметризованные модули, что позволяет гибко работать с различными диапазонами данных. При этом рассмотрены варианты как работы с фиксированной, так и с плавающей точкой. Основными вычислительными элементами классификатора являются:

Сумматор, выполнялся по классической реализации сложения двух чисел; при вычислении суммирования

многих слагаемых, например при расчете норм, использовалась древовидная структура Уоллеса.

Умножитель, применяемый при масштабировании и нормализации, был реализован в комбинационном и конвейерном вариантах, при ускорении можно использовать быстрые умножители при одном зафиксированном множителе [9].

Делитель, основанный на идеи нахождения взаимно обратных чисел делителя, а затем умножения на них для получения результата деления [10]. Недостатком такого подхода является его итерационность, но при использовании конвейеризации других модулей, этот недостаток компенсируется.

Вычислитель обратной функции, определяющий величину, обратную для заданного числа, был построен с помощью рассмотренных ранее пользовательских арифметических модулей с использованием заданного числа и предопределенных констант.

Описание слоев выполнено с использованием структурного подхода с применением оператора генерации. Полученные RTL описания весьма объемны, поэтому, поэтому на рис. 2 в качестве примера приводится только первый сгенерированный экземпляр нейрона слоя L1.

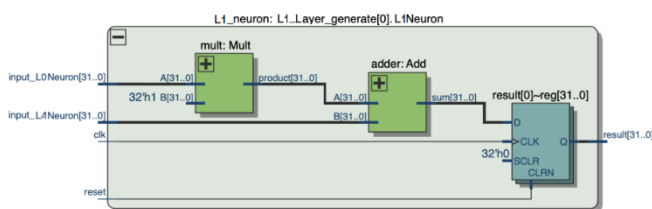


Рис. 2. Реализация нейрона L0 слоя

Полностью собранная архитектура модели была скомпилирована для имплементации в микросхему Intel FPGA Cyclone10LP. Для 32-разрядной имплементации классификатора на 10 входов был использован кристалл 10CL080, ограничением при выборе являлось количество пользовательских контактов ввода/вывода. Микросхема имеет 244 встроенных умножителя, что являлось недостаточным для имплементации классификатора, поэтому частично арифметические операции предполагают использование логических ячеек в арифметическом режиме, при котором возможна реализация ускоренной передачи сигналов. Конвейеризация устройства выполнялась с использованием 10 ступеней, что является компромиссом между частотой работы схемы и аппаратными затратами.

#### IV. ЗАКЛЮЧЕНИЕ

Одним из главных преимуществ архитектур ART является скорость, с которой они могут обрабатывать данные (особенно при параллельной аппаратной реализации), и относительно небольшое количество эпох, которые обычно требуются для сходимости. Это сочетается с работой в режиме реального времени, что делает их эффективными при работе с потоковыми данными или наборами данных, которые слишком велики, чтобы полностью поместиться в памяти.

Второе отличие – конфигурируемость. Архитектуры ART не требуют выбора количества кластеров, в отличие от многих других алгоритмов кластеризации. Между тем, выбор конкретной ART архитектуры и разумного значения бдительности может позволить обнаружить полезные кластеры без необходимости изменять множество чувствительных значений параметров. Это является актуальным для FPGA реализации, при которой желательно заранее определить количество кластеров без возможности добавления новых. При этом, возможно появление отрицательного результата, когда классификатор будет сообщать об отсутствии соответствующего входным сигналам шаблона. При заранее известном количестве выходных кластеров, такой подход является допустимым.

Важной особенностью является объяснимость работы ART классификатора. После достаточного обучения шаблоны классификатора позволяют интерпретировать результаты его обучения.

Распараллеливание при аппаратной реализации дает возможность снизить стоимость, энергопотребление и объем памяти по сравнению с аппаратными реализациями других нейронных сетей.

#### СПИСОК ЛИТЕРАТУРЫ

- [1] Grossberg S. Adaptive Resonance Theory: How a brain learns to consciously attend, learn, and recognize a changing world // *Neural Networks*. – 2013. – Vol. 37. – P. 1–47. DOI:10.1016/j.neunet.2012.09.017.
- [2] Carpenter G. A., Grossberg S. A Massively Parallel Architecture for a Self-Organizing Neural Pattern Recognition Machine. 1986.
- [3] Wunsch D. C. ART properties of interest in engineering applications // 2009 International Joint Conference on Neural Networks. 2009. DOI:10.1109/IJCNN.2009.5179094.
- [4] Silva L. E. B. D., Elnabarawy I., Wunsch D. C. A Survey of Adaptive Resonance Theory Neural Network Models for Engineering Applications // *Neural Networks*. 2019. Vol. 120 (2).
- [5] Milner P. A brief history of the Hebbian learning rule // *Canadian Psychology*. 2003. Vol. 44. P. 5–9.
- [6] Carpenter G. A., Grossberg S. ART 2: self-organization of stable category recognition codes for analog input patterns // *Applied Optics*. 1987. Vol. 26(23), P. 4919–30. DOI: 10.1364/AO.26.004919.
- [7] Bureneva O. I., Prasad M. S., Verma S. FPGA-based Hardware Implementation of the ART-1 classifier // 2023 XXVI International Conference on Soft Computing and Measurements (SCM). 2023. P. 171–174. DOI: 10.1109/SCM58628.2023.10159040.
- [8] Bukhanov D. G., Polyakov V. M. An Approach to Improve the Architecture of ART-2 Artificial Neural Network Based on Multi-Level Memory // *Fuzzy Technologies in the Industry (FTI 2018)*. Proceedings of the II International Scientific and Practical Conference. 2018, P. 235–242.
- [9] Bureneva O., Mironov S. Fast FPGA-Based Multipliers by Constant for Digital Signal Processing Systems // *Electronics*. 2023. Vol 12, 605. DOI: 10.3390/electronics12030605.
- [10] Bureneva O. I., Kaidanovich O. U. FPGA-based Hardware Implementation of Fixed-point Division using Newton-Raphson Method // 2023 IV International Conference on Neural Networks and Neurotechnologies (NeuroNT), Saint Petersburg, Russian Federation. 2023, P. 45–47. DOI: 10.1109/NeuroNT58640.2023.10175844.

# Trajectory Analysis of a Swarm of 3U CubeSats in Sun-Synchronous Orbit with Varying RAAN and Non-Overlapping Revisit Time

Shivani Verma

*Amity Institute of Space Science & Technology,  
Amity University  
Noida, India*

M. S. Prasad

*Amity Institute of Space Science & Technology,  
Amity University  
Noida, India*

<https://orcid.org/0000-0001-9057-9631>

**Abstract**—This paper explores the trajectory dynamics of a swarm of 3U CubeSats deployed in a Sun-synchronous orbit (SSO) at an altitude of 600 km with varying Right Ascension of the Ascending Node (RAAN). The primary objective is to achieve non-overlapping revisit times over a specified ground target located at a latitude of  $28.5355^\circ$  N and a longitude of  $77.3910^\circ$  E. By varying the RAAN values among the CubeSats, this paper analyzes the effectiveness of different orbital configurations in maximizing spatial and temporal coverage. The study utilizes astrodynamics libraries to propagate the orbits and compute the ground tracks and revisit times. Results indicate that strategic variation in RAAN significantly enhances the temporal resolution of ground observations, ensuring that each CubeSat revisits the target location at distinct, non-overlapping intervals. The findings contribute to the design and deployment strategies of CubeSat constellations for Earth observation, disaster management, and remote sensing missions, providing a robust framework for maximizing coverage and minimizing redundancy in data acquisition.

**Keywords**—*CubeSat, Sun-synchronous orbit, RAAN, revisit time, trajectory analysis, remote sensing, satellite swarm, Earth observation*

## I. INTRODUCTION

The proliferation of CubeSat missions has revolutionized the approach to space exploration and satellite deployment, particularly for Earth observation, communication, and scientific research. Among the various configurations, the 3U CubeSat, with its compact form factor and versatile capabilities, has emerged as a popular choice for a multitude of space missions. Leveraging the affordability and flexibility of CubeSats, researchers and engineers are exploring advanced deployment strategies, including the formation of swarms, to enhance mission capabilities and achieve objectives that would be infeasible with a single satellite.

Sun-synchronous orbits (SSO) are particularly advantageous for Earth observation missions due to their consistent lighting conditions, which enable continuous observation of the Earth's surface under nearly identical illumination. SSOs maintain a constant angle between the plane of the orbit and the direction to the Sun, allowing satellites to pass over the same area at approximately the same local solar time on each orbit. This consistency is particularly valuable for comparative analysis and long-term monitoring.

In the context of a swarm of 3U CubeSats, the deployment in a Sun-synchronous orbit offers additional opportunities to optimize coverage and revisit times. By carefully selecting the right ascension of the ascending node (RAAN) for each satellite within the swarm, it is possible to achieve a coordinated coverage pattern that maximizes efficiency and ensures non-overlapping revisit times. Varying the RAAN allows the satellites to distribute their ground tracks over the targeted regions more uniformly, thereby reducing redundancy and enhancing the temporal resolution of observations.

This study focuses on the trajectory analysis of a swarm of 3U CubeSats in a Sun-synchronous orbit at an altitude of 600 km, with varying RAAN values. The primary objective is to investigate the impact of RAAN variation on the orbital dynamics and revisit times of the CubeSat swarm. By leveraging numerical simulations and analytical techniques, this research aims to develop a comprehensive understanding of how different RAAN configurations can be utilized to optimize the coverage and revisit times for Earth observation missions.

The organization of the paper is as follows: Section II reviews the existing literature on CubeSat swarms and Sun-synchronous orbits. Section III describes the theoretical perspective of 3U cubesats and swarm techniques. Section IV presents proposed methodology and the simulations results obtained. Section V concludes the paper with a summary of the key contributions and potential directions for future research.

## II. LITERATURE REVIEW

Malkin, T. L. et al., in [1] provides an overview of recent developments in CubeSat constellations for Earth observation missions, highlighting the advantages of using swarms for improved coverage and revisit times. The survey in [2] offers insights into the challenges and advancements in formation flying and swarm control techniques for CubeSat missions, including coordination strategies and relative navigation algorithms. [3] explores various CubeSat constellation concepts, including distributed swarms and coordinated formations, and discusses their applications in Earth observation, communications, and scientific research. [4] provides a comprehensive overview of Sun-synchronous orbit design principles, orbital mechanics, and mission planning considerations, making it a valuable resource for understanding the theoretical foundations of SSOs. [5]

presents a detailed analysis of Sun-synchronous orbits, including their definition, properties, and applications in Earth observation missions. It discusses the advantages of SSOs for remote sensing and environmental monitoring. [6] offers insights into the design, coverage characteristics, and availability of Sun-synchronous orbits, addressing practical considerations for mission planners and satellite operators. The study in [7] investigates the optimization of CubeSat swarm configurations in Sun-synchronous orbits to enhance Earth observation capabilities, focusing on coverage, revisit times, and data acquisition strategies. [8] presents a detailed analysis of a CubeSat constellation in Sun-synchronous orbit for radio occultation measurements, demonstrating the feasibility and potential benefits of such a mission for atmospheric sensing. The study in [9] proposes a distributed formation control algorithm for CubeSats in Sun-synchronous orbit to achieve coordinated observation missions for Earth observation applications.

### III. THEORETICAL PERSPECTIVE

**CubeSat Swarms:** CubeSats, standardized small satellites with a typical size of 10 cm x 10 cm x 10 cm (1U), have gained significant traction in the space industry due to their low cost, rapid development cycles, and versatility. The concept of CubeSat swarms offers several advantages over traditional single-satellite missions. Firstly, swarms can provide redundancy and resilience, mitigating the impact of individual satellite failures. Secondly, they enable distributed sensing and sampling, facilitating higher spatial and temporal resolution for Earth observation and scientific research. Additionally, CubeSat swarms offer flexibility in mission design, allowing for adaptive and dynamic coverage of targeted areas.

Coordinating the motion and operation of CubeSat swarms presents unique challenges, including formation flying, relative navigation, and communication between satellites. Advanced algorithms and control strategies are employed to maintain desired formations and trajectories while avoiding collisions and maximizing mission efficiency.

**Sun-Synchronous Orbits (SSO):** Sun-synchronous orbits (SSOs) are a class of near-polar orbits commonly used for Earth observation missions. SSOs maintain a fixed angle between the orbital plane and the Sun, resulting in consistent lighting conditions over the Earth's surface. This characteristic enables satellites to pass over specific locations on the Earth's surface at approximately the same local solar time during each orbit.

The synchronization of orbital parameters with the Sun's position ensures consistent illumination and minimizes variations in lighting conditions between successive passes. This consistency is particularly valuable for applications such as remote sensing, where uniform lighting enhances the quality and comparability of acquired data.

**Integration of CubeSat Swarms in SSOs:** The integration of CubeSat swarms in Sun-synchronous orbits offers synergistic advantages for Earth observation missions. By leveraging the flexibility and modularity of CubeSats, swarms can be tailored to meet specific mission requirements and optimize coverage and revisit times over targeted areas.

Varying the RAAN of individual CubeSats within the swarm allows for distributed ground tracks and coordinated coverage of the Earth's surface. By carefully selecting RAAN

values, it is possible to achieve non-overlapping revisit times and maximize the temporal resolution of observations.

**Orbital Parameters:** SSOs are typically characterized by specific orbital parameters, including altitude, inclination, and right ascension of the ascending node (RAAN). Altitude determines the orbital period and coverage area, while inclination dictates the angle of the orbital plane relative to the Earth's equator. RAAN defines the orientation of the orbit in space, with different values resulting in variations in ground track and coverage pattern.

- **Altitude:** This is defined as the distance between the satellite and the Earth's surface along the line perpendicular to the Earth's surface. Altitude determines the period of the orbit and the coverage area of the satellite. CubeSats in Sun-synchronous orbits typically operate at altitudes ranging from 500 km to 800 km to achieve optimal coverage and revisit times.
- **Inclination:** This is defined as the angle between the orbital plane and the plane of the Earth's equator. Inclination determines the orientation of the orbit relative to the Earth's surface. For Sun-synchronous orbits, inclinations are typically near-polar, ensuring consistent lighting conditions over the Earth's surface.

$$i = \arccos \frac{h_z}{|h|} \quad (1)$$

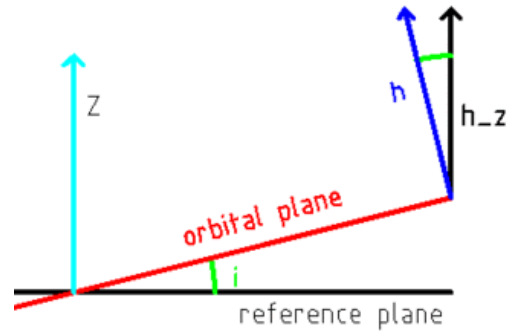


Fig. 1. Finding Inclination (Credit: Wikipedia)

- **Right Ascension of the Ascending Node (RAAN):** This is defined as the angle measured along the celestial equator from the vernal equinox to the ascending node, where the orbit crosses the equatorial plane from south to north. RAAN determines the orientation of the orbital plane in space. Varying RAAN values for CubeSat swarms enable distributed ground tracks and coordinated coverage of targeted areas.

$$\frac{d\Omega}{dt} = \frac{-3}{2} J_2 \left( \frac{R_E}{a(1-e^2)} \right)^2 \sqrt{\frac{\mu_E}{a^3}} \cdot \cos(i) \quad (2)$$

$$\frac{d\Omega}{dt} = \text{Nodal Precession} \left( \frac{\text{rad}}{\text{s}} \right)$$

$$J_2 = J_2 \text{ constant} (1.08262668 \cdot 10^{-3})$$

$$R_E = \text{Radius of Earth} (6378 \text{ km})$$

$$a = \text{Semi-Major Axis}$$

$$e = \text{Eccentricity}$$

$$\mu_E = \text{Standard Gravitational Parameter of Earth}$$

$$i = \text{Orbit Inclination (rad)}$$



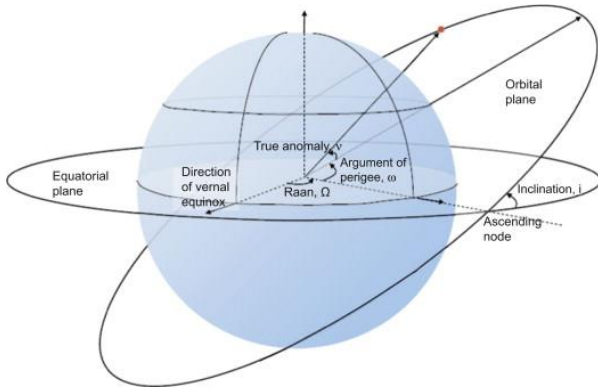


Fig. 2. Orbital elements (Credit: ScienceDirect.com)

- **Argument of Periapsis:** This is defined as the angle measured from the ascending node to the periapsis, the point in the orbit closest to the Earth. Argument of periapsis defines the orientation of the orbit within the orbital plane. While less critical for Sun-synchronous orbits, it can affect the shape and orientation of the ground track.

$$\omega = \arccos\left(\frac{n \cdot e}{|n| |e|}\right) \quad (3)$$

where:  $n$  is the unit vector pointing towards the ascending node of the orbit;  $e$  is the eccentricity vector of the orbit  $\arccos$  is the inverse cosine function.

**Eccentricity:** This is defined as the ratio of the distance between the foci of the orbit to the length of the major axis. Sun-synchronous orbits are typically near-circular (eccentricity close to zero) to maintain consistent lighting conditions and orbital stability.

$$e = \sqrt{1 - \left(\frac{b}{a}\right)^2} \quad (4)$$

where:  $e$  is the eccentricity of the orbit;  $a$  is the semi-major axis of the orbit;  $b$  is the semi-minor axis of the orbit.

For a circular orbit, the semi-major axis ( $a$ ) and semi-minor axis ( $b$ ) are equal, so the eccentricity is zero, indicating a perfect circle.

- **Orbital Period:** This is defined as the time taken for a satellite to complete one orbit around the Earth. Orbital period determines the frequency of revisits over specific geographic locations. For Sun-synchronous orbits, the orbital period is typically synchronized with the Earth's rotation to maintain consistent lighting conditions.

$$T^2 = \frac{4\pi^2}{G \cdot M} a^3 \quad (5)$$

Where:

- $T$  is the orbital period (in seconds),
- $G$  is the gravitational constant ( $6.67430 \times 10^{-11} \text{ m}^3 \text{ kg}^{-1} \text{ s}^{-2}$ ),
- $M$  is the mass of the central body (in kilograms), and
- $a$  is the semi-major axis of the orbit (in meters).

- **Revisit Time:** This is defined as the time interval between successive passes of the satellite over a

specific geographic location. Revisit time is a critical parameter for Earth observation missions, determining the frequency of data acquisition over targeted areas. Optimizing revisit times ensures timely and comprehensive coverage for monitoring and analysis purposes.

#### IV. PROPOSED METHODOLOGY AND SIMULATION RESULTS

The proposed methodology, implemented using Python, is shown in the flow chart in Fig. 3.

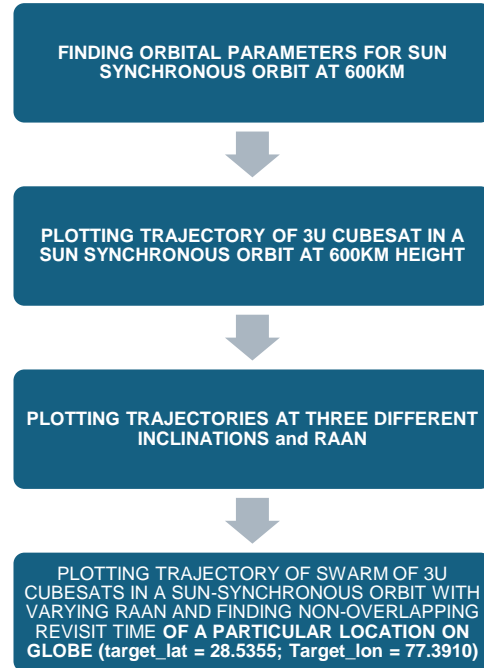


Fig. 3. Flow chart of the proposed methodology

- **FINDING ORBITAL PARAMETERS FOR SUN SYNCHRONOUS ORBIT AT 600 KM**

The implementation of this step starts by defining constants related to Earth's properties, such as its radius and gravitational parameter. The 'sun\_sync\_inclination' function is defined to calculate the inclination required for a Sun-synchronous orbit at a given semi-major axis. The altitude of the Sun-synchronous orbit is specified (600km), and the semi-major axis is calculated by adding the Earth's radius to the altitude. Other standard orbital parameters, including eccentricity, right ascension of the ascending node (RAAN), argument of periapsis, and true anomaly, are set to default values.

**Orbital Parameters for a Sun-synchronous orbit at 600 km altitude:**

**Semi-major axis (a): 6978.00 km**

**Eccentricity (e): 0.00**

**Inclination (i): 97.79 degrees**

**Right Ascension of the Ascending Node (RAAN): 0.00 degrees**

**Argument of Periapsis: 0.00 degrees**

**True Anomaly: 0.00 degrees**

- PLOTTING TRAJECTORY OF 3U CUBESAT IN A SUN SYNCHRONOUS ORBIT AT 600 KM HEIGHT

The Earth's radius and gravitational parameter were defined as constants. Additionally, the second zonal harmonic coefficient ( $J_2$ ) was included to account for the Earth's oblateness. Apart from calculating inclination and the orbital parameters such as eccentricity, right ascension of the ascending node (RAAN) and argument of periapsis, a function was developed to convert the orbital elements (semi-major axis, eccentricity, inclination, RAAN, argument of periapsis, and true anomaly) to Cartesian coordinates ( $x, y, z$ ). This conversion allows for the visualization of the satellite's trajectory in three-dimensional space. The trajectory of the CubeSat was generated by iterating over a range of true anomalies and computing the corresponding Cartesian coordinates. The trajectory represents the path traced by the CubeSat as it orbits the Earth. The generated trajectory was plotted on a three-dimensional globe using Matplotlib. Earth's surface was represented as a sphere, and the CubeSat's trajectory was overlaid on it. This visualization facilitates the interpretation of the satellite's orbit in the context of the Earth's geography.

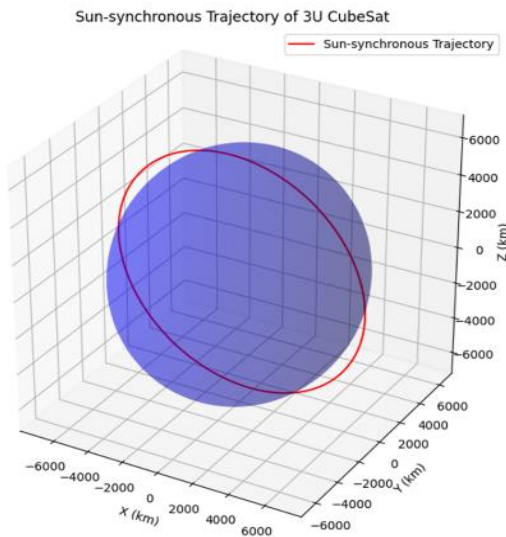


Fig. 4. Trajectory Of 3u CubeSat in a Sun Synchronous Orbit At 600km Height

- PLOTTING TRAJECTORIES AT THREE DIFFERENT INCLINATIONS and RAAN

The trajectories were plotted for different values of inclinations ( $97.6^\circ, 98.7^\circ$  and  $99.8^\circ$ ) and RAAN (Case 1: 5 cubesats at  $0.00^\circ, 2.50^\circ, 5.00^\circ, 7.50^\circ, 10.00^\circ$ ; Case 2: 3 cubesats at  $0^\circ, 120^\circ$  and  $240^\circ$ )

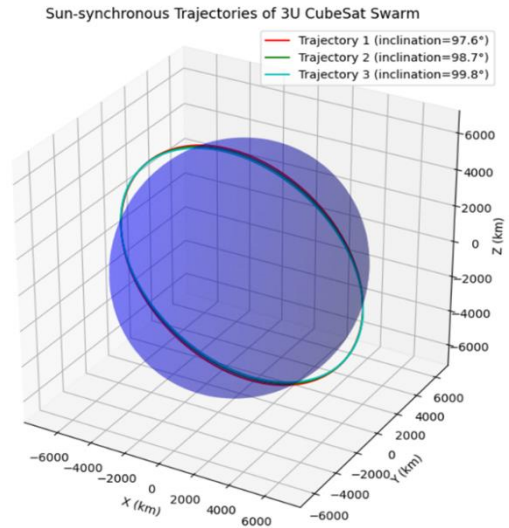


Fig. 5. Sun Synchronous trajectories of 3U CubeSats at different inclinations

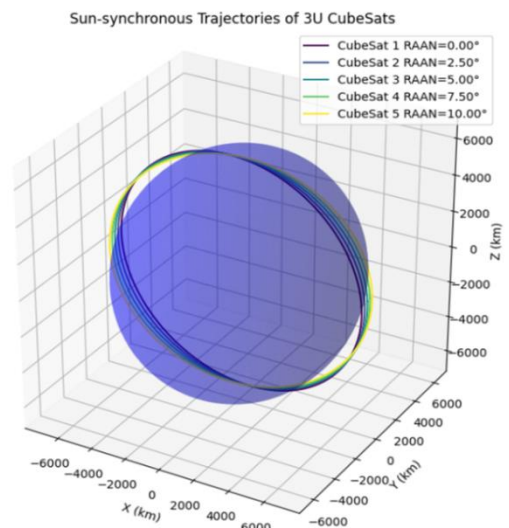


Fig. 6. Sun Synchronous trajectories of 3U CubeSats at different RAAN values (Case 1)

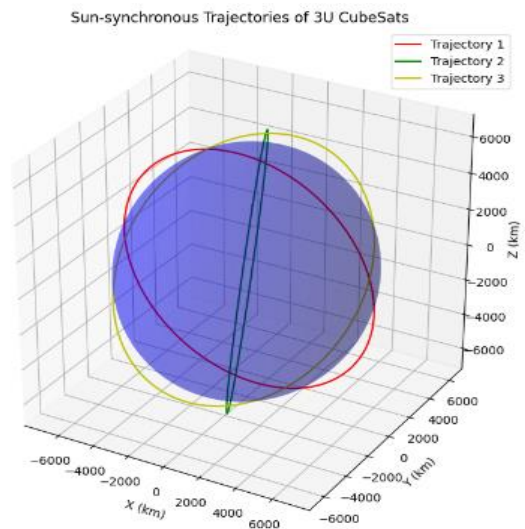


Fig. 7. Sun Synchronous trajectories of 3U CubeSats at different RAAN values (Case 2)

- PLOTTING TRAJECTORY OF SWARM OF 3U CUBESATS IN A SUN-SYNCHRONOUS ORBIT WITH VARYING RAAN AND FINDING NON-OVERLAPPING REVISIT TIME OF A PARTICULAR LOCATION ON GLOBE (target\_lat = 28.5355; Target\_lon = 77.3910)

This step involves determining the revisit times of a particular location on the Earth's surface by a swarm of CubeSats in a sun-synchronous orbit with varying Right Ascension of Ascending Node (RAAN). After calculating the semi-major axis of the sun-synchronous orbit 600km altitude above the Earth's surface, determining the inclination of the sun-synchronous orbit, setting the eccentricity of the orbit to zero (since it's a circular orbit) and taking different RAAN values, the number of points to sample along the orbit are specified. Finally, an array of time points and trajectory for each cubesat is generated by looping through the varying values of RAAN. After specifying the target location (latitude, longitude) on the Earth's surface, and setting a tolerance level to consider a CubeSat revisiting the target location, the revisit times for each CubeSat trajectory is calculated.

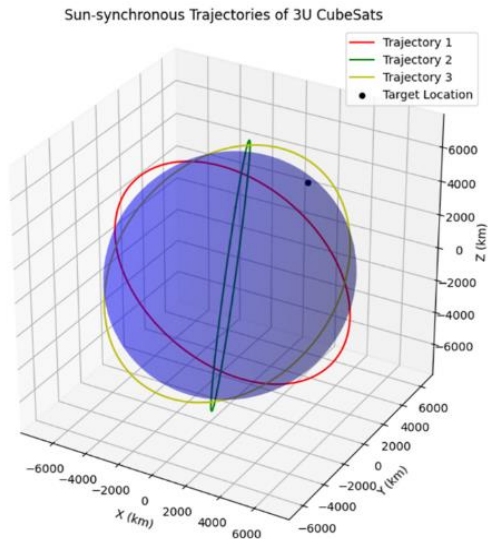


Fig. 8. Plotting Trajectory Of Swarm Of 3U Cubesats in a Sun-Synchronous Orbit with varying RAAN and finding Non-Overlapping Revisit time of a particular location on Globe (Target\_Lat = 28.5355; Target\_Lon = 77.3910)

**Tolerance=72**

Revisit times for CubeSat 1 (seconds): [5103.538587401902, 5115.163960038353, 5126.7893326748035, 5138.414705311255, 5150.040077947705, 5161.665450584156, 5173.290823220606, 5184.916195857058, 5196.541568493509, 5208.166941129959, 5219.79231376641, 5231.41768640286]

Revisit times for CubeSat 2 (seconds): [0.0, 11.625372636450802, 23.250745272901604, 34.8761179093524, 46.50149054580321, 58.12686318225401, 69.7522358187048, 81.37760845515561, 93.00298109160642, 104.62835372805722, 116.25372636450803, 127.8790990095882, 139.5044716374096, 151.1298442738604, 162.75521691031122, 174.38058954676202, 186.00596218321283, 197.63133481966364, 209.25670745611444, 220.88208009256525, 232.50745272901605, 244.13282536546683, 255.75819800191763, 267.38357063836844, 279.0089432748192, 290.63431591127005, 302.2596885477208, 313.88506118417166, 325.51043382062244,

337.1358064570733, 348.76117909352405, 360.3865517299749, 372.01192436642566, 383.63729700287644, 395.26266963932727, 406.88804227577805, 418.5134149122289, 430.13878754867966, 441.7641601851305, 453.38953282158127, 465.0149054580321, 476.6402780944829, 488.26565073093366, 499.8910233673845, 511.51639600383527, 523.141768640286, 534.7671412767369, 546.3925139131877, 558.0178865496384, 569.6432591860893, 581.2686318225401, 592.8940044589909, 604.5193770954417, 616.1447497318925, 627.7701223683433, 639.3954950047942, 651.0208676412449, 662.6462402776957, 674.2716129141465, 685.8969855505973, 697.5223581870481, 709.1477308234989, 720.7731034599498, 732.3984760964005, 744.0238487328513, 755.6492213693022, 767.2745940057529, 778.899666422037, 790.5253392786545, 802.1507119151054, 813.7760845515561, 825.4014571880069, 837.0268298244578, 848.6522024609085, 860.2775750973593, 871.9029477338102, 883.528320370261, 895.1536930067117, 906.7790656431625, 918.4044382796134, 930.0298109160642, 941.6551835525149, 953.2805561889658, 964.9059288254166, 976.5313014618673, 988.1566740983181, 999.782046734769, 1011.4074193712198, 1023.0327920076705, 1034.6581646441214, 1046.283537280572, 1057.908909917023, 1069.5342825534738, 1081.1596551899245, 1092.7850278263754, 1104.4104004628261, 1116.0357730992769, 1127.6611457357278, 1139.2865183721785, 1150.918910086295, 1162.5372636450802, 1174.162636281531, 1185.7880089179819, 1197.4133815544326, 1209.0387541908833, 1220.6641268273343, 1232.289499463785, 1243.9148721002357, 1255.5402447366866, 1267.1656173731374, 1278.7909900095883, 1290.416362646039, 1302.0417352824898, 1313.6671079189407, 1325.2924805553914, 1336.9178531918421, 1348.543225828293, 1360.1685984647438, 1371.7939711011945, 1383.4193437376455, 1395.0447163740962, 1406.670089010547, 1418.2954616469979, 1429.9208342834486, 1441.5462069198995, 1453.1715795563503, 1464.796952192801, 1476.422324829252, 1488.0476974657026, 1499.6730701021534, 1510.30701021534, 1511.63960038353, 15126.7893326748035, 15161.665450584156, 15173.290823220606, 5184.916195857058, 5196.541568493509, 5208.166941129959, 5219.79231376641, 5231.41768640286, 5243.043059039312, 5254.668431675763, 5266.293804312213, 5277.919176948664, 5289.5445495851145, 5301.169922221566, 5312.795294858017, 5324.420667494467, 5336.046040130918, 5347.671412767369, 5359.29678540382, 5370.922158040271, 5382.547530676721, 5394.172903313172, 5405.798275949623, 5417.423648586074, 5429.049021222524, 5440.674393858975, 5452.299766495426, 5463.925139131877, 5475.550511768328, 5487.175884404778, 5498.801257041229, 5510.4266296776805, 5522.052002314131, 5533.677374950582, 5545.302747587032, 5556.928120223483, 5568.5534928599345, 5580.178865496385, 5591.804238132836, 5603.429610769286, 5615.054983405737, 5626.680356042188, 5638.305728678639, 5649.93110131509, 5661.55647395154, 5673.1818465879915, 5684.807219224442, 5696.432591860893, 5708.057964497344, 5719.683337133794, 5731.3087097702455, 5742.934082406696, 5754.559455043147, 5766.184827679598, 5777.810200316048, 5789.4355729525, 5801.06094558895]

Revisit times for CubeSat 3 (seconds): [1453.1715795563503, 1464.796952192801, 1476.422324829252, 1488.0476974657026, 1499.6730701021534, 1511.2984427386043, 1522.923815375055, 1534.5491880115057, 1546.1745606479567, 1557.7999332844074, 1569.4253059208584, 1581.050678557309, 1592.6760511937598, 1604.3014238302108, 1615.9267964666615, 1627.5521691031122, 1639.1775417395631, 1650.8029143760139, 1662.4282870124646, 1674.0536596489155, 1685.6790322853662, 1697.304404921817, 1708.929777558268, 1720.5551501947186, 1732.1805228311696, 1743.8058954676203, 1755.431268104071, 1767.056640740522, 1778.6820133769727, 1790.3073860134234, 1801.9327586498744, 1813.558131286325, 1825.1835039227758, 1836.8088765592267, 1848.4342491956775, 1860.0596218321284, 1871.6849944685791, 1883.3103671050299, 1894.9357397414808, 1906.5611123779315, 1918.1864850143822, 1929.8118576508332, 1941.437230287284, 1953.0626029237346, 1964.6879755601856, 1976.3133481966363,

1987.938720833087, 1999.564093469538, 2011.1894661059887,  
 2022.8148387424396, 2034.4402113788904, 2046.065584015341,  
 2057.690956651792, 2069.3163292882427, 2080.9417019246935,  
 2092.567074561144, 2104.1924471975954, 2115.817819834046,  
 2127.443192470497, 2139.0685651069475, 2150.6939377433982,  
 2162.319310379849, 2173.9446830163, 2185.570055652751,  
 2197.1954282892016, 2208.8208009256523, 2220.446173562103,  
 2232.0715461985537, 2243.696918835005, 2255.3222914714556,  
 2266.9476641079063, 2278.573036744357, 2290.198409380808,  
 2301.823782017259, 2313.4491546537097, 2325.0745272901604,  
 2336.699899926611, 2348.325272563062, 2359.9506451995126,  
 2371.5760178359637, 2383.2013904724145, 2394.826763108865,  
 2406.452135745316, 2418.0775083817666, 2429.702881018218,  
 2441.3282536546685, 2452.9536262911192, 2464.57899892757,  
 2476.2043715640207, 2487.8297442004714, 2499.4551168369226,  
 2511.0804894733733, 2522.705862109824, 2534.3312347462747,  
 2545.9566073827255, 2557.5819800191766, 2569.2073526565273,  
 2580.832725292078, 2592.458097928529, 2604.0834705649795,  
 2615.70884320143, 2627.3342158378814, 2638.959588474332,  
 2650.584961110783, 2662.2103337472336, 2673.8357063836843,  
 2685.4610790201355, 2697.086451656586, 2708.711824293037,  
 2720.3371969294876, 2731.9625695659383, 2743.587942202389,  
 2755.2133148388402, 2766.838687475291, 2778.4640601117417,  
 2790.0894327481924, 2801.714805384643, 2813.340178021094,  
 2824.965550657545, 2836.5909232939957, 2848.2162959304464,  
 2859.841668566897, 2871.467041203348, 2883.092413839799,  
 2894.71778647625, 2906.3431591127005, 2917.968531749151,  
 2929.593904385602, 2941.2192770220527, 2952.844649658504,  
 2964.4700222949546, 2976.0953949314053, 2987.720767567856,  
 2999.3461402043067, 3010.971512840758, 3022.5968854772086,  
 3034.2222581136593, 3045.84763075011, 3057.473003386561,  
 3069.0983760230115, 3080.7237486594627, 3092.3491212959134,  
 3103.974493932364, 3115.599866568815, 3127.2252392052656,  
 3138.8506118417167, 3150.4759844781674, 3162.101357114618,  
 3173.726729751069, 3185.3521023875196, 3196.9774750239703,  
 3208.6028476604215, 3220.228220296872, 3231.853592933323,  
 3243.4789655697737, 3255.1043382062244, 3266.729710842675,  
 3278.3550834791263, 3289.980456115577, 3301.6058287520277,  
 3313.2312013884784, 3324.856574024929, 3336.4819466613803,  
 3348.107319297831, 3359.732691934282, 3371.3580645707325,  
 3382.983437207183, 3394.608809843634, 3406.234182480085,  
 3417.859555116536, 3429.4849277529866, 3441.1103003894373,  
 3452.735673025888, 3464.361045662339, 3475.98641829879,  
 3487.6117909352406, 3499.2371635716913, 3510.862536208142,  
 3522.4879088445928, 3534.113281481044, 3545.7386541174947,  
 3557.3640267539454, 3568.989399390396, 3580.614772026847,  
 3592.240144663298, 3603.8655172997487]

**Tolerance=12**

Revisit times (seconds): [2243.696918835005, 2255.3222914714556,  
 2266.9476641079063, 2278.573036744357, 2290.198409380808,  
 2301.823782017259, 2313.4491546537097, 2325.0745272901604,  
 2336.699899926611]

**V. CONCLUSION**

In this research paper, the trajectory of a swarm of 3U CubeSats in a sun-synchronous orbit with varying Right Ascension of the Ascending Node (RAAN) has been investigated and analyzed the non-overlapping revisit times over a specific target location. Sun-synchronous orbits are of particular interest due to their ability to maintain a fixed angle with respect to the Sun, allowing for consistent lighting conditions and optimal observation opportunities for Earth observation missions.

In conclusion, our work underscores the significance of considering orbital dynamics and parameter variations in the design and deployment of CubeSat constellations for Earth observation missions. By strategically manipulating orbital elements such as RAAN, we can enhance the effectiveness and efficiency of CubeSat swarms in acquiring timely and valuable data for a wide range of scientific, environmental, and commercial applications. Our findings contribute to advancing the field of small satellite mission planning and optimization, paving the way for future innovations in space-based remote sensing and monitoring technologies.

**REFERENCES**

- [1] Malkin, T. L., et al. "Advances in CubeSat Constellations for Earth Observation." *Journal of Small Satellites*, vol. 8, no. 2, 2019, pp. 369-384.
- [2] Lee, D., et al. "Formation Flying and Swarm Control for CubeSats: A Survey." *Journal of Aerospace Engineering*, vol. 30, no. 2, 2017, pp. 04016044.
- [3] Bristow, C. S., et al. "A Survey of CubeSat Constellation Concepts." *Acta Astronautica*, vol. 170, 2020, pp. 536-547.
- [4] Chobotov, V. A. *Sun-Synchronous Orbit Design and Analysis*. AIAA Education Series, 2004.
- [5] Bretagnon, P., et al. "Sun-Synchronous Orbits: Definition, Properties and Missions." *Celestial Mechanics and Dynamical Astronomy*, vol. 85, no. 1, 2003, pp. 1-18.
- [6] Sankaranarayanan, K., et al. "Sun-Synchronous Orbit: Design, Coverage, and Availability." *Journal of Spacecraft and Rockets*, vol. 56, no. 3, 2019, pp. 870-880.
- [7] Garrigues, L., et al. "Optimization of CubeSat Swarm Configurations in Sun-Synchronous Orbits for Earth Observation." *IEEE Transactions on Aerospace and Electronic Systems*, vol. 57, no. 6, 2021, pp. 4563-4576.
- [8] Xu, S., et al. "Design and Analysis of a Sun-Synchronous CubeSat Constellation for Global Navigation Satellite System Radio Occultation." *Journal of Geophysical Research: Atmospheres*, vol. 125, no. 11, 2020, e2020JD033438.
- [9] Li, Z., et al. "Distributed Formation Control of CubeSats in Sun-Synchronous Orbit for Earth Observation." *IEEE Transactions on Aerospace and Electronic Systems*, vol. 54, no. 2, 2018, pp. 932-944.

# Performance Evaluation of Ramanujan Transform with OMP and BP Algorithms

Shivani Verma

*Amity Institute of Space Science & Technology,  
Amity University,  
Noida, India*

M. S. Prasad

*Amity Institute of Space Science & Technology,  
Amity University  
Noida, India*  
<https://orcid.org/0000-0001-9057-9631>

Olga I. Bureneva

*Department of Computer Science and Engineering  
Saint Petersburg Electrotechnical University "LETI"  
Saint Petersburg, Russia*  
<https://orcid.org/0000-0002-6169-454X>

Yulia A. Shichkina

*Department of Computer Science and Engineering  
Saint Petersburg Electrotechnical University "LETI"  
Saint Petersburg, Russia*  
<https://orcid.org/0000-0001-7140-1686>

**Abstract**—In this study, we explore the application of compressed sensing techniques to Synthetic Aperture Radar (SAR) signal reconstruction using the Ramanujan Transform. SAR signal processing is critical for various remote sensing applications, necessitating efficient and accurate reconstruction methods. We implement two reconstruction algorithms: Orthogonal Matching Pursuit (OMP) and Basis Pursuit (BP), both leveraging the sparse representation capabilities of the Ramanujan Transform. The SAR signals are initially transformed using the Ramanujan Transform to obtain a sparse representation, which is then processed using OMP and BP for signal reconstruction. We evaluate the performance of both methods by comparing their Mean Squared Error (MSE) in reconstructing the original SAR signals. Our results demonstrate that while both methods are effective, there are notable differences in their reconstruction accuracy and computational efficiency. This comparative analysis provides valuable insights into the suitability of OMP and BP for SAR signal reconstruction, highlighting the potential of the Ramanujan Transform in enhancing compressed sensing techniques for remote sensing applications.

**Keywords**—Ramanujan Transform, Orthogonal Matching Pursuit, Basis Pursuit, Reconstruction, SAR

## I. INTRODUCTION

Synthetic Aperture Radar (SAR) imaging has emerged as a pivotal technology in remote sensing, offering high-resolution images independent of weather conditions and daylight. SAR images are invaluable for a range of applications including environmental monitoring, earth observation, and military surveillance. However, the voluminous nature of SAR data poses significant challenges for storage, transmission, and processing, necessitating the development of effective image compression and reconstruction techniques.

In recent years, sparse representation theory has gained traction as a powerful tool for signal and image processing. Sparse representation leverages the inherent sparsity of signals in certain transform domains, enabling efficient compression and accurate reconstruction. Two prominent algorithms in sparse representation are Orthogonal Matching Pursuit (OMP) and Basis Pursuit (BP).

Orthogonal Matching Pursuit (OMP) is a greedy algorithm that iteratively selects the dictionary atoms that best correlate with the residual signal. OMP is renowned for its computational efficiency and simplicity, making it suitable for real-time applications. Despite its heuristic nature, OMP often yields satisfactory results in various practical scenarios, including SAR image reconstruction.

Basis Pursuit (BP), on the other hand, formulates the sparse representation problem as a convex optimization problem, typically solved using L1-norm minimization. BP aims to find the sparsest solution by balancing the fidelity to the original signal and the sparsity of the representation. Although BP is computationally more intensive than OMP, it often provides superior reconstruction quality, especially in cases where the signal is highly sparse.

The Ramanujan Transform has recently been explored for its potential in signal processing. The Ramanujan Transform is based on Ramanujan sums, which exhibit unique properties that can be leveraged for analyzing periodic components in signals. The transform has shown promise in various domains, including image processing, due to its ability to capture intricate patterns and structures.

This paper investigates the reconstruction of SAR images using OMP and BP algorithms within the framework of sparse representation. Furthermore, the application of the Ramanujan Transform in conjunction with these algorithms

to enhance the reconstruction process has been explored. The integration of the Ramanujan Transform with OMP and BP aims to exploit the complementary strengths of these methods, potentially leading to improved reconstruction performance.

The remainder of this paper is organized as follows: Section 2 provides a review of the relevant literature work. Section 3 gives the theoretical perspective, providing an overview of SAR imaging, the Ramanujan Transform and the theoretical foundations of sparse representation algorithms, OMP and BP. Section 4 presents the proposed methodology, simulation results, comparing the performance of the proposed methods on synthetic SAR dataset. Finally, Section 5 concludes the paper with a discussion on the findings and future research directions.

## II. LITERATURE REVIEW

OMP is a greedy algorithm used for sparse approximation. It iteratively selects atoms from a dictionary that best match the residual part of the signal. OMP is popular due to its simplicity and efficiency. Zhang et al in [1] introduced the concept of Matching Pursuit (MP) and its variant, OMP. The paper also Highlighted the effectiveness of OMP in decomposing signals into linear combinations of waveforms. Tropp et al in [2] provided a rigorous theoretical foundation for OMP demonstrating its performance in various signal reconstruction scenarios. Tropp in [3] analyzed the convergence properties and performance bounds of OMP and Showed that OMP can recover sparse signals effectively under certain conditions.

Chen et al in [4] Introduced the Basis Pursuit framework and demonstrated its effectiveness in signal and image processing applications. Donoho et al in [5] discussed the stability and robustness of BP in noisy environments and showed BP's superiority over traditional methods in achieving sparse representations. Candes et al in [6] linked BP to the theory of compressed sensing and highlighted the advantages of BP in reconstructing sparse signals from a small number of measurements. R.G. Meditch in [7] introduced Ramanujan sums in the context of Fourier analysis and highlighted their potential in signal processing. Gopalan et al in [8] developed the theoretical foundations of RT and explored its applications in signal processing and analysis. Shanbhag et al in [9] investigated the sparse representation capabilities of RT and demonstrated its effectiveness in compressing and reconstructing signals. [10] discusses the efficiency of OMP for group sparse recovery, extending traditional OMP to handle group structures. The authors prove that OMP can achieve instance optimality under the Restricted Isometry Property (RIP) and show that it is robust under various noise scenarios and sparsity levels. [11] focusses on the efficiency of BP in solving large-scale optimization problems related to sparse recovery. It discusses the implementation of BP in various domains, highlighting its robustness and ability to recover sparse signals accurately even with fewer measurements. [12] explores the application of BP in denoising signals, showing how it can be extended to various contexts such as image processing and compressed sensing. The authors emphasize BP's ability to provide sparse solutions that are both accurate and computationally feasible. [13] explores research on the application of RT in signal processing, highlighting its unique properties that make it

suitable for analyzing periodic and quasi-periodic signals. The paper demonstrates the transform's efficiency in compressing and reconstructing signals with high accuracy. [14] introduces a new framework for signal analysis using Ramanujan filter banks, showing improved performance over traditional Fourier and wavelet transforms in specific applications such as biomedical signal processing and communication systems.

## III. THEORETICAL PERSPECTIVE

### 3.1 SYNTHETIC APERTURE RADAR (SAR) IMAGING

Synthetic Aperture Radar (SAR) is a form of radar technology that is used to create high-resolution images of landscapes. Unlike traditional optical imaging systems that rely on visible light, SAR can operate in all weather conditions and during both day and night, making it extremely valuable for various applications, including military surveillance, environmental monitoring, and geological mapping.

The core principle behind SAR involves using the movement of the radar antenna over a target area to simulate a large antenna aperture. This synthetic aperture creates the high-resolution images that are the hallmark of SAR technology. Fig. 1 shows step-by-step breakdown of how SAR works:



Fig. 1. Working of SAR

**Radar Signal Transmission:** A radar system on an aircraft or satellite transmits a series of microwave pulses towards the ground. These pulses are characterized by their wavelength, frequency, and polarization. The transmitted pulses travel to the ground, where they interact with the surface and other objects.

**Signal Reflection and Reception:** The transmitted radar pulses are reflected back to the radar antenna from the ground and objects. The reflected signals, or echoes, are received by the radar system. The time delay between transmission and reception of each pulse is used to determine the distance to the target (range).

**Antenna Movement and Synthetic Aperture:** As the aircraft or satellite moves, the radar antenna captures multiple echoes from each point on the ground. By combining these echoes, the radar system synthesizes a larger effective aperture. This synthetic aperture improves the azimuth (along-track) resolution of the radar image, enabling the creation of detailed images. The length of this synthetic aperture determines the azimuth resolution, which can be much finer than the physical antenna's resolution.

As the radar platform moves, the frequency of the received signal varies due to the Doppler effect. This frequency shift is used to determine the relative motion of the radar and the target, which is crucial for azimuth compression and image formation.

Signal Processing: The received radar echoes are processed using various signal processing techniques to form a high-resolution image. This involves:

- **Range Compression:** Enhances the range resolution by processing the received signals to distinguish between closely spaced targets along the range direction.
- **Azimuth Compression:** Enhances the azimuth resolution by combining the echoes received at different positions of the moving antenna.

Matched filtering techniques are employed to maximize the signal-to-noise ratio (SNR) of the received echoes. This is essential for accurately resolving the target features in both range and azimuth directions.

**Image Formation:** The processed signals are then converted into a two-dimensional image, with one dimension representing the range and the other representing the azimuth. The intensity of each pixel in the image corresponds to the strength of the radar echo from that point on the ground.

A SAR radar emits pulses using a linear FM chirp signal. This signal is defined by (1).

$$(\tau)=rect(\tau/T_r) \times \cos(2\pi f_0\tau + \pi K_r\tau^2) \quad (1)$$

where  $K_r$  is the chirp rate,  $f_0$  is the carrier frequency, and  $\tau$  is the range time. The pulse envelope can be approximated using a rectangular function,  $re(\tau T_r)$ , where  $T_r$  is the pulse duration. The received signal, or echo, is given by the convolution of the pulse and the ground reflectivity,  $g_r$ , in the illuminated section,

$$(\tau)=g_r(\tau) \otimes s_t(\tau) \quad (2)$$

Considering the time delay between the emission and reception of the pulse, given by  $2R_t/c$ , where  $R_t$  is the distance from the antenna to the target and  $A_0'$  is the magnitude, the received signal is given by

$$(\tau)=A_0'rect(\tau-2R_t/c) \times \cos[2\pi f_0(\tau-2R_t/c) + \pi K_r(\tau-2R_t/c)^2 + \phi] \quad (3)$$

where  $\phi$  represents the phase change that may result from the scattering due to the roughness of the reflecting surface. Since  $R_t$  changes with azimuth time, it is referred to now as  $(\eta)$ . Before sampling the signal, it is necessary to remove the radar carrier,  $\cos(2\pi f_0\tau)$ . This is done with a demodulation process, resulting in the following demodulated signal from a single point,

$$(\tau,\eta)=A_0'rect(\tau-2R_t(\eta)/c)w_a(\eta-\eta_c) \times \exp[-j4\pi f_0R_t(\eta)/c] \times \exp[j\pi K_r(\tau-2R_t(\eta)/c)^2] \quad (4)$$

where  $w_a$  is the antenna pattern in the azimuth direction,  $\eta$  is the azimuth time and  $A_0$  is a complex constant,  $A_0 = A_0' \exp(i\phi)$ .

### 3.2 RAMANUJAN TRANSFORM

The Ramanujan transform is a relatively new mathematical tool developed for signal processing and analysis, inspired by the work of the renowned Indian mathematician Srinivasa Ramanujan. It is particularly useful for analyzing periodic and quasi-periodic signals.

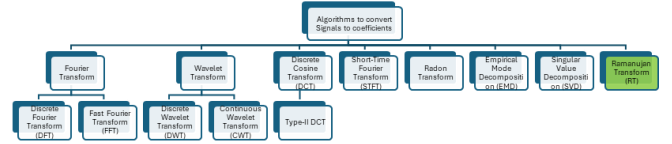


Fig. 2. Different Signal Processing Algorithms

The Ramanujan Transform (RT) offers the following advantages over other transforms, particularly in applications involving the analysis of periodicities and sparse representation of signals.

- **Intrinsic Periodicity Extraction:** RT is specifically designed to identify and extract periodic components in signals. Unlike Fourier Transform, which represents the signal in terms of sinusoidal basis functions, RT uses Ramanujan sums that are well-suited for identifying exact periodicities.
- **Sparse Representation:** RT provides a sparse representation of signals with strong periodic components, meaning that the transformation often results in a small number of non-zero coefficients, which is beneficial for compressed sensing and signal compression.
- **Robustness to Noise:** RT is particularly effective in noisy environments. Since it focuses on periodic components, random noise (which typically lacks a periodic structure) has less impact on the transform coefficients.
- **High Resolution:** RT can resolve closely spaced periodicities better than other transforms like the Discrete Fourier Transform (DFT), making it advantageous in applications where distinguishing between closely spaced periodic components is crucial.
- **Adaptability to Non-Stationary Signals:** RT is well-suited for non-stationary signals where the periodic components may change over time. This makes it a useful tool in fields like biomedical signal processing (e.g., ECG, EEG) where signal characteristics can vary.
- **Efficient Computation:** For signals with strong periodic components, RT can be computed more efficiently compared to other transforms that might require more complex operations or produce less sparse representations.
- **Compressed Sensing:** The sparse representation provided by RT aligns well with the principles of compressed sensing, allowing for efficient reconstruction of signals from fewer samples than traditional methods would require.

The Ramanujan transform provides a unique and efficient way to represent signals using Ramanujan sums, which are sums of exponential functions with periods that are related to the divisors of integers.

**Ramanujan Sums:** Ramanujan sums,  $c_q(n)$ , are defined as  $\sum_{\substack{k=1 \\ \gcd(k,q)=1}}^q e^{2\pi i \frac{kn}{q}}$  where:

- $q$  is a positive integer.
- $k$  ranges over integers from 1 to  $q$  that are coprime with  $q$  (i.e.,  $\gcd(k,q)=1$ )
- $n$  is an integer.
- $e^{2\pi i \frac{kn}{q}}$  represents the complex exponential function.

Ramanujan sums are known for their orthogonality and periodicity properties, making them useful for signal decomposition.

**Ramanujan Space:** The Ramanujan space for a given period  $q$ , denoted as  $R_q$ , is the space of all sequences that can be represented using Ramanujan sums with period  $q$ . For a signal  $x(n)$ , the Ramanujan space decomposition allows representing  $x(n)$  as a sum of components from different Ramanujan spaces.

**Ramanujan Transform:** The Ramanujan transform of a signal  $x(n)$  decomposes it into components corresponding to different periods  $q$ . The transform coefficients,  $X_q(n)$ , are calculated using the inner product of the signal with the Ramanujan sums.

**Forward Ramanujan Transform:** Given a signal  $x(n)$ , the Ramanujan transform  $R_x(q)$  for a period  $q$  is  $R_x(q) = \sum_{n=0}^{N-1} x(n) \cdot c_q(n)$ , where  $N$  is the length of the signal  $x(n)$ .

**Inverse Ramanujan Transform:** To reconstruct the original signal from its Ramanujan transform, the inverse Ramanujan transform is used,  $x(n) = \sum_{q=1}^Q R_x(q) \cdot c_q(n)$  where  $Q$  is the maximum period considered for the decomposition.

Fig. 3 shows the steps followed in Ramanujan Transform.

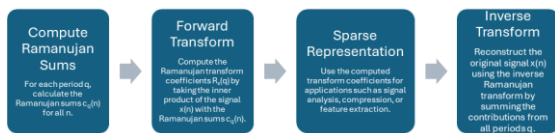


Fig. 3. Steps followed in Ramanujan Transform

### 3.3 SPARSE REPRESENTATION OF SAR DATA

Compressed sensing (CS) is particularly relevant in SAR to reduce the amount of data collected while still allowing accurate image reconstruction. CS theory states that if a signal is sparse in some domain, it can be recovered from far fewer samples than traditionally required, using  $y=Ax$  where  $A$  is a measurement matrix. The signal  $x$  can be recovered by solving  $\min_x \|x\|_1$  subject to  $y=Ax$ .

Fig. 4 shows the basic steps involved in Compressed Sensing and Reconstruction of SAR Data.



Fig. 4. Basic steps involved in Compressed Sensing and Reconstruction of SAR Data

- **Dictionary Learning:** Learn a dictionary  $D$  from SAR training data using methods like K-SVD.

- **Sparse Coding:** Encode new SAR data  $y$  using the learned dictionary by solving  $\min_x \|x\|_1$  subject to  $y \approx Dx$ .
- **Compression:** Store or transmit only the non-zero coefficients and their indices.
- **Reconstruction:** Reconstruct the SAR image by multiplying the dictionary with the sparse coefficients.

Sparse representation of Synthetic Aperture Radar (SAR) data refers to expressing radar signals or images in a domain where the majority of the signal coefficients are zero or close to zero. This representation leverages the fact that SAR images often contain significant amounts of structure and redundancy, meaning that they can be efficiently described using a few non-zero coefficients in an appropriate basis or dictionary. Sparse representation enables efficient data compression, which is crucial for reducing storage and transmission costs, particularly for large-scale SAR datasets. By focusing on the most significant components of the signal, sparse representation can enhance signal-to-noise ratio, improving image quality.

There are several algorithms designed to achieve sparse representation of SAR data. These algorithms can be broadly categorized based on their approaches and the nature of the optimization problems they solve.

Fig. 5 shows the different algorithms used for Sparse representation of SAR signals.

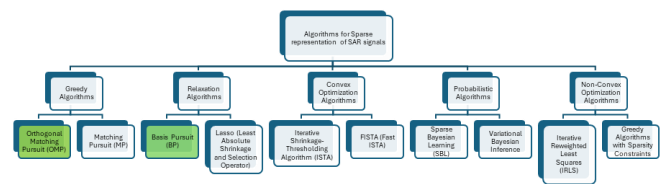


Fig. 5. Algorithms used for Sparse representation of SAR signals.

#### 3.3.1 ORTHOGONAL MATCHING PURSUIT (OMP) ALGORITHM

The Orthogonal Matching Pursuit (OMP) algorithm is a greedy algorithm used for sparse approximation. Its primary goal is to find the best sparse representation of a signal using a dictionary of basis functions. OMP is particularly useful in compressive sensing and signal processing due to its efficiency and simplicity.

Steps of OMP Algorithm:

Consider the following sparse representation problem:  $y=D \cdot x+e$  where  $y$  is the observed signal (vector of length  $n$ ),  $D$  is the dictionary matrix (size  $n \times m$ ), where each column is a basis function,  $x$  is the sparse coefficient vector (vector of length  $m$ ), which is required to be computed and  $e$  is the error or noise vector. The goal of OMP is to find a sparse vector  $x$  such that  $y \approx D \cdot x$ .

- **Initialization:**
  - Initialize the residual  $r=y$  (the part of  $y$  that has not been explained by the selected basis functions).
  - Initialize the set of selected indices  $\Lambda=\emptyset$ .



- Initialize the sparse coefficient vector  $x=0$ .

Find the Best Matching Atom:

- Compute the inner products between the residual  $r$  and all columns of the dictionary  $D$ .
- Select the index  $j$  of the column (atom) that has the maximum absolute inner product with the residual.  $j=\arg \max_k |\langle D_k, r \rangle|$
- Add the index  $j$  to the set of selected indices  $\Lambda$ .

Update the Coefficient Vector:

- Solve a least squares problem to find the new coefficients for the selected atoms.
- Denote  $D_\Lambda$  as the submatrix of  $D$  containing only the columns indexed by  $\Lambda$ .
- Solve for the coefficients  $x_\Lambda : x_\Lambda=\arg \min_z \|y-D_\Lambda z\|_2$
- Update  $x$  such that the coefficients corresponding to  $\Lambda$  are  $x_\Lambda$  and the rest are zero.

Update the Residual:

- Update the residual  $r$  by subtracting the contribution of the selected atoms:  $r=y-D_\Lambda x_\Lambda$

- **Termination:** Repeat the iterative selection step until one of the stopping criteria is met:

- The residual  $rrr$  is sufficiently small (i.e.,  $\|r\|_2$  is below a threshold).
- A predefined number of iterations have been reached.
- The desired sparsity level (number of non-zero elements in  $x$ ) is achieved.

OMP is easy to implement and understand, and is computationally efficient, especially for problems where the number of non-zero coefficients (sparsity level) is low.

### 3.3.2 BASIS PURSUIT (BP) ALGORITHM:

The Basis Pursuit (BP) algorithm is a method for solving the sparse representation problem. It aims to find the sparsest solution to a linear system of equations. Unlike greedy algorithms like Orthogonal Matching Pursuit (OMP), Basis Pursuit relies on convex optimization, specifically linear programming, to find the sparse solution.

Considering the same sparse representation problem as mentioned in section 3.3.1, the goal is to find the sparsest vector  $x$  that satisfies the equation above.

Sparsity here means that the solution  $x$  has as many zeros as possible, i.e., only a few of its components are non-zero. The  $\ell_0$ -norm directly counts the number of non-zero entries in  $x$ . Minimizing the  $\ell_0$ -norm would be the ideal approach for enforcing sparsity. However, the  $\ell_0$ -norm minimization problem is non-convex and combinatorial, making it computationally infeasible to solve for large-scale problems. The  $\ell_1$ -norm serves as the closest convex relaxation of the  $\ell_0$ -norm. Minimizing the  $\ell_1$ -norm has been shown to encourage sparsity in the solution. This is because the  $\ell_1$ -norm penalizes all non-zero entries linearly, effectively reducing the number of non-zero components in the optimal solution.

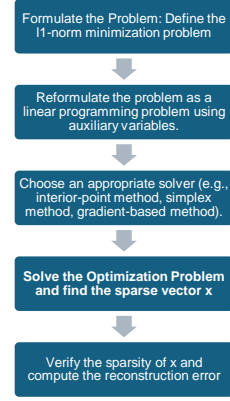


Fig. 6. Steps of BP Algorithm

## IV. PROPOSED METHODOLOGY

This research proposes a methodology for compressing and reconstructing Synthetic Aperture Radar (SAR) signals using the Ramanujan Transform. The reconstruction is performed using two different sparse representation algorithms: Orthogonal Matching Pursuit (OMP) and Basis Pursuit (BP). The performance of both methods is evaluated and compared based on their reconstruction accuracy, measured by the Mean Squared Error (MSE).

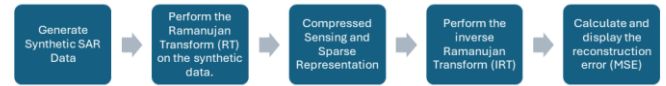


Fig. 7. Proposed Methodology

**Synthetic SAR Signal Generation:** To validate the proposed methodology, a synthetic SAR signal is generated. The synthetic signal is a sum of multiple sinusoidal components to simulate the periodic nature of real SAR data. The signal is defined as follows:

$$x(n) = \sum_{i=1}^k A_i \sin(2\pi f_i n + \varphi_i) + \eta(n)$$

where  $A_i, f_i$  and  $\varphi_i$  are the amplitude, frequency, and phase of the  $i$ -th sinusoidal component, respectively, and  $\eta(n)$  is additive white Gaussian noise. The signal length  $N$  is set to 512 samples for computational efficiency.

**Ramanujan Transform:** The Ramanujan Transform decomposes the SAR signal into periodic components. The transform coefficients for different periods  $q$ , are computed using Ramanujan sums.

**Compressed Sensing and Sparse Representation:** The compressed sensing framework involves sampling the SAR signal at a lower rate than the Nyquist rate and reconstructing it using sparse representation algorithms. The sparse representation is achieved using two methods: OMP and BP.

**Inverse Ramanujan Transform:** After obtaining the sparse coefficients from OMP or BP, the inverse Ramanujan Transform reconstructs the signal

**Performance Evaluation:** The reconstructed signals from both OMP and BP are compared with the original synthetic SAR signal. The performance metric used for comparison is the Mean Squared Error (MSE), defined as:

$$MSE = \frac{1}{N} \sum_{n=0}^{N-1} (x(n) - \widehat{x(n)})^2,$$

where  $\widehat{x(n)}$  is the reconstructed signal.

**Simulation Results:** The simulation results after process of 128x128 and 512x512 images with Ramanujan Transform and OMP Algorithm are shown in Fig. 8 and Fig. 9. The results for images of the same sizes with Ramanujan Transform and BP Algorithm are shown in Fig. 10 and Fig. 11.

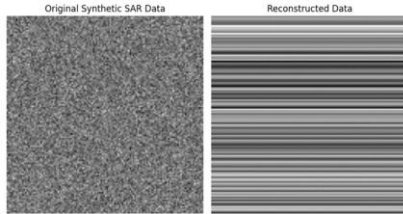


Fig. 8. Simulation Results for Ramanujan Transform with OMP for 128x128 image (MSE=206722.678419)

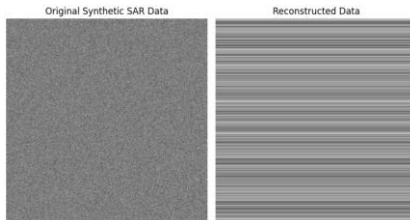


Fig. 9. Simulation Result for Ramanujan with OMP for 512x512 image (MSE= 827061.796474)



Fig. 10. Simulation Result for Ramanujan with BP for 128X128 image (MSE=225550.831061)

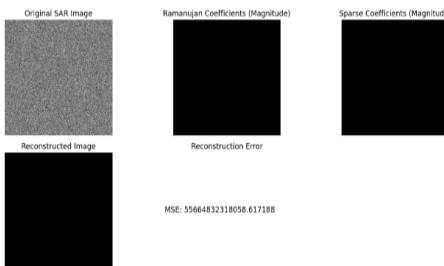


Fig. 11. Simulation Result for Ramanujan with BP for 512X512 image (MSE=55664832318058.61)

Table 1 gives a comparison of the MSE values obtained after Simulation.

TABLE I. SIMULATION RESULTS

Synthetic Image size	Ramanujan Transform + OMP Algorithm, MSE	Ramanujan Transform + Basis Pursuit Algorithm, MSE
128X128	206722.678419	225550.831061
512X512	827061.796474	55664832318058.61

As shown by Table 1, the MSE obtained by using Ramanujan Transform with Basis Pursuit Algorithm is much higher than Ramanujan Transform used with OMP algorithm.

**Interpretation of Results:** From a mathematical perspective, the MSE is defined as:

- $MSE = \frac{1}{N} \sum_{n=0}^{N-1} (x(n) - \widehat{x(n)})^2$  where  $x(n)$  are the original pixel values and  $\widehat{x(n)}$  are the reconstructed pixel values. As  $n$  (the number of pixels) increases, the impact of errors in accumulates, leading to higher MSE.
- OMP achieves lower MSE compared to BP in the context of SAR signal reconstruction when using RT due to its greedy nature that effectively captures significant signal features with lower computational overhead. BP, while theoretically providing a sparse optimal solution, might suffer in practical scenarios due to its strict sparsity enforcement and sensitivity to regularization parameters. BP's global optimization can sometimes get stuck in local minima or take a longer path to converge to the solution, potentially increasing MSE in finite iterations.

## V. CONCLUSION

This study provides a comprehensive approach to evaluating the effectiveness of the Ramanujan Transform in compressive sensing and sparse reconstruction of SAR signals using OMP and BP algorithms. By comparing the MSE, the study aims to determine the most efficient algorithm for SAR signal reconstruction.

## REFERENCES

- [1] Mallat, S.G. and Zhang, Z. (1993) Matching Pursuit with Time-Frequency Dictionaries. *IEEE Transactions on Signal Processing*, 41, 3397-3415.
- [2] Tropp, J. A. (2004). Greed is Good: Algorithmic Results for Sparse Approximation. *IEEE Transactions on Information Theory*, 50(10), 2231-2242. doi:10.1109/TIT.2004.834793.
- [3] Tropp, J. A., & Gilbert, A. C. (2007). Signal Recovery from Random Measurements via Orthogonal Matching Pursuit. *IEEE Transactions on Information Theory*, 53(12), 4655-4666. doi:10.1109/TIT.2007.909108.
- [4] Chen, S. S., Donoho, D. L., & Saunders, M. A. (1998). Atomic Decomposition by Basis Pursuit. *SIAM Journal on Scientific Computing*, 20(1), 33-61. doi:10.1137/S1064827596304010.
- [5] Donoho, D. L., Elad, M., & Temlyakov, V. N. (2006). Stable Recovery of Sparse Overcomplete Representations in the Presence of Noise. *IEEE Transactions on Information Theory*, 52(1), 6-18. doi:10.1109/TIT.2005.860430.
- [6] Candes, E. J., & Wakin, M. B. (2008). An Introduction To Compressive Sampling. *IEEE Signal Processing Magazine*, 25(2), 21-30. doi:10.1109/MSP.2007.914731.
- [7] Meditch, J. S. (1972). The Ramanujan sums: A novel approach to Fourier analysis. *IEEE Transactions on Computers*, 21(9), 1090-1096. doi:10.1109/TC.1972.5009074.
- [8] Gopalan, R., & Agarwal, P. (2010). The Ramanujan-Fourier Transform: Theory and Applications. *IEEE Transactions on Signal Processing*, 58(10), 5110-5122. doi:10.1109/TSP.2010.2055887.
- [9] Shanbhag, U. V., Banerjee, A., & Vaidyanathan, P. P. (2017). Sparse Signal Representation using Ramanujan Sums. 2017 IEEE International Symposium on Circuits and Systems (ISCAS), 1-4. doi:10.1109/ISCAS.2017.8050560.

- [10] Hekmati, A., & Zahiri, S. H. (2021). Group Sparse Recovery Using OMP. *MDPI Axioms*, 10(3), 33.
- [11] Chen, S. S., Donoho, D. L., & Saunders, M. A. (2001). Atomic decomposition by basis pursuit. *SIAM Journal on Scientific Computing*, 20(1), 33-61.
- [12] Figueiredo, M. A. T., Nowak, R. D., & Wright, S. J. (2007). Gradient projection for sparse reconstruction: Application to compressed sensing and other inverse problems. *IEEE Journal of Selected Topics in Signal Processing*, 1(4), 586-597.
- [13] Anand, S., & Soman, K. P. (2021). Ramanujan Subspace Decomposition for Sparse Signal Representation. *arXiv preprint arXiv:2106.00124*.
- [14] Tsai, C. Y., & Lin, Y. (2019). Ramanujan Filter Banks: Construction and Application to Periodic Signal Processing. *IEEE Transactions on Signal Processing*, 67(12), 3203-3218.

# Оценка влияния снижения размерности пространства признаков на эффективность классификации движений на основе поверхностной электромиографии

И. В. Кузнецов, Ю. В. Пономарчук

Дальневосточный государственный университет путей сообщения

kyzaiv1@yandex.ru

**Аннотация.** Одной из актуальных задач классификации движений на основе электромиографии является выбор наиболее информативного набора признаков, характеризующего движения. Увеличение числа признаков может привести к уменьшению скорости обучения классификаторов и снижению эффективности классификации жестов. В работе исследуются возможности повышения эффективности решения задачи классификации движений при использовании методов снижения размерности пространства признаков. Были рассмотрены классификаторы на основе метода опорных векторов (Support Vector Machines, SVM), решающих деревьев (Decision Trees, DT), случайного леса (Random Forest, RF) и K-ближайших соседей (KNN). В качестве метода снижения размерности был использован метод главных компонент (Principal Component Analysis, PCA). Достигнуто повышение эффективности классификации на 3% при использовании PCA и KNN.

**Ключевые слова:** классификация движений; электромиография; метод главных компонент; снижение размерности; машинное обучение; SVM; DT; RF; KNN

## I. ВВЕДЕНИЕ

Электромиография (ЭМГ) – метод регистрации разности потенциалов, возникающих при активации скелетных мышц – является одним из перспективных источников управляющего сигнала для реализации человеко-машинных интерфейсов (ЧМИ). ЭМГ-сигнал собирается при разработке ЧМИ поверхностными электродами и представляет собой результат интерференции множества ЭМГ-сигналов отдельных двигательных единиц (ДЕ).

Для определения жестов, способных выступать в роли команд для конечных устройств, из ЭМГ-сигнала может быть извлечено большое число характеристик. Последние, как правило, выступают в роли признаков классов жестов и используются для обучения классификаторов на основе методов машинного обучения и глубокого обучения.

Признаки, которые могут быть извлечены из ЭМГ-сигнала, относятся к временной (Time Domain, TD), частотной (Frequency Domain, FD) или частотно-временной (Time-Frequency Domain, TFD) областям. TD-признаки характеризуются простотой вычисления, а потому часто используются [1]. Недостатком TD-

признаков является изменчивость их статистических свойств из-за нестационарности ЭМГ-сигнала [2, 3]. В свою очередь, применение FD-признаков осложняется большой чувствительностью ЭМГ-сигнала к шуму [4].

Перспективным направлением исследований является анализ TFD-признаков: считается, что они позволяют избавиться от ограничений частотной и временной пространств [4, 5], недостатком же является большая размерность пространства признаков, что ограничивает качество обучения классификаторов. При этом невозможно гарантировать, что все извлекаемые TFD-признаки достаточно информативны.

Для повышения эффективности классификации жестов применяются методы снижения размерности, в частности, метод главных компонент. Целью данной работы является исследование эффективности классификации при использовании PCA по сравнению с классификацией без него. Рассматривается предположение об увеличении эффективности классификации при использовании PCA.

## II. МАТЕРИАЛЫ И МЕТОДЫ ИССЛЕДОВАНИЯ

### A. Рассматриваемые классификаторы

Для проверки гипотезы выполнен сравнительный анализ эффективности четырех наборов классификаторов. Каждый содержал четыре классификатора на основе следующих методов:

- машина опорных векторов (SVM);
- решающее дерево (DT);
- случайный лес (RF);
- K-ближайших соседей (KNN).

Во всех наборах классификаторов были использованы одинаковые архитектуры. Первая группа строилась на базе TD-признаков. Вторая группа классификаторов дополнительно использовала TFD-признаки. Для третьей и четвертой групп классификаторов, реализованных аналогичным образом, дополнительно применен PCA для снижения размерности пространства признаков.

Перед обучением классификаторов набор данных был нормализован с использованием  $l_2$  нормы и

стандартизован. Для проведения расчетов использована библиотека Scikit-learn и язык программирования Python.

### B. Используемый набор данных

Для проведения эксперимента использовался NinaPro DB5 – набор данных ЭМГ-сигнала, полученный от 10 субъектов с помощью двух браслетов Thalmic Labs Myo [6]. Движения в нем описываются с помощью данных 16-канальной электромиограммы (по восемь каналов на каждый браслет). В рамках эксперимента для обучения классификаторов не использовалась иная информация (например, данные кинематики движений), представленная в наборе.

В целом, набор данных содержит информацию о пятидесяти двух движениях, из них для классификации выбрано 12, которые представлены на рис. 1.

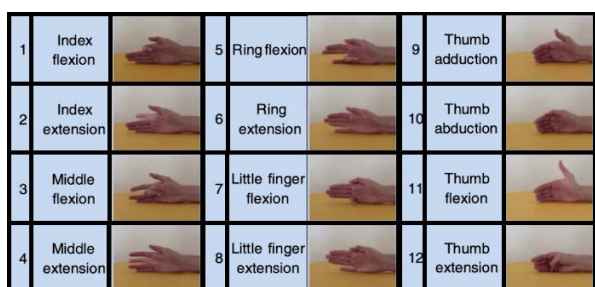


Рис. 1. Движения, выбранные для эксперимента

В ходе эксперимента принималось, что представленная в наборе данных информация не требует дополнительной предобработки.

### C. Признаки временной области

Для описания движений использовались следующие TD-признаки: Zero Crossing (ZC), Variance (Var), Difference Absolute Standard Deviation Value (DASDV).

Zero Crossing представляет собой количество пересечений нулевого уровня ЭМГ-сигналом. Способы вычисления ZC рассмотрены в [7], где указано, что в ходе анализа требуется учитывать влияние шумов. Для вычисления использовались следующие выражения:

$$ZC = \sum_{i=1}^{N-1} ZC(x_i), \quad T = 4 \left( \frac{1}{10} \sum_{j=1}^{10} x_j \right),$$

$$ZC(x_i) = \begin{cases} 1, & (x_i \geq T) \text{ and } (x_{i+1} < T) \\ 1, & (x_i < T) \text{ and } (x_{i+1} \geq T) \\ 0, & \text{в иных случаях} \end{cases}$$

где  $T$  – пороговый коэффициент.

Variance, согласно Spiewak [8], описывает мощность ЭМГ-сигнала. Признак может быть вычислен следующим образом:

$$VAR = \frac{1}{N-1} \sum_{i=1}^N \sqrt{x^2}$$

Признак Mean Absolute Value использовался в [1]. В [8] отмечается, что MAV – это метод определения и

измерения уровня мышечного сокращения, который можно вычислить с помощью следующего выражения:

$$MAV = \frac{1}{N} \sum_{i=1}^N |x_i|$$

Признак Difference Absolute Standard Deviation Value применялся в исследовании Phimyomark [9]. Он представляет собой стандартное абсолютное отклонение для разницы соседних значений:

$$DASDV = \sqrt{\frac{1}{N-1} \sum_{i=1}^{N-1} (x_{i+1} - x_i)^2}$$

Вычисленные признаки использовались во всех группах классификаторов.

### D. Признаки частотно-временной области

Для вычисления признаков частотно-временной области данные ЭМГ-сигнала были подвергнуты дискретному вейвлет-преобразованию с четырехуровневой декомпозицией. В качестве материнского использован вейвлет Добеши пятого порядка. Далее, для полученных коэффициентов вычислены признаки Variance, Root Mean Square (RMS) и Energy Discrete Wavelet Transform (EDWT) [10]:

$$RMS = \sqrt{\frac{1}{N} \sum_{i=1}^N x_i^2}$$

$$EDWT = \sum_{i=1}^N c_i^2$$

где  $c_i$  – полученный в результате вейвлет-преобразования коэффициент.

Представленные признаки использовались в классификаторах третьей и четвертой группы.

## III. РЕЗУЛЬТАТЫ

Для оценки эффективности работы были рассчитаны метрики эффективности Recall и Precision:

$$Precision = \frac{TP}{TP + FP}, \quad Recall = \frac{TP}{TP + FN}$$

где  $TP$  – доля объектов, корректно отнесенных к заданному классу,  $TN$  – доля объектов, корректно не отнесенных к классу,  $FN$  – доля объектов, некорректно не отнесенных к заданному классу,  $FP$  – доля объектов, некорректно отнесенных к заданному классу.

Представленные в наборе классы не сбалансированы: объектов, описывающих состояние покоя значительно больше, нежели объектов жестов. Все классы, несмотря на частоту появления соответствующих объектов, имеют одинаковую важность – при проектировании ЧМИ нельзя считать одну команду более ценной по сравнению с другой. Это требуется учитывать при оценке эффективности классификаторов.

Значения метрик эффективности классификаторов представлены в табл. 1. При расчете метрик использовано макроусреднение.

ТАБЛИЦА 1. МЕТРИКИ ЭФФЕКТИВНОСТИ РАБОТЫ КЛАССИФИКАТОРОВ

	SVM	DT	RF	KNN
<b>Precision(TD)</b>	0.71	0.56	0.77	0.76
<b>Recall(TD)</b>	0.51	0.57	0.69	0.73
<b>Precision(TD+PCA)</b>	0.72	0.58	0.81	0.78
<b>Recall(TD+PCA)</b>	0.52	0.59	0.71	0.77
<b>Precision(TD+TFD)</b>	0.78	0.63	0.81	0.82
<b>Recall(TD+TFD)</b>	0.76	0.62	0.80	0.82
<b>Precision(TD+TFD+PCA)</b>	0.78	0.66	0.83	0.85
<b>Recall(TD+TFD+PCA)</b>	0.76	0.63	0.81	0.85

Матрицы ошибок первой и третьей групп классификаторов для KNN представлены на рис. 2. Можно заметить, что при использовании TD-признаков классификаторы некорректно распознают движения большого пальца, движения мизинца и безымянного пальца, а также указательного пальца. Можно предположить, что причиной этому является использование близкорасположенных двигательных

единиц (ДЕ): в силу того, что в случае поверхностной электромиографии ЭМГ-сигнал представляет собой интерференцию сигналов отдельных ДЕ, классификаторам не удастся различать движения, которые используют близкорасположенные ДЕ, на основе лишь только TD-признаков. Лучшую эффективность показали KNN и RF. Эффективность RF оказалась значительно выше, чем DT. Это позволяет утверждать, что применение ансамблевых методов для классификации движений позволяет повысить эффективность решения задачи классификации.

Применение PCA приводит к увеличению общей эффективности классификации. При этом эффективность классификации не изменяется для большого пальца, падает для безымянного и растет для мизинца и указательного пальца. Это позволяет утверждать, что при использовании PCA теряется часть полезной информации, а также сделать вывод о необходимости уточнения расположения считывающих электродов для увеличения доли успешно классифицированных движений большого пальца.

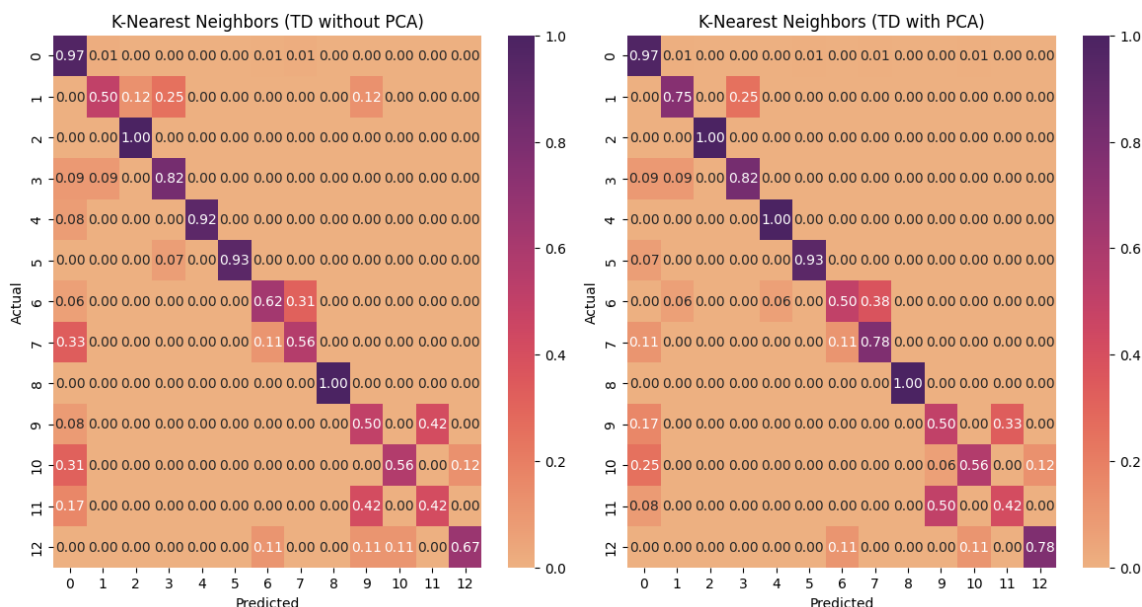


Рис. 2. Матрица ошибок KNN первой и второй групп классификаторов

При использовании признаков TFD дополнительно к TD для всех классификаторов фиксируется повышение эффективности классификации. При дополнительном использовании PCA для классификатора на основе SVM значения Precision и Recall не изменились.

Классификаторы на основе DT и RF значительно увеличили эффективность относительно аналогов, использующих лишь TD-признаки. При использовании PCA для классификаторов с TFD-признаками наблюдается увеличение эффективности классификации.

Наилучшие показатели были достигнуты для KNN и составляют 85% для Precision и Recall при использовании TFD признаков и PCA. Матрица ошибок этого классификатора, а также его аналога, не использующего PCA, представлена на рис. 3. Можно заметить, что увеличение эффективности достигается за счет уменьшения доли ошибочной классификации движений указательного, безымянного пальцев и мизинца. Доля успешной классификации жестов большого пальца, изменилась незначительно.

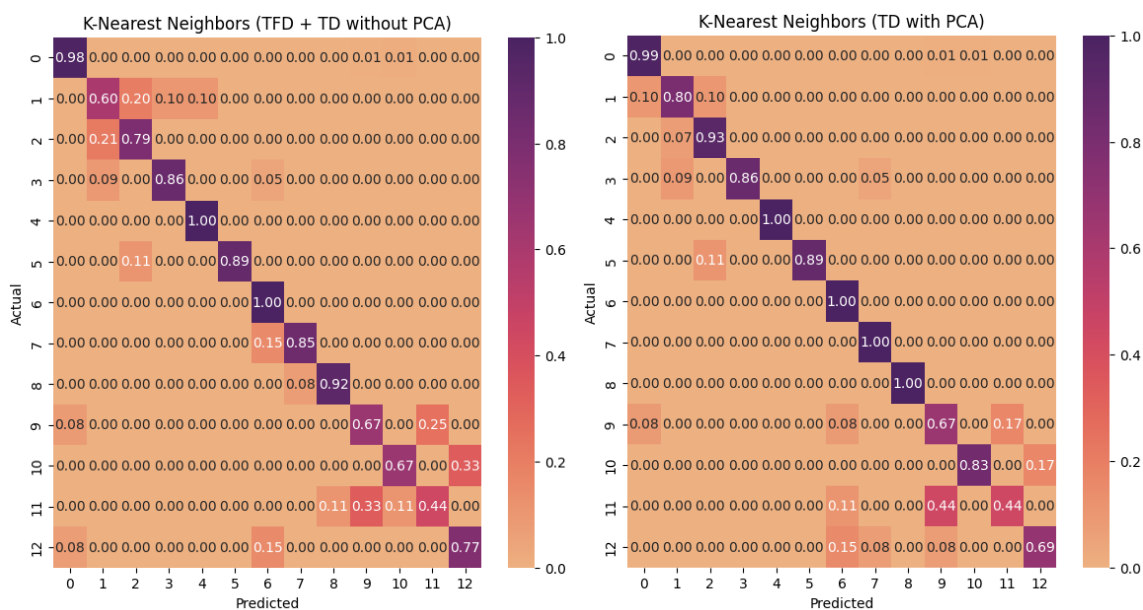


Рис. 3. Матрицы ошибок KNN третьей и четвертой группы классификаторов

Увеличение общей эффективности для KNN при использовании TFD-признаков и PCA составляет порядка 3 % относительно аналога, не использующего PCA, что позволяет утверждать, что применение PCA оправдано для увеличения эффективности классификации движений.

#### IV. ЗАКЛЮЧЕНИЕ

В ходе эксперимента удалось достичь хорошей доли успешной классификации движений мизинца, безымянного, указательного и среднего пальцев. Для большого пальца удалось достичь роста эффективности классификации, однако задача повышения эффективности распознавания жестов остается открытой. В качестве одного из решений предлагается накладывать считывающие электроды с учетом расположения ДЕ, отвечающих за рассматриваемые движения.

Использование PCA позволило увеличить долю успешно классифицированных движений для классификаторов на основе RF, DT и KNN. В работе рассмотрен лишь один из множества возможных наборов признаков, что ставит вопрос о поиске более эффективных наборов.

При этом невозможно гарантировать того, что для новых наборов признаков PCA окажется эффективным в той же мере. Это делает актуальным вопрос поиска других эффективных методов снижения размерности, а также сравнения результатов их применения. Ожидается, что эффективными окажутся методы анализа независимых компонент, а также стохастического вложения соседей с t-распределением.

На основе выполненных исследований также можно выбирать жесты с высокой эффективностью классификации для разработки ЧМИ.

#### СПИСОК ЛИТЕРАТУРЫ

- [1] Abbaspour S. и др. Evaluation of surface EMG-based recognition algorithms for decoding hand movements // *Medical & Biological Engineering & Computing*. 2019. Т. 58. № 1. С. 83–100.
- [2] Karheily S. и др. sEMG time–frequency features for hand movements classification // *Expert Systems with Applications*. 2022. Т. 210. С. 118282.
- [3] Lee K.H., Min J.Y., Byun S. Electromyogram-Based Classification of Hand and Finger Gestures Using Artificial Neural Networks // *Sensors*. 2021. Т. 22. № 1. С. 225.
- [4] Shair E.F., Jamaluddin N.A., Abdullah A.R. Finger Movement Discrimination of EMG Signals Towards Improved Prosthetic Control using TFD // *International Journal of Advanced Computer Science and Applications*. 2020. Т. 11. № 9.
- [5] Nazmi N. и др. A Review of Classification Techniques of EMG Signals during Isotonic and Isometric Contractions // *Sensors*. 2016. Т. 16. № 8. С. 1304.
- [6] Pizzolato S. и др. Comparison of six electromyography acquisition setups on hand movement classification tasks // *PLOS ONE*. 2017. Т. 12. № 10. С. e0186132.
- [7] Toledo-Perez D.C., Rodriguez-Resendiz J., Gomez-Loenzo R.A. A Study of Computing Zero Crossing Methods and an Improved Proposal for EMG Signals // *IEEE Access*. 2020. Т. 8. С. 8783–8790.
- [8] Spiewak C. A Comprehensive Study on EMG Feature Extraction and Classifiers // *Open Access Journal of Biomedical Engineering and Biosciences*. 2018. Т. 1. № 1.
- [9] Phinyomark A., Phukpattaranont P., Limsakul C. Feature reduction and selection for EMG signal classification // *Expert Systems with Applications*. 2012. Т. 39. № 8. С. 7420–7431.
- [10] Ren L., Xu D., Li S. Feature extraction of surface EMG signal // *Вестник Тихоокеанского государственного университета*. 2019. Т. 2. С. 9–16.

# Оценка производительности оптимизатора в сверточных нейронных сетях для классификации микрокальцинатов молочной железы: сравнительное исследование

Э. Ч. Угвеек<sup>1</sup>, Д. Палогиианнидис<sup>2</sup>

Санкт-Петербургский государственный электротехнический университет  
«ЛЭТИ» им. В.И. Ульянова (Ленина)

<sup>1</sup>ugwuekeestherc@gmail.com, <sup>2</sup>dimitris.palogiannidis@gmail.com

**Аннотация.** Оптимизаторы оказывают значительное влияние на глубокое обучение нейронной сети, что влияет на способность сети к обобщению, скорость конвергенции и общую эффективность модели. Выбор наилучшего оптимизатора имеет решающее значение в контексте классификации микрокальцификации молочной железы, где точный диагноз необходим для успешного лечения. В этой работе сверточные нейронные сети (CNN) были использованы для изучения влияния различных оптимизаторов на классификацию микрокальцификаций молочной железы. В частности, китайская база данных маммографии (CMMD) была использована для оценки эффективности известных оптимизаторов, включая Adam, RMSProp, Adadelta и Adagrad, с точки зрения оптимизации архитектуры CNN для точной классификации микрокальцификации. В этой работе набор данных аннотированных маммограмм использовался для обучения нескольких моделей CNN, каждая из которых использовала свой оптимизатор. Были проведены комплексные эксперименты для оценки производительности этих оптимизаторов с точки зрения адаптивности, скорости сходимости и точности классификации. Выводы исследования содержат полезную информацию о выборе оптимизаторов для оптимизации CNN в контексте анализа медицинских изображений, особенно для ранней диагностики рака молочной железы на основе классификации микрокальцификации.

**Ключевые слова:** рак молочной железы, ранняя диагностика, микрокальцинаты, обнаружение, классификация, глубокое обучение, сверточная нейронная сеть, CNN, Adam, Rmsprop, Adadelta, Adagrad

## I. ВВЕДЕНИЕ

Одним из наиболее распространенных видов рака у женщин является рак молочной железы, который часто имеет неблагоприятный прогноз на высоких стадиях диагностики и лечения [1]. Раннее выявление рака молочной железы с помощью маммографии имеет важное значение для снижения смертности и заболеваемости, что является серьезной проблемой глобального здравоохранения. Микрокальцификации на маммограммах проявляются в виде более ярких пятен, чем нормальные ткани молочной железы, и тесно связаны с раком молочной железы [2]. В течение многих лет микрокальцификации были очень полезным

показателем рака молочной железы; фактически, они были ответственны за идентификацию от 30 до 50 % непальпируемых опухолей, обнаруженных во время скрининга. Кроме того, в большинстве случаев протоковой карциномы in situ (DCIS) они присутствуют [3].

Несколько исследований показали связь микрокальцификации и рака молочной железы. Исследование, проведенное [4] показало, что у пациенток с мелкими линейными микрокальцификациями риск местного рецидива был в 5,2 раза выше. Они подтвердили, что эти маммографические характеристики имеют клиническое значение как маркеры более тяжелого заболевания. По данным Ling и соавторов, у женщин с кальцификациями на момент постановки диагноза риск смерти от рака молочной железы был в 2,4 раза выше, а вероятность рецидива – в 2 раза выше [5].

Используя комплексный метод на основе изображений, а не методологии, основанные на фрагментах, использованные в предыдущих исследованиях, Хаки и др. представили CNN для автоматической идентификации микрокальцификации на основе U-net. Они обучили сеть сегментации и оценили результаты, используя показатель среднеквадратичной логарифмической ошибки (MSLE), который количественно определяет разницу между ожидаемыми и фактическими значениями. При использовании предложенного подхода были получены многообещающие результаты в обнаружении микрокальцификации с наилучшим значением потери MSLE, равным 0,05, и точностью 95 % [1].

Кумар Сингх и др. al классифицировал три класса микрокальцификаций с использованием четырех различных оптимизаторов (Adam, Adagrad, Adadelta и RMSProp) на специально отобранном подмножестве изображений молочной железы из набора данных маммограммы DDSM (CBIS-DDSM) с использованием предварительно обученной модели InceptionResnetV2. Чтобы CNN достиг хорошей производительности классификации, его параметры были скорректированы с использованием каждого из этих оптимизаторов. Как следствие, результаты показали, что оптимизатор



ADADelta со скоростью обучения 0,001 имел наибольшую скорость обучения 98 % и точность проверки 94 % [6].

В этой статье демонстрируется использование глубокого обучения для автоматической идентификации и категоризации двух классов микрокальциатов: (0) доброкачественной и (1) злокачественной с использованием предложенной нами архитектуры CNN. Область интереса (ROI), которая состоит из микрокальцификации, аннотируется вручную для получения масок сегментации. Наше исследование направлено на оценку производительности различных оптимизаторов и роли, которую они играют в точности классификации микрокальцификатов.

## II. МАТЕРИАЛЫ И МЕТОДЫ

### A. Материалы

Набор данных CMD, использованный для этого исследования, был собран из Архива изображений рака (TCIA) [7]. Набор данных был дополнительно предварительно обработан [2] для преобразования изображений Digital Imaging and Communications in Medicine (DICOM) в формат Portable Network Graphics (PNG) с сохранением значений пикселей без влияния на разрешение. Новый набор данных состоит из 288 доброкачественных и 1158 злокачественных кальцификаций с разрешением 1914 x 2294.

Чтобы устранить проблему дисбаланса исходных данных, когда было 288 доброкачественных изображений и 1158 злокачественных, был применен простой метод увеличения, включающий поворот на 45°, 90° и 180°, чтобы увеличить количество доброкачественных изображений, в результате чего в общей сложности получилось 1152 доброкачественных изображения. Благодаря получению более сбалансированного набора данных становится проще разделить данные на обучающие и тестовые наборы, и ожидается, что показатели оценки будут давать более точные результаты. На данный момент сбалансированный набор данных состоит в общей сложности из 4540 изображений с соответствующими масками аннотаций сегментации, которые были созданы для разных классов (доброкачественных и злокачественных) с помощью инструмента аннотирования. Набор данных был разделен на 3 части в соотношении 70:15:15 для обучения, тестирования и валидации модели и реализован с использованием библиотеки Keras в Google Colab на персональном компьютере с процессором Intel Core i5 и 24 ГБ оперативной памяти.

Ниже описаны различные этапы разработки методов и концепций кода:

### B. Подготовка данных

На этом этапе код определяет пути к папкам обучения, проверки и тестирования, содержащим изображения различных классов. Метод `flow_from_directory` в `ImageDataGenerator` используется для загрузки изображений и соответствующих меток из предоставленных каталогов,

что позволяет увеличивать данные «на лету» и пакетную загрузку и особенно полезно для больших наборов данных, которые не могут поместиться в память все сразу.

### C. Предварительная обработка данных

Предварительная обработка является важным этапом подготовки данных для моделей машинного обучения. Были определены три отдельных объекта `ImageDataGenerator` для обучения, проверки и тестирования данных. К обучающим данным были применены преобразования изображений, такие как изменение масштаба, сдвиг, масштабирование по горизонтали. Изображения предварительно обрабатываются путем изменения их размера до согласованного размера 256x256 пикселей и нормализации значений пикселей от 0 до 1. Это гарантирует, что все изображения будут иметь одинаковые размеры для ввода в нейронную сеть, и помогает улучшить сходимость во время обучения. Другие преобразования помогают дополнить обучающие данные, делая модель более устойчивой к изменениям во входных изображениях.

### D. Построение модели

Модель CNN построена с использованием контейнера «Sequential». Она начинается со сверточного слоя (`Conv2D`) с 16 фильтрами размером 3x3 и применяет функцию активации ReLU. Далее следует максимальное объединение (`MaxPooling2D`) с размером пула 2x2, выполняющее пространственную нисходящую выборку. Этот процесс повторяется с дополнительными сверточными слоями и слоями максимального объединения для извлечения объектов из входных изображений. Затем карты объектов сглаживаются в вектор и пропускаются через полностью связанные ("плотные") слои из 128 единиц измерения, которые выполняют классификацию на основе извлеченных объектов. Выходной слой, который является конечным плотным слоем, состоит из 2 блоков с функцией активации softmax для обеспечения вероятности класса.

### E. Обучение модели

Модель сначала компилируется с помощью оптимизатора Adam и категориальной функции потерь перекрестной энтропии. Затем выполняется обучение в течение 20 периодов с использованием данных обучения и проверкой производительности на основе данных проверки и размера пакета, равного 32. Оптимизатор Adam обычно используется из-за его возможностей адаптивной скорости обучения, в то время как категориальная потеря перекрестной энтропии подходит для задач классификации по нескольким классам. Затем обучение выполняется путем вызова метода `fit` для объекта модели, передачи генераторам данных обучения и проверки. Метод `fit` запускает цикл обучения в течение указанного количества эпох, обновляя параметры модели, чтобы минимизировать функцию потерь. Другими оптимизаторами, использованными для обучения модели, были RMSProp, Adagrad и Adadelata.

После обучения строится график истории обучения, содержащий значения точности и потерь за разные

эпохи, и сохраняется в формате ".h5" для дальнейшего использования. Сохраненная модель вместе с историей обучения загружается для дальнейшего анализа или вывода. Это позволяет повторно использовать обученную модель для составления прогнозов или дальнейшего анализа без необходимости ее переобучения.

**F. Прогнозирование и оценка модели**

Тестовые изображения и метки считываются из тестовой папки с использованием тех же шагов предварительной обработки, которые применяются к данным обучения и проверки. Затем загружается модель и используется для прогнозирования меток классов для предварительно обработанных тестовых изображений. Прогнозируемые метки кодируются обратно в числовой формат с помощью кодировщика меток. Матрица путаницы генерируется путем сравнения истинных меток с предсказанными метками.

Матрица путаницы визуализируется с помощью тепловой карты, где строки представляют истинные метки, а столбцы представляют прогнозируемые метки. Тепловая карта снабжена комментариями с указанием количества прогнозов каждого класса. Подзаголовки создаются для визуализации точности обучения и потерь при обучении в разные эпохи. Кроме того, создается область под кривой (ROC) для визуализации AUC двух классов.

**III. РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ**

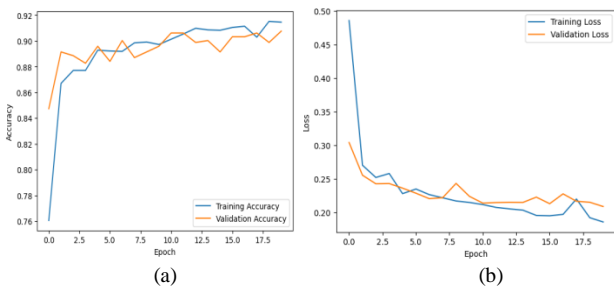


Рис. 1. Обучение и валидация оптимизатора Adam. (a) Точность обучения и валидации оптимизатора Adam. (б) Потеря обучения и валидации оптимизатора Adam

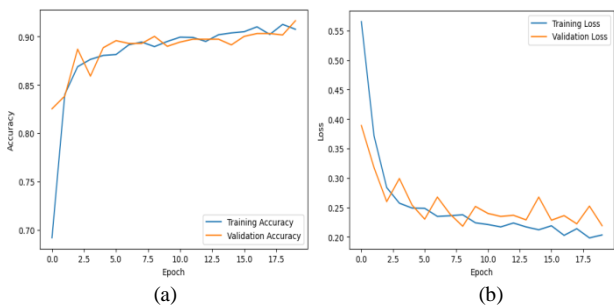


Рис. 2. Обучение и валидация оптимизатора RMSProp. (a) точность обучения и валидации оптимизатора RMSProp. (б) Потеря обучения и валидации оптимизатора RMSProp

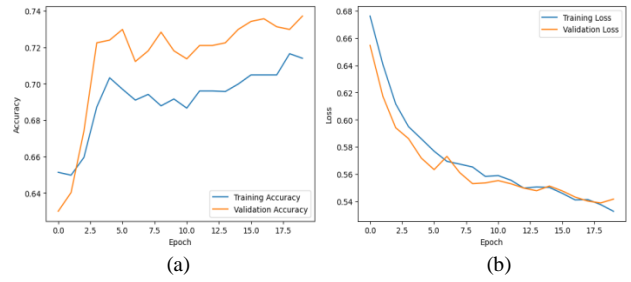


Рис. 3. Обучение и валидация оптимизатора Adagrad. (a) точность обучения и валидации оптимизатора Adagrad. (б) Потеря обучения и валидации оптимизатора Adagrad.

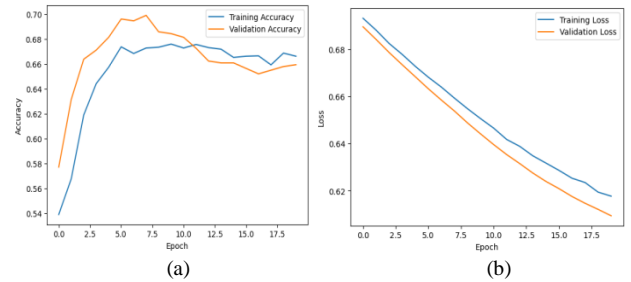


Рис. 4. Обучение и валидация оптимизатора Adadelta. (a) Точность обучения и валидации оптимизатора Adadelta. (б) потеря обучения и валидации оптимизатора Adadelta.

На точность модели нейронной сети во время обучения и валидации влияют различные факторы, включая выбор оптимизатора. В данном случае точность обучения и валидации составила около 0,91 как в оптимизаторах Adam, так и в RMSProp, что значительно выше, чем у других.

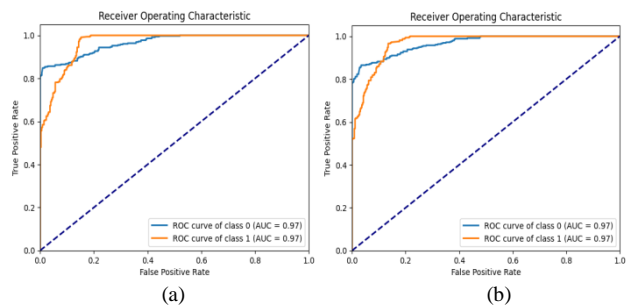


Рис. 5. ROC-кривая (a) ROC-кривая оптимизатора Adam. (б) ROC-кривая оптимизатора RMSProp

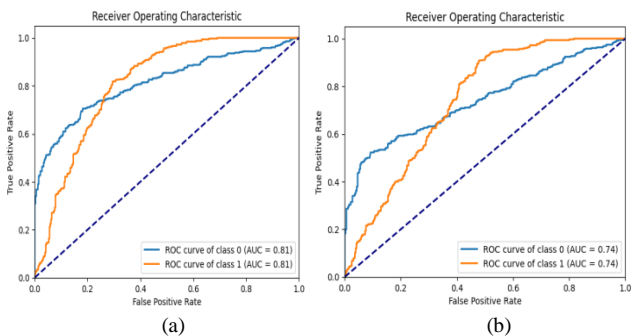


Рис. 6. ROC-кривая (a) ROC-кривая оптимизатора Адаграда. (б) ROC-кривая оптимизатора Ададельта

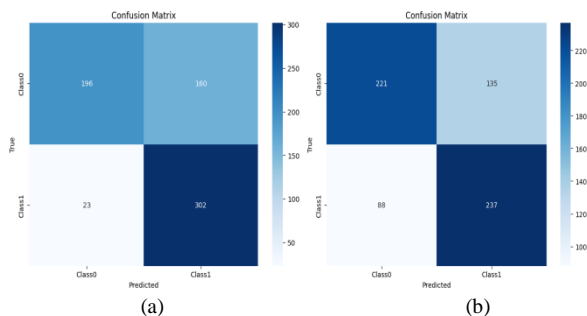


Рис. 7. (a) Матрица путаницы для оптимизатора Adam (b) Матрица путаницы для оптимизатора RMSProp

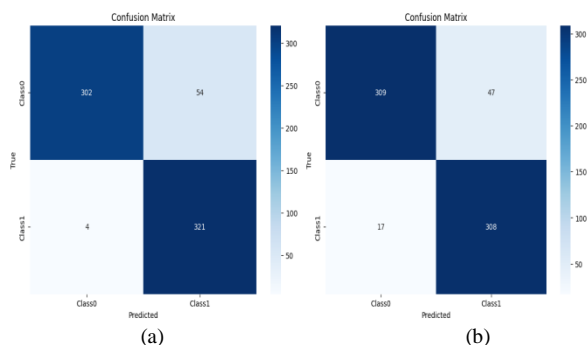


Рис. 8. (a) Матрица путаницы для оптимизатора Adagrad (b) Матрица путаницы для оптимизатора Adadelata

Скорость обучения каждого оптимизатора показала, что общее время обучения составило приблизительно 10224 секунды для Adam, 10014 секунд для RMSProp, 10239 секунд для Adagrad и 10422 секунды для Adadelata. RMSProp продемонстрировал сопоставимую производительность с Adam, потребовав при этом немного меньше времени на обучение.

Результаты тестирования различных оптимизаторов обобщены в таблице 1 на основе нашей сети CNN. Adam optimizer добился наивысшей общей производительности по всем показателям с точностью 91 %, прецизионностью 0,92, отзывом 0,91, оценкой F1 0,91, специфичностью для класса 0 0,86 и специфичностью для класса 1 0,99. Следовательно, Adam будет считаться лучшим выбором, за которым следует RMSProp. Учитывая точность, он дал более высокую точность для обоих классов по сравнению с RMSProp, предполагая, что он может работать немного лучше с точки зрения предотвращения ложных срабатываний.

ТАБЛИЦА I. РЕЗУЛЬТАТЫ ТЕСТИРОВАНИЯ РАЗЛИЧНЫХ ОПТИМИЗАТОРОВ С ОДИНАКОВОЙ СКОРОСТЬЮ ОБУЧЕНИЯ, ПАКЕТАМИ И ЭПОХАМИ

Optimizers	Metrics					
	Accuracy	Precision	Recall	F1 score	Specificity (Class 0)	Specificity (Class 1)
Adam	0.91	0.92	0.91	0.91	0.86	0.99
RMSprop	0.91	0.91	0.91	0.91	0.87	0.95
Adagrad	0.73	0.78	0.73	0.72	0.65	0.89
Adadelata	0.67	0.68	0.67	0.67	0.64	0.72

#### IV. ЗАКЛЮЧЕНИЕ

Основываясь на результатах, хотя Adam и RMSProp достигли одинаковой точности, видно, что Adam optimizer показал лучший показатель точности по сравнению с RMSProp optimizer. На эффективность оптимизаторов, таких как RMSProp и Adam, может повлиять расширение набора данных или количество периодов. Adam работает лучше, чем другие оптимизаторы, когда речь заходит об общей эффективности модели для обнаружения микрокальцификации на маммографических изображениях, согласно сравнению производительности. Он обладает большей способностью точно распознавать как положительные, так и отрицательные примеры. Кроме того, его исключительная способность снижать количество ложноположительных и отрицательных результатов, что имеет решающее значение для правильной диагностики при медицинской визуализации, о чем свидетельствуют его высокие значения специфичности для обоих классов (0,86 для класса 0 и поразительные 0,99 для класса 1).

В то время как Adam был наиболее эффективным оптимизатором в целом для выявления микрокальцификации, когда принимались во внимание показатели производительности и время обучения, интересно отметить, что RMSProp работал аналогично и требовал немного меньше времени на обучение. Это показывает, что RMSProp является надежной заменой, которая обеспечивает эффективные результаты без ущерба для точности или специфичности. Таким образом, при взвешивании компромисса между временем обучения и производительностью, RMSProp представляется хорошим вариантом для максимального использования компьютерных систем обнаружения при диагностике рака молочной железы.

#### СПИСОК ЛИТЕРАТУРЫ

- [1] A.N.R. Hakim, P. Prajitno, and D.S. Soejoko, "Microcalcification detection in mammography image using computer-aided detection based on convolutional neural network," AIP Conference Proceedings, vol. 2346, pp. 040001-1 –040007, March 2021. <https://aip.scitation.org/doi/10.1063/5.0047828>
- [2] E. C. Ugwueke and D. Palogiannidis, "Preparation of Patient Mammogram Images for Segmentation," in 2023 XXVI International Conference on Soft Computing and Measurements (SCM), pp. 305–307, May 2023. doi: 10.1109/SCM58628.2023.10159092.
- [3] S. O’Grady and M. P. Morgan, "Microcalcifications in breast cancer: From pathophysiology to diagnosis and prognosis," Biochim. Biophys. Acta BBA - Rev. Cancer, vol. 1869, pp. 310–320, April 2018. doi: 10.1016/j.bbcan.2018.04.006.
- [4] G. M. Rauch et al., "Microcalcifications in 1657 Patients with Pure Ductal Carcinoma in Situ of the Breast: Correlation with Clinical, Histopathologic, Biologic Features, and Local Recurrence," Ann. Surg. Oncol., vol. 23, pp. 482–489, February 2016. doi: 10.1245/s10434-015-4876-6.
- [5] H. Ling, Z.-B. Liu, L.-H. Xu, X.-L. Xu, G.-Y. Liu, and Z.-M. Shao, "Malignant calcification is an important unfavorable prognostic factor in primary invasive breast cancer," Asia Pac. J. Clin. Oncol., vol. 9, pp. 139–145, June 2013. doi: 10.1111/j.1743-7563.2012.01572.x.
- [6] K. Kumar Singh et al., "Deep Learning Capabilities for the Categorization of Microcalcification," Int. J. Environ. Res. Public Health, vol. 19, pp. 2159, February 2022.
- [7] C. Cui et al., "The Chinese Mammography Database (CMMD): An online mammography database with biopsy confirmed types for machine diagnosis of breast," The Cancer Imaging Archive, 2021

# Применение алгоритма стекинга для идентификации отклонений в данных газотранспортной сети

А. К. Петрова, С. Е. Абрамкин

Санкт-Петербургский государственный электротехнический университет  
«ЛЭТИ» им. В.И. Ульянова (Ленина)

ak72p@yandex.ru, seabramkin@etu.ru

**Аннотация.** Для решения задач идентификации отклонений в данных все чаще применяются алгоритмы машинного обучения. При этом точность алгоритмов различна, и не все алгоритмы одинаково полезны. В ряде задач целесообразно применение ансамблей алгоритмов, среди которых особое место занимает алгоритм стекинга, особенность которого состоит в том, что он может использовать алгоритмы разного типа. В статье рассматривается применение алгоритма стекинга в сравнении с LSTM нейронной сетью.

**Ключевые слова:** классификация, LSTM нейросеть, ансамбли, стекинг, матрица спутанности, гиперпараметры, предсказание

## I. ВВЕДЕНИЕ

Отклонения во временных рядах представляют собой аномалии или аномальные подпоследовательности, которые могут указывать на наличие неисправностей в системе. На основе алгоритмов идентификации аномалий построены методики предсказательной аналитики и систем поддержки принятия решений, мониторинга состояния и раннего оповещения систем реального времени.

Разработано множество методов предиктивного анализа: деревья решений, модель регрессии, искусственная нейронная сеть, байесовская статистика, ансамбль методов, градиентный бустинг, метод опорных векторов, анализ временных рядов, метод k-ближайших соседей, метод главных компонент и другие [1–12].

В статье рассматривается применение классификационных LSTM – нейронных сетей и ансамбля алгоритмов машинного обучения, для классификации данных на предмет наличия или отсутствия отклонений [13].

## II. ПРИМЕНЕНИЕ LSTM – НЕЙРОННЫХ СЕТЕЙ

К статистическим данным по расходу газа Калининградского участка Север-Западной ГТС за 2015 год, при помощи приложения Classification Learner среды Matlab были применены различные алгоритмы машинного обучения, для выбора наилучшего (табл. 1) Для применения алгоритмов данные результатов измерения были размечены.

В качестве вектора входных факторов использовались следующие данные:

- расход с шагом – 1,  $Q_{\text{расх}}$  тыс. м<sup>3</sup>;
- фактор времени (дата, час);
- разметка отклонений.

Результаты сравнения классификационных алгоритмов машинного обучения иллюстрирует табл. 1.

ТАБЛИЦА I. РЕЗУЛЬТАТЫ СРАВНЕНИЯ РЕЗУЛЬТАТОВ КЛАССИФИКАЦИИ АЛГОРИТМАМИ МАШИННОГО ОБУЧЕНИЯ

Алгоритм машинного обучения	Точность (test)
LSTM нейросеть	76%
ИНС прямого распространения	68%
К ближайших соседей	61%
Деревья решений	53%
SVM	45%
Наивный Байес	38%

Сравнение показало, что LSTM нейросеть классифицирует более точно.

Следующим этапом на размеченных данных была обучена классификационная LSTM нейросеть [14–21], данные предварительно были нормированы по математическому ожиданию и среднеквадратическому отклонению. Архитектуру сети иллюстрирует рис. 1.

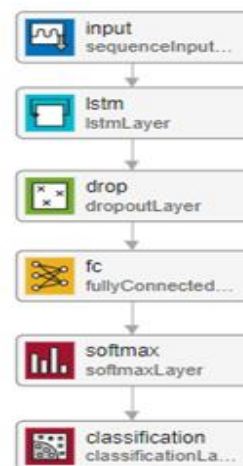


Рис. 1. Архитектура LSTM-сети для детектирования отклонений в данных

Параметры LSTM классификатора:

- первый слой – вектор входных факторов;
- один рекуррентный с модулями LSTM, число нейронов в LSTM модуле – 128;
- третий слой dropout используется для предотвращения переобучения сети.
- четвертый слой – полносвязный слой прямого распространения, с функцией активации RELU;
- пятый слой – слой активации softmax;
- последний, выходной слой выводит результат классификации по наличию / отсутствию отклонений в данных, вероятность от 0 до 1 в пользу того или иного варианта;
- алгоритм обучения – адаптивная оценка момента ADAM;
- время обучения: 100 эпох;
- начальная скорость обучения (initial learning rate) – 0.01;
- количество наблюдений в каждом пакете, MiniBatchSize = 1.

Итогом работы нейросети являются значения вероятности наступления события, в данном случае, отклонения, для каждого анализируемого значения измерения. Если рассчитанная нейросетью вероятность превышает некоторое заданное пороговое значения, то такой результат измерений классифицируется как отклонение.

Следующим этапом обученная LSTM нейросеть была применена к тестовой выборке для классификации данных.

Результаты LSTM – классификации представлены в матрице спутанности (confusion matrix), которую иллюстрирует рис. 2. Строки матрицы спутанности соответствуют истинному классу, а столбцы – предсказанному классу. Диагональные и недиагональные ячейки соответствуют правильно и неправильно классифицированным наблюдениям соответственно.

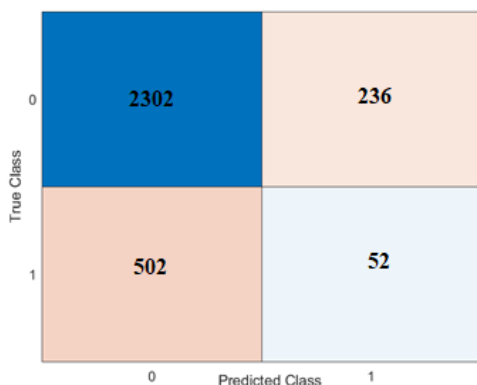


Рис. 2. Результаты классификации при помощи LSTM нейросети

Матрица показывает, что нейросеть классифицирует наличие (52) или отсутствие (2302) отклонений, однако, присутствует большое количество нераспознанных отклонений (236) или принятых за отклонения данных, таковыми не являющихся (502). Эти недостатки могут быть нейтрализованы применением стекинга, одного из алгоритмов ансамблирования [22–24], то есть дополнительного второго классификатора для данных, полученных из первого, реализуя стекинг (stacking), бэггинг и бустинг.

### III. ПРИМЕНЕНИЕ АЛГОРИТМА СТЕКИНГА

Для реализации дополнительного алгоритма обучения выбран стекинг, алгоритм ансамблирования, основные преимущества которого состоят в том, что он может использовать алгоритмы разного типа [22–24]. Стекинг – обучение нескольких разных алгоритмов и передача их результатов на вход последнему, который называют мета-алгоритмом. Мета-алгоритм объединяет результаты работы других алгоритмов и принимает итоговое решение.

Ключевые особенности стекинга:

- использование различных базовых алгоритмов;
- наличие мета-алгоритма.

Финальное решение мета-алгоритм принимает, опираясь на результаты базовых алгоритмов.

Для применения второго алгоритма машинного обучения, результаты классификации первого были обработаны следующим образом:

- выбраны 5 результатов первой нейронной сети до часа, где точно есть отклонение (этот гиперпараметр можно настраивать);
- выбраны 10 точек значений, где точно нет отклонения (этот гиперпараметр также можно настраивать);
- в полученной таким образом выборке найдена скорость изменения данных, находящихся до и после отклонения, при помощи функции diff, Matlab;
- созданы случайные последовательности по 10 интервалов, длиной 5 точек, для получения большего объема выборки и улучшения качества обучения;
- выборка скоростей изменения данных, находящихся непосредственно перед отклонением и данных, где отклонений точно нет, поданы на Classification Learner;
- определение, при помощи Classification Learner алгоритма, обеспечивающего максимальную точность обучения и предсказания отклонения.

Алгоритмом, обеспечившим максимальную точность предсказания 98.8 %, оказалась трехслойная нейронная сеть прямого распространения (рис. 3):

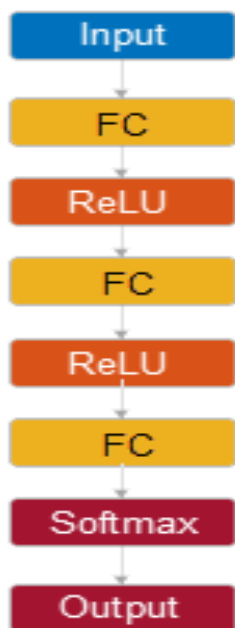


Рис. 3. Архитектура сети прямого распространения для детектирования скорости изменения отклонений

Параметры классификатора алгоритмом стекинга:

- первый слой – вектор входных факторов: данные перед отклонением, данные, где отклонений точно нет; скорости изменения данных;
- три рабочих полносвязных слоя прямого распространения, число нейронов в каждом слое: 10;
- функция активации: RELU, определяет выходное значение нейрона в зависимости от результата взвешенной суммы входов и порогового значения. ReLU имеет формулу  $f(x) = \max(0, x)$  и реализует простой пороговый переход в нуле;
- алгоритм обучения: LBFGS (квазиньютоновский метод, аппроксимирующий алгоритм Бroyдена–Флетчера–Гольдфарба–Шенно (BFGS). Алгоритм L-BFGS лучше всего подходит для небольших сетей и наборов данных, которые можно обрабатывать за один пакет) [25];
- на выходе – результат классификации по наличию / отсутствию отклонений в данных, вероятность от 0 до 1 в пользу того или иного варианта.

Результаты классификации скоростей изменения выходных данных LSTM-нейросети при помощи стекинга со второй нейросетью прямого распространения, представлены в матрице спутанности (confusion matrix), которую иллюстрирует рис. 4.

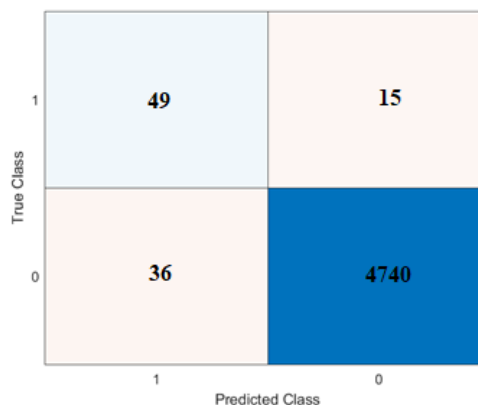


Рис. 4. Результаты применения алгоритма стекинга

Матрицы спутанности иллюстрируют, (49), их отсутствия (4740). количество нераспознанных отклонений (15) или принятых за отклонения данных, таковыми не являющихся (36), значительно снились. Сравнение данных, результатов работы первого и второго классификаторов и разметки, на тестовой выборке ГРС Краснознаменская иллюстрирует рис. 5. Для комфорта визуализации использовались сравнительные шкалы, справа шкала для нормированного расхода, единица измерения млн.м3, слева – безразмерная шкала для разметки и результатов работы алгоритмов машинного обучения, значение разметки при построении графика умножено на 2, чтобы можно было отличить ее от остальных данных на рисунке, по горизонтальной оси отложены часы. Пунктиром обозначены значения первого классификатора, точками – значения второго, вертикальными линиями – разметка, сплошной неровной линией – реальные данные по объемному расходу газа.

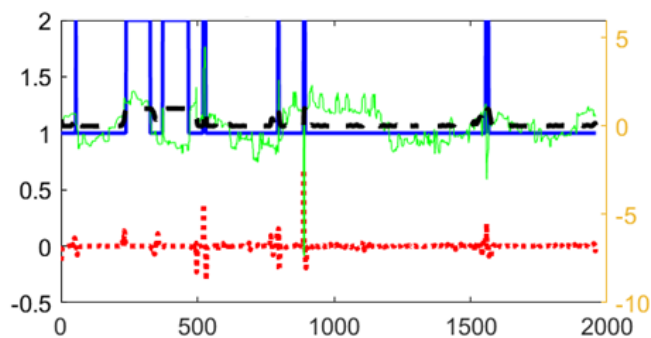


Рис. 5. Сравнение реальных данных, результатов работы первого и второго классификаторов и разметки, на тестовой выборке Краснознаменской газораспределительной станции

Полученный набор алгоритмов машинного обучения (классификационная LSTM нейросетью и нейросеть прямого распространения) можно применять вместе с регрессионной LSTM нейросетью, полученной ранее [13]. Это позволит одновременно выявлять некоторые виды отклонений заблаговременно, а также определять их значения и динамику, и классифицировать их.

## IV. ЗАКЛЮЧЕНИЕ

В статье рассматривается применение алгоритма стекинга в сравнении с LSTM нейронной сетью. Матрица спутанности результаты классификации при помощи стекинга иллюстрирует, что набор алгоритмов точнее классифицирует данные на наличие отклонений. Доля ошибок, при этом, варьируется от объекта к объекту ГТС. Это может быть направлением для дальнейшего изучения причин таких различий. Указанные ошибки могут быть нейтрализованы качеством разметки и подбором гиперпараметров ансамбля стекинга. Благодаря заблаговременному обнаружению и классификации аномалий, удастся принимать адекватные меры, оптимальные с точки зрения соотношения потерь газа и затрат на профилактику и устранения небаланса.

## СПИСОК ЛИТЕРАТУРЫ

- [1] Haykin S. *Neural networks. Complete course*. Williams, 2018, 1104 p.
- [2] Vladimir N. Vapnik. *The nature of statistics. Learning theory*. Second edition. Springer Verlag NY, 2005.
- [3] Назаров А.В., Лоскутов А.А. *Нейросетевые алгоритмы прогнозирования и оптимизации систем: Наука и Техника, СПб, 2003, 384 с.*
- [4] Holland J.H. *Adaptation in Natural and Artificial Systems*, University of Michigan Press, Ann Arbor, 1975.
- [5] Круглов В.В., Борисов В.В. *Искусственные нейронные сети: теория и практика*. М.: Горячая линия – Телеком, 2002.
- [6] M. Moller, "A Scaled Conjugate Gradient Algorithm for Fast Supervised Learning," *Neural Networks*, Vol. 6, No. 4, 1993, pp. 525-533.
- [7] Shashkov V.B. *Applied regression analysis. Multifactorial regression: Textbook*. Orenburg: GOU VPO OSU, 2003. Russian. Erkan Erdogdu, *Natural gas demand in Turkey*, *Appl. Energy* 87 (1) (2010) 211–219.
- [8] Vincenzo Bianco, Federico Scarpa, Luca A. Tagliafico, *Scenario analysis of nonresidential natural gas consumption in Italy*, *Appl. Energy* 113 (2014) 392–403.
- [9] Jolanta Szoplik, *Forecasting of natural gas consumption with artificial neural networks*, *Energy* 85 (2015) 208–220.
- [10] Coskun Hamzacebi, Huseyin Avni Es, *Forecasting the annual electricity consumption of Turkey using an optimized grey model*, *Energy* 70 (2014) 165–171.
- [11] Hagan M.T., and M. Menhaj, "Training feed-forward networks with the MarQuardt algorithm," *IEEE Transactions on Neural Networks*, Vol. 5, No. 6, 1999, pp. 989–993, 1994.
- [12] Srivastava, Hinton, Krizhevsky, Sutskever and Salakhutdinov, *Dropout: A Simple Way to Prevent Neural Networks from Overfitting* // *Journal of Machine Learning Research* 15 (2014) 1929-1958.
- [13] Петрова А.К. *Детектирование аномалий в технологических процессах газотранспортной системы с использованием сетей с долгой краткосрочной памятью (LSTM)* // Сборник докладов Международной конференции «Проектирование и обеспечение качества информационных процессов и систем» СПбГЭТУ «ЛЭТИ», 2022, с. 95-98.
- [14] J.A.K. Suykens, J. De Brabanter, L. Lukas, J. Vandewalle, *Weighted least sQuares support vector machines: robustness and sparse approximation*, *Neurocomputing* 48 (1–4) (2002) 85–105.
- [15] Fazil Kaytez, M. Cengiz Taplamacioglu, Ertugrul Cam, and Firat Hardalac, *Forecasting electricity consumption: A comparison of regression analysis, neural networks and least sQuares support vector machines*, *Int. J. Electr. Power Energy Syst.* 67 (2015) 431–438.
- [16] Gang Xie, Shouyang Wang, Kin Keung Lai, *Short-term forecasting of air passenger by using hybrid seasonal decomposition and least sQuares support vector regression approaches*, *J. Air Transp. Manag.* 37 (2014) 20–26.
- [17] You Lv, Jizhen Liu, Tingting Yang, Deliang Zeng, *A novel least sQuares support vector machine ensemble model for NOx emission prediction of a coal-fired boiler*, *Energy* 55 (2013) 319–329.
- [18] P.J. García Nieto, J.R. Alonso Fernández, V.M. González Suárez, C. Díaz Muñiz, E. García-Gonzalo, R. Mayo Bayón, *A hybrid PSO optimized SVM-based method for predicting of the cyanotoxin content from experimental cyanobacteria concentrations in the Trasona reservoir: A case study in Northern Spain*, *Appl. Math. Comput.* 260 (2015) 170–187
- [19] Hochreiter S. *Long short-term memory* // *Neural Computation*. 1997. 9 (8). P.1735–1780. DOI:10.1162/neco.1997.9.8.1735. PMID 9377276.
- [20] Gers F.A. *Learning to Forget: Continual Prediction with LSTM* // *Neural Computation*. 2000. 12 (10). P.2451–2471. DOI: 10.1162/089976600300015015.
- [21] Hochreiter S. *Gradient flow in recurrent nets: the difficulty of learning long-term dependencies* [Text] / S. Hochreiter, Y. Bengio, P. Frasconi, J. Schmidhuber // In S.C. Kremer and J.F. Kolen, editors, *A Field Guide to Dynamical Recurrent Neural Networks*. IEEE Press. 2001. P. 464–467.
- [22] *Ensemble Machine Learning, A beginner's guide that combines powerful machine learning algorithms to build optimized models.*, Copyright©2017 Packt Publishing, ISBN 978-1-78829-775-2.
- [23] Cha Zhang, Yunqian Ma Editors, *Ensemble Machine Learning, Methods and Applications*, ISBN 978-1-4419-9325-0 e- ISBN 978-1-4419-9326-7. © Springer Science+Business Media, LLC 2012
- [1] David H. Wolpert, *Stacked generalization*, *Neural Networks*, V. 5, Issue 2, 1992, Pages 241-259, ISSN 0893-6080, [электронный ресурс] режим доступа [https://doi.org/10.1016/S0893-6080\(05\)80023-1](https://doi.org/10.1016/S0893-6080(05)80023-1). (дата обращения 11.03.2024).
- [24] Liu, Dong C., and Jorge Nocedal. "On the limited memory BFGS method for large scale optimization." *Mathematical programming* 45, no. 1 (August 1989): 503-528. <https://doi.org/10.1007/BF01589116>.

# Применение методов машинного обучения для приведения параметров работы ГТУ для диагностики турбокомпрессора

В. С. Карелин, И. В. Попов, П. В. Соколов

Санкт-Петербургский государственный электротехнический университет  
«ЛЭТИ» им. В.И. Ульянова (Ленина)

karelinvld@gmail.com, popov\_iv2@power-m.ru, pvsokolov@etu.ru

**Аннотация.** С 2017 года развиваются отечественные системы мониторинга и предиктивной диагностики энергетического оборудования. Развитию данных систем поспособствовала четвертая промышленная революция "Индустрия 4.0», а также широкая цифровизация и увеличение разнообразных контрольно-измерительных приборов, а значит и увеличение статистических массивов данных, без которых развитие данных систем было бы невозможно. Помимо этого, с всеобщим курсом импортозамещения, активно начали разрабатываться и реализовываться российские газовые турбины, требующие повышенного внимания систем диагностики для нормальной эксплуатации и контроля работы. В данной работе описывается алгоритм приведения параметров работы ГТУ, построенный на базе машинного обучения. Данное приведение необходимо для дальнейшей диагностики турбокомпрессора ГТУ и выявления его основных неисправностей.

**Ключевые слова:** предиктивная диагностика, газотурбинная установка, машинное обучение, энергетическое оборудование

## I. ВВЕДЕНИЕ

С 2017 года развиваются отечественные системы мониторинга и предиктивной диагностики энергетического оборудования. Развитию данных систем поспособствовала четвертая промышленная революция "Индустрия 4.0», а также широкая цифровизация и увеличение разнообразных контрольно-измерительных приборов, а значит и увеличение статистических массивов данных, без которых развитие данных систем было бы невозможно. Однако не стоит забывать и о работах прошлого. Так, в [1] дается определение технической диагностики следующим образом: «техническая диагностика – наука о распознавании состояния технической системы, включающая широкий круг проблем, связанных с получением и оценкой диагностической информации». Помимо этого, с всеобщим курсом импортозамещения, активно начали разрабатываться и реализовываться российские газовые турбины, требующие повышенного внимания систем диагностики для нормальной эксплуатации и контроля работы.

Все вышеуказанные предпосылки «подталкивают» современных производителей энергетического оборудования и других игроков IT-рынка к созданию и реализации систем мониторинга и диагностики на самых

ранних этапах разработки энергетических установок. Эксплуатация, в свою очередь, заинтересована в данных системах не меньше, так как по средствам мониторинга и диагностики решаются задачи повышения эксплуатационной надежности, экономичности и доступности энергетических установок, а также предотвращение возможных аварий и планирование соответствующих ремонтов.

Газовая турбина – это энергетическая установка, необходимая для обеспечения движения турбогенератора [2]. Основным преимуществом газовых турбин по сравнению с паровыми можно выделить следующие пункты:

- уменьшенные габариты при одинаковой мощности с паровыми;
- быстрый запуск;
- высокая манёвренность;
- широкий спектр используемого топлива;
- постоянство электрической частоты.

Как известно, внедрение газотурбинных установок позволяет сократить потребность экономики региона в энергетическом топливе и обеспечить прирост энергетических мощностей без строительства новых неэкономичных и сложных паротурбинных ГРЭС[3]. Тем самым становится ясно, что бесперебойная работа ГТУ важна для всех участников процесса выработки электроэнергии. А для обеспечения бесперебойной работы необходим соответствующий инструмент, чем являются системы мониторинга и предиктивной диагностики оборудования.

К наиболее нагруженным элементам ГТУ относятся компрессор, турбина и камера сгорания. В данной работе описан процесс приведения режимов работы газотурбинной установки для дальнейшей диагностики турбокомпрессора.

## II. ПРЕДОБРАБОТКА ИСХОДНЫХ ДАННЫХ

Для анализа состояния турбокомпрессора ГТУ определим перечень необходимых исходных параметров:



ТАБЛИЦА I. ИСХОДНЫЕ ПАРАМЕТРЫ

Наименование параметра	Обозначение	Единица измерения
Скорость (частота) вращения ротора	nГ	Об/мин
Мощность на клеммах генератора	$N_{эл}$	Вт
Температура воздуха перед компрессором	$T_n$	К
Давление воздуха перед компрессором	$P_n$	Па
Полная температура за компрессором	$T_k$	К
Полное давление за компрессором	$P_k$	Па
Расход топлива	$G_m$	Кг/с
Температура выхлопных газов	$T_v$	К
Угол установки ВНА компрессора	$\alpha_T, \alpha_P, \alpha_G$	°

Для совместной обработки параметров значения должны регистрироваться в один и те же моменты времени.

Валидация данных может выполняться по следующим критериям:

- по ограничениям на порядок значений измеренных параметров;
- по ограничениям на количество возможных повторений значений в реализации;
- по скорости изменения параметров в течение установленных интервалов времени;
- по данным об изменении наработки (контроль хронологической последовательности, контроль потери информации).

Фактически применяемый алгоритм очистки временных рядов параметров разработан с учетом особенностей эмпирических данных. Блок-схема представлена на рис. 1.

Упрощённая блок-схема алгоритма:



Рис. 1. Блок-схема алгоритма

Измеряемые параметры работы ГТУ проходят первичную обработку:

- Выполняется разделение по характерным рабочим режимам А и Б, таким, что:

А – это режимы работы с малой нагрузкой, с полностью закрытым ВНА ( $\alpha_{ВНА} = -30^\circ$ ), расход компрессора минимален и неизменен, а регулированием мощности осуществляется путём изменения температуры на выходе КС;

Б – это расчётные режимы работы с поддержанием постоянной температуры на выходе турбины, когда регулирование мощности осуществляется путём изменения положения ВНА ( $\alpha_{ВНА} = -30^\circ \dots 0^\circ$ ), т.е. за счёт изменения расхода воздуха и давления за компрессором.

- Выделяются точки, соответствующие установившемуся режиму, т.е. такому, для которого выполняется условие:

$$[(\Delta n < 1 \%) \text{ И } (\Delta T_k < 2 \%) \text{ И } (\Delta P_k < 2 \%)$$

$$\text{И } (\Delta G_m < 10 \%) \text{ И } (\Delta T_v < 2 \%) \text{ И } (\Delta \alpha_{ВНА} < 1^\circ)] \quad (1)$$

в течение интервала времени  $t \geq 15$  минут (интервал может корректироваться в процессе отладки алгоритма).

Затем значения рабочих параметров ГТУ усредняются на интервалах  $t$ , для которых режим работы интерпретируется как установившийся, тем самым формируются характерные средние значения на интервалах – реперные точки. Значения параметров для интервалов, не подпадающих под определение установившегося режима, отбрасываются и в дальнейшей обработке не участвуют.

Далее, на начальном этапе функционирования алгоритма осуществляется его обучение, настройка на конкретный диагностируемый образец ГТУ. Методология подразумевает приведение параметров работы ГТУ к стандартной атмосфере и номинальной мощности следующим образом:

$$T_{в_{CApr}} = T_{в_{CA}} + \alpha_T * (N_{e_{ном}} - N_{e_{изм}}) \quad (2)$$

$$P_{к_{пр}} = P_{к_{изм}} + \alpha_P * (N_{e_{ном}} - N_{e_{изм}}) \quad (3)$$

$$G_{m_{пр}} = G_{m_{изм}} + \alpha_G * (N_{e_{ном}} - N_{e_{изм}}) \quad (4)$$

где,

$$T_{в_{CA}} = T_{в_{изм}} * (288,15 / T_{н_{изм}}) \quad (5)$$

$$N_{e_{изм}} = N_{эл_{изм}} / \eta_{ТГ} \quad (6)$$

$$N_{e_{пр}} = N_{e_{изм}} * (288,15 / T_{н_{изм}}) \quad (7)$$

$\alpha_T, \alpha_P, \alpha_G$  – коэффициенты приведения для температуры за турбиной, давления за компрессором и расхода топливного газа к базовой мощности.

### III. ПРИВЕДЕНИЕ ПАРАМЕТРОВ К ПОСТОЯННОМУ РЕЖИМУ РАБОТЫ ГТУ

Приведение параметров возможно произвести двумя способами: по средствам расчетного подхода, либо эмпирическим путем. Оба подхода имеют свои сильные стороны и свои ограничения, которые необходимо принимать во внимание при выборе расчётного или эмпирического подхода. Краткое сопоставление указанных двух подходов представлено в табл. 2.

ТАБЛИЦА II. СРАВНЕНИЕ МЕТОДОВ ПОСТРОЕНИЯ МОДЕЛИ

Метод построения модели	Расчётный подход – использование математической модели ГТУ	Эмпирический подход – использование опытных данных на начальном этапе
Возможность осуществления диагностики с момента ввода ГТУ в эксплуатацию	Имеется. Аппроксимация взаимосвязей рабочих параметров строится по имеющейся математической модели ГТУ	Отсутствует. Для построения модели необходимо накопление опытных данных, охватывающих широкий диапазон по атмосферным условиям и мощности (полгода – год работы)
Возможность диагностирования ГТУ без расчётных данных	Отсутствует, так как отсутствуют расчётные данные, математическая модель для ГТУ стороннего производителя	Имеется, диагностирование возможно после накопления начальной выборки эмпирических данных
Настройка под конкретный образец ГТУ	Имеется, осуществляется путём корректировки соответствующих конфигурационных параметров, включаемых в используемую математическую модель ГТУ	Имеется, осуществляется путём корректировки модели на имеющихся опытных данных
Уязвимость, связанная с наличием дефектов ГТУ изначально, при настройке и вводе модели в работу	Отсутствует. Благодаря использованию математической модели именно исправной ГТУ, существенное расхождение между моделью и фактическим поведением ГТУ станет сразу очевидным и будет свидетельствовать о проблемах с ГТУ.	Имеется, так как модель строится на эмпирических данных. Если ГТУ изначально неисправна, модель, обученная на опытных данных такой ГТУ, будет воспринимать неисправное состояние как норму.

Кроме формирования самих зависимостей оценивается также и нормальный разброс значений приведённых параметров относительно среднего значения, используемый далее в работе алгоритма для классификации наблюдаемых выбросов параметров как допустимых или аномальных отклонений, свидетельствующих о возникновении дефекта.

В связи с отсутствием математической модели ГТУ в данной работе был использован эмпирический подход для приведения параметров работы ГТУ.

Далее, после настройки модели, на основном этапе работы алгоритма, выполняется формирование реперных точек и их приведение к номинальному режиму и стандартным условиям с использованием зависимостей, сформированных на начальном этапе. При этом контролируется:

- появление существенного отклонения, выпадения реперной точки от полученной базовой зависимости, что может быть вызвано наличием дефекта;
- слом ранее наблюдавшегося тренда изменения реперных точек во времени, свидетельствующий о нарушении нормальных темпов естественной деградации, износа узлов ГТУ и ускоренном развитии какого-либо дефекта.

При обнаружении таких отклонений по их характерным сочетаниям классифицируется вероятный дефект с использованием экспертных диагностических правил, заложенных в знаковую матрицу. Сама матрица может быть настроена как с использованием имеющихся опытных данных по ранее наблюдавшимся отказам, так и экспертно, на основе глубокого анализа рабочих процессов ГТУ, выполнения тестовых расчётов с использованием имеющихся математических моделей и т. п.

Для оценки взаимного влияния различных параметров работы ГТУ был проведён статистический анализ данных на предмет корреляции значений различных параметров работы между собой. Результаты данного анализа представлены на графике на рис. 2.

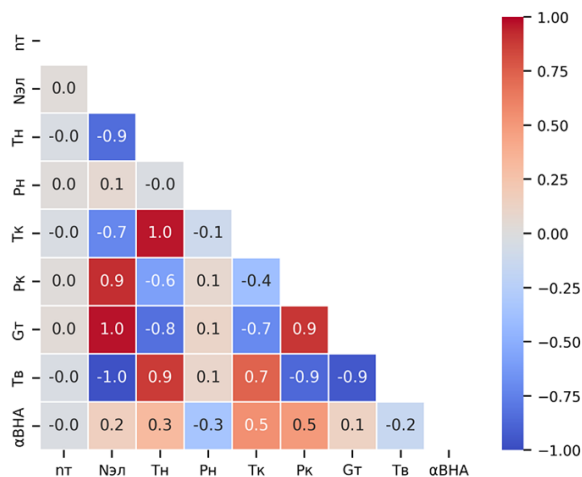


Рис. 2. Корреляционная матрица

По представленным данным можно сделать следующие выводы:

- появление существенного отклонения, выпадения реперной точки от полученной базовой зависимости, что может быть вызвано наличием дефекта;

- при анализе коэффициентов корреляции необходимо учитывать условия, при которых они были получены путём статистической обработки данных, физику самого объекта диагностирования. Машинная обработка данных иногда может выявить ложные зависимости, характерные только для отдельных частных случаев. Например, коэффициент  $-0,9$  между начальной температурой  $T_n$  и мощностью  $N_e$  связан с эксплуатацией ГТУ вблизи кривой пиковой мощности, для которой такая взаимосвязь справедлива. В общем же случае, ГТУ может эксплуатироваться на множестве режимов, лежащих ниже предельно достижимого максимума, поэтому такой однозначной взаимосвязи  $T_n$  и  $N_e$  не существует;
- с другой стороны, начальное давление  $P_n$ , при прочих равных оказывающее прямое влияние на развиваемую мощность  $N_e$ , как и температура воздуха на входе, формально имеет малый коэффициент корреляции с мощностью  $N_e$ . Малое влияние на мощность имеет также и угол установки ВНА, при этом как раз и определяющий значение мощности на данных режимах в широком диапазоне регулирования примерно от 50 до 100 % нагрузки.

#### IV. ЗАКЛЮЧЕНИЕ

В результате работы был определен алгоритм приведения данных работы ГТУ. Был выбран набор исходных данных, необходимый для функционирования и последующего использования алгоритма для диагностики турбокомпрессора ГТУ. В статье были рассмотрены два подхода. Первый основывается на математической модели ГТУ. Второй и выбранный для данной работы опирается на методы машинного обучения. Оба подхода имеют свои преимущества и недостатки. Основное преимущество метода, основанного на машинном обучении, заключается в возможности определения основных зависимостей эмпирическими путем, не имея расчетных характеристик объекта диагностики.

Данный алгоритм является подготовительной частью диагностического алгоритма для турбокомпрессора ГТУ.

#### СПИСОК ЛИТЕРАТУРЫ

- [1] Биргер И.А. Техническая диагностика. Москва: Машиностроение, 1978. 240 с.
- [2] Газовая турбина: назначение и области применения // АТФ URL: [https://atf.ru/articles/materialy\\_dlya\\_promyshlennogo\\_oborudovaniya/gazovaya-turbina-naznachenie-i-oblasti-primeneniya/](https://atf.ru/articles/materialy_dlya_promyshlennogo_oborudovaniya/gazovaya-turbina-naznachenie-i-oblasti-primeneniya/) (дата обращения: 08.04.2024).
- [3] Буланин В.А. Использование газовых турбин для комбинированного производства энергии. Вестник Дагестанского государственного технического университета. Технические науки. 2020; 47 (1): 8-18. DOI:10.21822/2073-6185-2020-47-1-8-18

# Обнаружение аномалий на медицинских изображениях с использованием трансферного обучения и оптимизации данных

Я. А. Низамли<sup>1</sup>, В. В. Фадел, А. Ю. Филатов, Ю. А. Шичкина

Санкт-Петербургский государственный электротехнический университет  
«ЛЭТИ» им. В.И. Ульянова (Ленина)

<sup>1</sup>yanizamli@stud.etu.ru

**Аннотация.** Сегодня методы медицинской визуализации широко используются для выявления различных состояний и заболеваний человека. Чтобы ускорить процесс диагностики, системы часто автоматизируются с использованием методов глубокого обучения, которые, как доказано, дают выдающиеся результаты. Однако эти модели могут быть ограничены конкретной областью, подвержены переобучению и их трудно обновлять. В этой статье мы представляем подход к трансферному обучению, который обеспечивает высокую точность обнаружения аномалий как на изображениях МРТ, так и на КТ. Наша система использует сеть ConvNeXt-XL, вдохновленную архитектурой трансформатора зрения, в качестве средства извлечения признаков высокого уровня из аугментированных и обработанных медицинских изображений. Полученные данные затем оптимизируются с использованием алгоритма Edited Nearest Neighbors, который выбирает наиболее эффективные наблюдения и отбрасывает остальные. Наконец, данные передаются в классификатор K-Nearest Neighbors для обнаружения аномалий в изображениях, представленных в числовом виде. Наша модель обеспечивает высокую производительность на двух разных типах медицинских изображений и может быть легко обновлена без необходимости повторного обучения.

**Ключевые слова:** МРТ, КТ, трансферное обучение, оптимизация данных

## I. ВВЕДЕНИЕ

Медицинская визуализация считается одним из наиболее распространенных анализов, выполняемых с целью получить более глубокое представление о человеческом теле без какого-либо хирургического вмешательства. Методы медицинской визуализации позволяют специалистам медицинских учреждений просматривать и диагностировать широкий спектр патологических состояний и, таким образом, назначать наиболее подходящий курс лечения и оценивать потенциальные риски. Посредством повторной визуализации специалисты могут измерить прогресс выбранного лечения, оценить его эффективность и при необходимости скорректировать план лечения пациента [1, 2].

Анамнез и текущее состояние пациента, помимо некоторых особых диагностических требований, являются факторами, влияющими на выбор подходящего метода диагностики. Среди наиболее часто используемых методов визуализации в

медицинских учреждениях – магнитно-резонансная томография (МРТ) и компьютерная томография (КТ). По сути, МРТ использует магнитные поля и радиоволны для создания изображений внутренней структуры тела, тогда как КТ опирается на рентгеновские снимки, полученные под разными углами и объединенные компьютером. Оба обеспечивают точные изображения структур и тканей организма, и каждый из них имеет конкретное клиническое применение. Преимущество МРТ заключается в том, что она показывает лучший контраст мягких тканей, тогда как КТ предпочтительна для костных структур и в экстренных случаях [1–3].

Ручная диагностика с использованием медицинской визуализации, которая представляет собой процесс интерпретации и изучения изображений специалистами с академической подготовкой, считается дорогостоящей, требующей усилий и времени процедурой. Эта нагрузка на организации здравоохранения критически возрастает, особенно при возникновении необычных и неожиданных обстоятельств, таких как эпидемии и новые широко распространенные заболевания. Для преодоления этих проблем в качестве перспективного варианта и активного направления исследований появляются компьютерные системы, основанные на методах машинного обучения и глубокого обучения, отличающиеся своей производительностью и способностью обрабатывать огромные объемы данных [2–4].

Чтобы добиться максимальной эффективности в диагностике заболеваний с использованием медицинских изображений, исследователи внедряют системы, основанные на различных подходах искусственного интеллекта и машинного обучения. Исследователи в [3, 4] используют ручное извлечение признаков для преобразования медицинских изображений в числовое представление, которое можно обрабатывать с помощью классического машинного обучения. Среди этих методов – гистограммы ориентированных градиентов (HOG), матрица совпадения уровней серого (GLCM), локальный бинарный шаблон (LBP) и метод разницы уровней серого (GLDM). Полученные признаки передаются известным алгоритмам, таким как машина опорных векторов, дерево решений, случайный лес, наивный Байес и другие, для выполнения задачи распознавания или классификации состояния, выраженного на

медицинском изображении. В другом подходе в [4–9] исследователи используют модели глубокого обучения, сформированные в основном из серии последовательных блоков свертка-пулинга, за которыми следует полносвязанная нейронная сеть. Модели могут отличаться друг от друга количеством слоев, размером окон свертки, количеством фильтров, размером шага, типом пулинга, типом функций потерь, типом функций активации, количеством нейронов и многие другие. Некоторые из этих моделей просты по структуре и созданы с нуля, тогда как другие созданы на основе стандартных моделей, таких как LeNet.

Чтобы преодолеть недостатки, которые могут возникнуть при обучении глубоких моделей, исследователи обращаются к подходам трансферного обучения для решения задач, связанных с обнаружением и классификацией аномалий на медицинских изображениях. В [10–15] глубокие модели, такие как VGG, ResNet, AlexNet, Inception, DenseNet и Xception, используются для извлечения признаков высокого уровня из изображений МРТ и КТ. В этой методологии параметры моделей, обученных на наборе данных ImageNet, фиксируются, а слои классификации переобучаются после их изменения в соответствии с решаемой задачей. Верхние слои моделей также можно удалить и заменить различными алгоритмами или методами, выполняющими задачу классификации извлеченных признаков.

Недостаточная точность является одной из основных проблем представленных ранее систем, особенно с учетом важности и деликатности задач, связанных с жизнью человека, что ставит под сомнение надежность этих систем и возможность внедрения их в реальное использование. Кроме того, невозможность обобщения, возникающая в результате переобучения при обучении глубоких моделей с большим количеством параметров на ограниченном объеме данных, приведет к низкой производительности для новых реальных данных. Эта проблема остается актуальной даже при использовании трансферного обучения, когда извлеченные крупномерные признаки передаются для обучения алгоритмам машинного обучения. В любом случае обновление предыдущих систем считается сложной процедурой, которая обычно требует повторного обучения некоторых или всех элементов модели, чтобы снова подогнать параметры под новые добавленные данные.

В этой статье мы стремимся решить затронутые проблемы, поэтому наш основной вклад заключается в следующем: (1) представить подход к трансферному обучению, который обеспечивает требуемую высокую производительность при обнаружении аномалий как на изображениях МРТ, так и на КТ, (2) использовать и оценить эффективность революционной сетевой архитектуры ConvNeXt как компонента извлечения признаков в задачах обработки медицинских изображений, (3) повысить производительность за счет обработки данных с использованием алгоритма edited nearest neighbors, (4) разработать систему, которую можно легко обновить для новых добавленных данных без повторного обучения, (5) сравнить

производительность разработанной модели с теми, которые доступны в недавних работах.

## II. МАТЕРИАЛЫ И МЕТОДЫ

На рис. 1 представлена архитектура предлагаемого подхода, где операции системы можно резюмировать следующим образом: (1) изображения МРТ головного мозга и КТ грудной клетки подвергаются серии операций аугментации и обработки, (2) обработанные изображения в каждом наборе данных передаются отдельно в предобученную модель ConvNeXt с исключенным классификационным слоем для извлечения числовых признаков, (3) признаки, извлеченные из каждого изображения, преобразуются в одномерный вектор с использованием выравнивающего слоя (flatten layer), (4) сформированные данные подвергаются процессу оптимизации путем отбора лучших наблюдений с использованием метода edited nearest neighbors, (5) алгоритм  $k$ -nearest neighbors используется для классификации изображений МРТ на нормальные или показывающие опухоль, тогда как изображения КТ классифицируются как указывающие на COVID-19 или другие случаи.

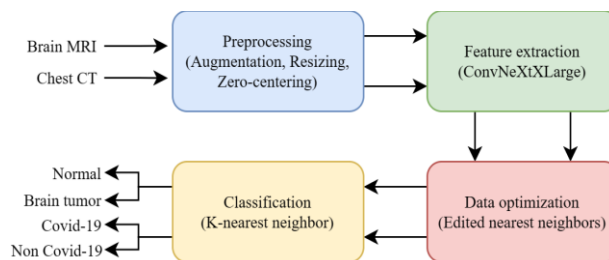


Рис. 1. Схема предлагаемой системы

### A. Набор данных и предобработка

В нашем исследовании, чтобы протестировать предложенную модель, мы используем два разных набора данных. Первый включает изображения МРТ [16], а второй – компьютерную томографию [17]. Набор данных МРТ содержит 253 изображения головного мозга, разделенных на 98 нормальных изображений и 155 изображений, показывающих опухоль. Второй набор данных содержит 746 КТ грудной клетки, разделенных на две категории. Первая категория включает 349 изображений, показывающих инфекцию COVID-19, а вторая – 397 изображений других случаев. На рис. 2 показаны образцы из обоих наборов данных.

Размер двух наборов данных относительно невелик, поэтому мы прибегаем к применению ряда операций аугментации, чтобы увеличить данные примерно в три раза от их первоначального размера. Количество наблюдений в наборе данных МРТ увеличивается до 725 изображений, а размер набора данных КТ – до 2236 выборов. Выполняемые операции аугментации включают вертикальные и горизонтальные перевороты, вращения по часовой стрелке и против часовой стрелки. За этим следует изменение размера изображений до размеров  $224 \times 224 \times 3$  и центрирование по нулю относительно набора данных ImageNet перед передачей их в качестве входных данных в модель.

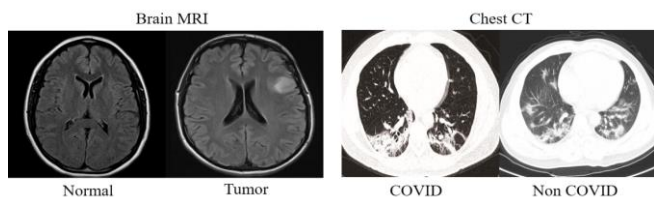


Рис. 2. Выборки из наборов данных

### В. Извлечение признаков

В нашем исследовании мы используем модель ConvNeXt для выполнения задачи извлечения признаков из обработанных медицинских изображений. ConvNeXt – это чистая сверточная нейронная модель, вдохновленная архитектурой иерархического трансформатора зрения (Swin Transformer). Новая революционная сеть состоит из серии последовательных остаточных блоков, каждый из которых начинается с depthwise свертки с относительно большим окном. За этим следует подход с обратным узким местом (inverted bottleneck), который включает выполнение двух операций свертки  $1 \times 1$  (pointwise свертки). Первый увеличивает размеры, а второй уменьшает их. Использование depthwise свертки и pointwise свертки в сети означает, что операции выполняются либо в пространственном измерении, либо в канальном измерении, но не в обоих вместе. Это свойство, в дополнение к подходу с обратным узким местом, можно обнаружить в каждом трансформаторе [18].

На микроуровне ConvNeXt, как и в случае с трансформаторами, использует меньшее количество функций активации и слоев нормализации, чем в других моделях, таких как ResNet. Функция активации ReLU также заменяется функцией GELU, а слой BatchNorm заменяется LayerNorm [18]. На рис. 3 показана структура блока ConvNeXt. Предыдущая архитектура продемонстрировала превосходную производительность по сравнению со всеми известными глубокими моделями, обученными на стандартных наборах данных, таких как ImageNet.

Основываясь на нашем обширном обзоре исследовательской литературы, мы обнаружили, что использование модели ConvNeXt в качестве средства извлечения признаков для задач обнаружения аномалий в медицинских изображениях еще не получило широкого признания. В нашем исследовании мы используем модель в ее самой большой по количеству фильтров и блоков версии под названием ConvNeXtXLarge. Верхние слои классификации удаляются, а веса фиксируются, чтобы модель выполняла задачу преобразования изображений в числовые признаки высокого уровня.

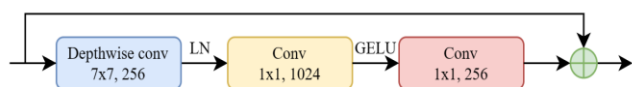


Рис. 3. Структура блока ConvNeXt

### С. Оптимизация данных

Новые числовые точки данных были созданы в результате передачи изображений из двух наборов данных через модель, обученную для совершенно

другой задачи и набора данных. Набор данных ImageNet, на котором обучалась модель, не является набором медицинских изображений [10, 13]. Это может привести к созданию зашумленных данных, находящихся далеко от областей плотности класса, к которому они принадлежат. Решить эту проблему можно с помощью простого алгоритма отредактированных ближайших соседей (ENN), который будет выполнять задачу очистки выборочного пространства путем исключения наблюдений, которые расположены в недопустимых областях классов и не согласуются со своими соседями [19]. Простой механизм используемого алгоритма можно объяснить следующей процедурой, применяемой к каждому классу данных:

- Для каждой выборки  $x$  в классе с:
  - Найти  $k$ -ближайших соседей;
  - Удалить выборку  $x$  из набора данных, если она не соответствует своим соседям по метке класса.

Параметр  $k$  играет очень важную роль в алгоритме ENN. По мере увеличения  $k$  количество учитываемых ближайших соседей увеличивается, а вероятность того, что выборка соответствует всем своим соседям, уменьшается, что означает исключение большего количества точек данных. Этот подход обеспечивает более чистые области данных и более строгие границы принятия решений, что, естественно, приводит к более высокой точности, но оказывает существенное влияние на распределение данных. На рис. 4 показано влияние выбора разных значений  $k$  на размер двух используемых наборов данных, где  $k = 0$  указывает на то, что ENN не применяется. Для достижения баланса было выбрано значение  $k = 1$ , что позволило выполнить минимально возможную обработку.

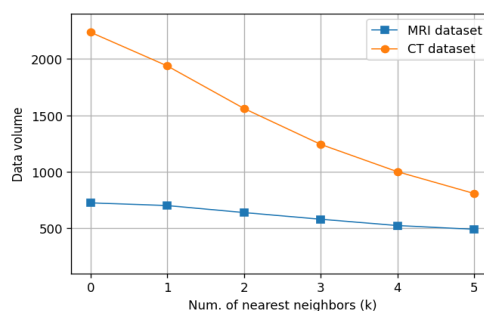


Рис. 4. Объем данных, полученный в результате применения ENN при разных значениях  $k$

### Д. Классификация

В нашей работе мы используем хорошо известный алгоритм  $k$ -ближайшего соседа (KNN), чтобы различать аномалии, выраженные в медицинских изображениях и представленные в виде числовых признаков. Принцип работы алгоритма основан на классификации новой выборки по голосам выбранных ближайших соседей. Следовательно, реального процесса обучения нет, а есть только простой расчет расстояния между выборкой и ее соседями. Такой подход позволяет легко обновлять систему во время работы, включая в набор данных новые классифицированные выборки и используя их

для принятия решений. В нашем исследовании мы изучаем использование 1-NN и 3-NN для выполнения окончательной классификации.

### III. РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

Для получения оценки, отражающей реальную производительность модели, разработанной в сервисе Jupyter Notebook Colab, использовался подход 10-кратной перекрестной проверки. Для каждого набора данных было проведено четыре эксперимента, включая оценку системы без и с обработкой данных с использованием ENN и для классификации на основе одного или трех соседей. В табл. 1 показан отчет о классификации предлагаемой системы. Хорошая точность наблюдается для всех экспериментов, особенно для набора данных МРТ, что указывает на эффективность глубоких признаков, извлеченных с помощью модели ConvNeXt, даже без использования ENN. Выполнение улучшения с минимальной обработкой с использованием ENN и в сочетании с классификацией на основе 1-ближайшего соседа позволяет достичь наилучшей точности 98.72% и 95.61% для наборов данных МРТ и КТ соответственно, без смещения в сторону определенных классов.

ТАБЛИЦА I. ОТЧЕТ О КЛАССИФИКАЦИИ ПРЕДЛОЖЕННОЙ МОДЕЛИ

MRI dataset					
Model	Class	PRE	SEN	F1	ACC
ConvNeXt+1NN	Normal	92.73	98.08	96.30	96.55
	Tumor	98.89	95.70		
ConvNeXt+ENN+1NN	Normal	96.93	99.61	98.62	98.72
	Tumor	99.77	98.21		
ConvNeXt+3NN	Normal	87.11	96.15	93.10	93.52
	Tumor	97.72	92.04		
ConvNeXt+ENN+3NN	Normal	90.51	97.64	95.14	95.44
	Tumor	98.59	94.18		
CT dataset					
ConvNeXt+1NN	Non Covid	87.78	85.71	86.01	86.05
	Covid	84.17	86.42		
ConvNeXt+ENN+1NN	Non Covid	95.95	95.85	95.59	95.61
	Covid	95.23	95.34		
ConvNeXt+3NN	Non Covid	83.93	87.31	84.23	84.35
	Covid	84.87	80.98		
ConvNeXt+ENN+3NN	Non Covid	88.31	91.80	89.02	89.11
	Covid	90.12	86.02		

На рис. 5 и рис. 6 показаны матрицы ошибок предлагаемой системы (ConvNeXtXLarge+ENN+1NN) для наборов данных МРТ и КТ соответственно. Для задачи обнаружения опухоли головного мозга только 9 выборки из всего набора данных были предсказаны неправильно, тогда как для задачи обнаружения COVID-19 общая ошибка не превысила 4.39%. В табл. 2 показано сравнение предлагаемой системы и моделей в литературе, которые используют те же наборы данных. Для набора данных МРТ представленная система значительно превосходит модели в [6, 8, 10, 11]. Хотя производительность близка к моделям в [5, 7], они страдают от серьезных проблем. Первый достиг значения F1, не превышающего 85%, что означает смещение в сторону одного из классов и точность не

отражает истинную производительность системы, а второй достиг точности 97.33% при выделении 10% данных для тестирования и только 94.70% при выделении 20%. Для набора данных КТ предложенная система достигла высокой точности, превосходящей доступные модели.

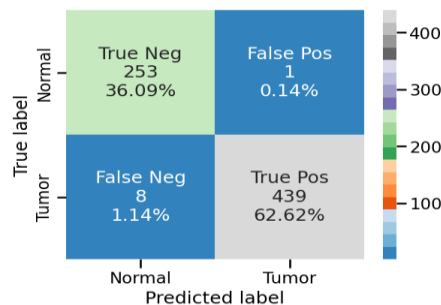


Рис. 1. Матрица ошибок предлагаемой модели (набор данных МРТ)

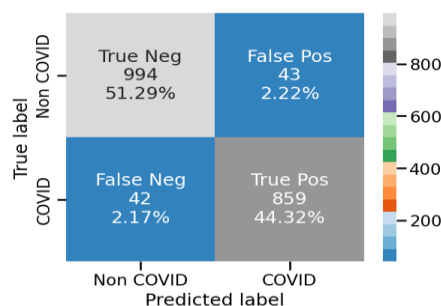


Рис. 5. Матрица ошибок предлагаемой модели (набор данных КТ)

ТАБЛИЦА II. СРАВНЕНИЕ ПРЕДЛОЖАЕМОЙ МОДЕЛИ И СУЩЕСТВУЮЩЕЙ ЛИТЕРАТУРЫ

MRI dataset		
Ref.	Accuracy	F1-score
[5]	98	85
[6]	94	94.3
[7]	97.33	97.50
[8]	96	94
[10]	88.26	-
[11]	91.94	91.80
Our model	98.72	98.62
CT dataset		
[9]	86.06	87
[12]	82.64	-
[13]	85.2	85.2
[14]	85	86
[15]	-	82
Our model	95.61	95.59

### IV. ЗАКЛЮЧЕНИЕ

В этом исследовании была предложена модель трансферного обучения с целью повышения эффективности обнаружения аномалий как на изображениях МРТ, так и на КТ. В нашем подходе после обработки и аугментации численное представление медицинских изображений извлекается с использованием предварительно обученной модели ConvNeXtXLarge. Полученные данные подвергаются процессу оптимизации, включающему очистку выборочного пространства с помощью алгоритма Edited Nearest Neighbors, после чего классификатор K-Nearest

Neighbors берет на себя задачу обнаружения опухоли головного мозга и COVID-19 на изображениях МРТ и КТ соответственно. Предложенная модель превзошла свои аналоги в литературе, достигнув высочайшей точности тестирования без необходимости обучения каких-либо параметров.

#### СПИСОК ЛИТЕРАТУРЫ

- [1] Kasban H., El-Bendary M. A. M., Salama D. H. A comparative study of medical imaging techniques // *International Journal of Information Science and Intelligent System*, 2015, vol. 4, no. 2, pp. 37-58.
- [2] Azizi A., Azizi M., Nasri M. Artificial intelligence techniques in medical imaging: A systematic review // *International Journal of Online and Biomedical Engineering*, 2023, vol. 19, no. 17, pp. 66-97.
- [3] Alavi S. E., Zare E., Rashti M. Brain tumors detection on MRI images through extracting HOG features // *Journal of Advanced Applied Scientific Research*, 2018, vol. 2, no. 1, pp. 9-25.
- [4] Sowrirajan S. R., Balasubramanian S. Brain tumor classification using machine learning and deep learning algorithms // *International Journal of Electrical and Electronics Research (IJEER)*, 2022, vol. 10, no. 4, pp. 999-1004.
- [5] Das T. K., Roy P. K., Uddin M., Srinivasan K., Chang C.-Y., Syed-Abdul S. Early tumor diagnosis in brain MR images via deep convolutional neural network model // *Computers, Materials & Continua*, 2021, vol. 68, no.2, pp. 2413–2429.
- [6] Gupta I., Singh S., Gupta S., Nayak S. R. Classification of brain tumours in MRI images using a convolutional neural network // *Current Medical Imaging*, 2024, vol. 20, no.1, pp. 1-9.
- [7] Rahman T., Islam M. S. MRI brain tumor detection and classification using parallel deep convolutional neural networks // *Measurement: Sensors*, 2023, vol. 26.
- [8] Einy S., Saygin H., Hivehch H., Navaei Y. D. Local and deep features based convolutional neural network frameworks for brain MRI anomaly detection // *Complexity*, 2022, vol. 2022.
- [9] Islam M. R., Matin A. Detection of COVID 19 from CT image by the novel LeNet-5 CNN architecture // *23rd International Conference on Computer and Information Technology (ICIT)*, 2020, pp. 1-5.
- [10] Srinivas C., Prasad N., Zakariah M., Alothaibi Y. A., Shaukat K., Partibane B., Awal H. Deep transfer learning approaches in performance analysis of brain tumor classification using MRI images // *Journal of Healthcare Engineering*, 2022, vol. 2022.
- [11] Asif S., Yi W., Ain Q. U., Hou J., Yi T., Si J. Improving effectiveness of different deep transfer learning-based models for detecting brain tumors from MR images // *IEEE Access*, 2022, vol. 10, pp. 34716-34730.
- [12] Loey M., Manogaran G., Khalifa N. E. M. A deep transfer learning model with classical data augmentation and CGAN to detect COVID-19 from chest CT radiography digital images // *Neural Computing and Applications*, 2020.
- [13] Gifani P., Shalhaf A., Vafaezadeh M. Automated detection of COVID-19 using ensemble of transfer learning with deep convolutional neural network based on CT scans // *International Journal of Computer Assisted Radiology and Surgery*, 2021, vol. 16, pp. 115-123.
- [14] Huang L., Ruan S., Denoeux T. Covid-19 classification with deep neural network and belief functions // *5th International Conference on Biological Information and Biomedical Engineering (BIBE2021)*, 2021, pp. 1-4.
- [15] Horry M. J., Chakraborty S., Paul M., Ulhaq A., Pradhan B., Saha M., Shukla N. COVID-19 detection through transfer learning using multimodal imaging data // *IEEE Access*, 2020, vol. 8, pp. 149808-149824.
- [16] Chakraborty N. Brain MRI images for brain tumor detection (2020). Available at : <https://www.kaggle.com/datasets/navoneel/brain-mri-images-for-brain-tumor-detection> (accessed 26 December 2023).
- [17] Yang X., He X., Zhao J., Zhang Y., Zhang S., Xie P. COVID-CT-dataset: a CT scan dataset about COVID-19 // *arXiv*, 2020.
- [18] Liu Z., Mao H., Wu C.-Y., Feichtenhofer C., Darrell T., Xie S. A ConvNet for the 2020s // *arXiv*, 2022.
- [19] Wilson D. L. Asymptotic properties of nearest neighbor rules using edited data // *IEEE Transactions on Systems, Man, and Cybernetics*, 1972, pp. 408-421.



# Предиктивная диагностика отказов энергогенерирующего оборудования на основе нейронной сети

И. В. Некрасов<sup>1</sup>, Ю. Д. Константиновский<sup>2</sup>, Н. С. Кукин<sup>3</sup>

<sup>1</sup>ИПУ РАН им. В.А. Трапезникова

<sup>2</sup>МГТУ им. Н.Э. Баумана

<sup>3</sup>АНО «Институт инженерной физики»

ivannekr@mail.ru

**Аннотация.** В работе рассматривается построение системы диагностики агрегатов и систем теплоэлектростанции на основе полносвязной нейронной сети. Констатируется распространенная на практике ситуация недостаточности статистических данных об аварийных остановах, что не позволяет эффективно обучать систему диагностики по каждому типу отказов. Для исправления ситуации реализуется идея дополнительного использования данных предаварийных режимов работы, присутствующих в измерительных системах в достаточном количестве. Приведен метод формирования и предварительной обработки обучающей выборки на основе предаварийных измерений, включающий в себя нахождение корреляции, определение разброса данных для каждого из параметров и методику выбора наиболее качественных данных для обучения сети. Приведен результат обучения и тестирования системы на реальных данных питающего электронасоса турбоагрегата ТЭЦ, по которым сделаны выводы и предложены дальнейшие действия для улучшения результата.

**Ключевые слова:** теплоэлектростанция; турбоагрегат; питающий электронасос; предиктивная диагностика; нейронные сети; машинное обучение; корреляционный анализ

## I. ВВЕДЕНИЕ

Своевременное обнаружение неисправностей и дефектов в агрегатах и сопутствующих им системах является важным аспектом в обеспечении безопасной, надежной и эффективной работы теплоэлектростанции. В настоящее время разработано значительное количество расчетных методов, позволяющих заранее определить развивающийся дефект, которые можно условно объединить в следующие группы:

- классические предупредительные сигнализации на основании прямых измерений и логических срабатываний [1], используемые в системах управления в течение более 50 последних лет;
- методы, основанные на так называемых экспертных правилах – типовых сочетаниях технологических параметров, характеризующих нештатное функционирование оборудования. Например, [2] для оборудования распределительных сетей в Великобритании, [3] для дизель-генераторного оборудования в США, и многие другие базы дефектов оборудования по

разным типам и отраслям применения, накапливаемые производителями и эксплуатирующими организациями;

- алгоритмы, осуществляющие сравнение текущего состояния объекта с его эталонной математической моделью штатной работы. При этом эталонная модель строится на основе известных физических закономерностей и конструктивных особенностей агрегата [4];
- методы, основанные на классификаторах текущего состояния по степени его «близости» к области возможных штатных/нештатных состояний, наблюдавшихся ранее [5].

Первые три подхода по сути все являются «экспертными» и различаются лишь степенью погружения эксперта в физические процессы агрегата (от простых пороговых наблюдений за критическими параметрами на уровне эксплуатации до полной модели физических процессов в дифференциальных уравнениях). Последний тип методов основывается на чисто «математических» методах обработки ранее измеренных значений наблюдаемых параметров и не использует накопленные инженерные (например, при построении агрегата) и эксплуатационные знания. Методы, основанные на данных, в свою очередь фундаментально делятся на две ветки [6]:

- агрегация данных о штатной работе оборудования и построение системы с общим индикатором «ненормальная работа»;
- поиск паттернов отказов за счет обработки пред-отказных измерений, создание адресных алармов «неисправность конкретного типа».

В идеале, для корректной диагностики оборудования, все три подхода (экспертный и оба математических) необходимо сочетать в единой информационной системе – баланс применения методов экспертного и математического типа в системах диагностики рассмотрен, например, в [7].

В данной работе разрабатывается и апробируется на реальных данных один из математических методов второго типа (на основе обработки измерений и статистики отказов), а именно метод ранней диагностики на основе нейронной сети с использованием перечня

зафиксированных типовых отказов и максимального набора имеющихся телеметрических данных, полученных в предшествующих им периодах времени.

## II. ОПИСАНИЕ ОБЪЕКТА И ОСОБЕННОСТЕЙ ЕГО ДИАГНОСТИРОВАНИЯ

Теплоэлектростанция (далее – ТЭЦ) представляет собой сложную многосвязную систему в классическом определении, данном, например, в [8]: работа агрегатов выстроена в сильно связанный процесс с большим количеством последовательных включений [9], когда непредвиденный выход из строя одного из агрегатов может остановить работу всей станции. Данный факт делает ТЭЦ одним из перспективных объектов для внедрения систем раннего обнаружения неисправностей, которая позволит заблаговременно проводить технические работы, а также замену неисправных агрегатов и узлов на новые. Многосвязность электростанции (энергоблока) как объекта управления [9] естественным образом делает процесс диагностики составным, то есть имеющим многие точки приложения с разбивкой по отдельным агрегатам и узлам, а также типам неисправностей.

В дополнение к вышесказанному, с практической точки зрения, диагностика крупного технологического объекта, такого как ТЭЦ, в виде общего признака «объект неисправен» не несет полезной нагрузки для оператора. Для эффективной работы требуется система автоматической диагностики, которая будет показывать проблемный узел системы и давать предположительный диагноз [10]. Это упрощает ее использование и снижает требования к персоналу, т. к. оператору остается лишь выполнять и передавать рекомендации, отображающиеся у него на экране. Таким образом, при внедрении системы диагностики в рамках настоящей работы необходимо выбрать конкретный агрегат/узел ТЭЦ. Исходя из имеющихся в наличии измерений, без ограничения общности подхода, в качестве объекта исследования использован питательный электронасос (далее – ПЭН), предназначенный для подачи питательной воды к котлоагрегатам электростанции. Данный узел является важным элементом технологической цепочки производства электроэнергии, отказ которого приводит к снижению объема выпуска вплоть до полного выведения энергоблока из работы (рис. 1).

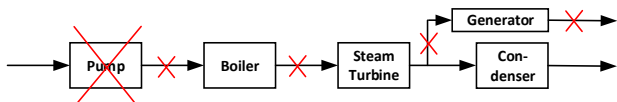


Рис. 1. Отказ ПЭН и его влияние на выработку электроэнергии

Как было сказано выше, в работе реализуется математическое диагностическое правило, основанное на анализе измерений предаварийных состояний и статистики отказов. Выбранной технологией построения диагностического правила выбраны нейронные сети, т. к. технология их обучения на исторических телеметрических данных в настоящее время хорошо отработана [11]. Однако существует ряд трудностей при реализации системы диагностики на основе нейронной сети:

- Низкое качество телеметрических данных, в частности присутствие выбросов и отсутствие значений за некоторый временной период. Данная проблема решается традиционными методами очистки и фильтрации данных [12].
- Неопределенность временного периода развития неисправности/отказа, который напрямую определяет границы выборки данных, используемых для обучения сети. В настоящей работе использовался экспертно заданный фиксированный период развития одни календарные сутки, однако для обеспечения оптимальной работы внедряемой диагностики требуется дополнительное численное моделирование, выходящее за рамки данной статьи.
- Малая выборка по отказам конкретного диагностического оборудования, – как правило, единицы зафиксированных случаев – не позволяющая осуществить полноценное обучение нейросети на типовых событиях [13]. Способ преодоления подобного дефицита данных предложен и реализован авторами данной статьи, описание представлено далее в разделе III.

Плановым результатом настоящей работы является алгоритм, способный обнаружить на ранних стадиях развитие отказа конкретного типа и выдать сигнал предупреждения. При этом ключевым требованием к алгоритму является способность его обучения на малой выборке отказов.

## III. ПОСТРОЕНИЕ ДИАГНОСТИЧЕСКОЙ НЕЙРОСЕТИ В УСЛОВИЯХ НЕДОСТАТОЧНОСТИ ДАННЫХ ПО ОТКАЗАМ

При создании системы диагностики за основу возьмем полносвязную нейронную сеть (далее – ПНС) как наиболее распространенную технологию, широко используемую для решения задачи классификации и работы с многомерными данными [11]. Такая нейронная сеть состоит из трех частей: входной слой (представляет собой набор параметров агрегата, обозначенных на рис. 1 как  $x_n$ ); скрытые слои; выходной слой (обозначен на рис. 1 как  $y_m$  и отображает результат работы сети: в нашем случае, признак наличия или отсутствия неисправности). Связью между слоями являются взвешенные веса  $k_i$ , значения которых определяются во время обучения.

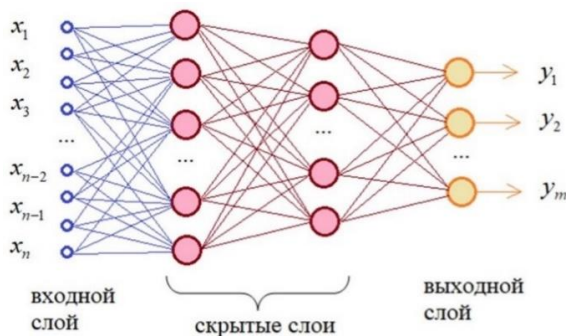


Рис. 2. Пример полносвязной нейронной сети

В качестве входных данных для обучения используются тренды измерений технологических параметров ПЭН за три года, список которых представлен в табл. 1.

ТАБЛИЦА I. СПИСОК ТЕХНОЛОГИЧЕСКИХ ПАРАМЕТРОВ

Наименование параметра	Параметр
Напряжение СК-1, [мВ]	x1
Температура фазы А-В медь, [°C]	x2
Температура фазы А-В железо, [°C]	x3
Температура фазы В-С медь, [°C]	x4
Температура фазы В-С железо, [°C]	x5
Температура фазы С-А медь, [°C]	x6
Температура фазы С-А железо, [°C]	x7
Температура холодного воздуха точка 2, [°C]	x8
Напряжение СК-2, [мВ]	x9
Температура верха корпуса насоса, [°C]	x10
Температура низа корпуса насоса, [°C]	x11
Температура воды на всасе, [°C]	x12
Температура воды в разгрузочной камере, [°C]	x13
Температура масла за маслоохладителем, [°C]	x14
Температура горячего воздуха, [°C]	x15
Температура холодного воздуха точка 1, [°C]	x16
Напряжение СК-3 [мВ]	x17
Температура подшипника №1 - электродвигателя, [°C]	x18
Температура подшипника №2 - электродвигателя, [°C]	x19
Температура подшипника №3 - промежуточного вала, [°C]	x20
Температура подшипника №4 - промежуточного вала, [°C]	x21
Температура подшипника №5 - насоса, [°C]	x22
Температура подшипника №6 - насоса, [°C]	x23
Температура торцевого уплотнения 5-го подшипника, [°C]	x24
Температура торцевого уплотнения 5-го подшипника, [°C]	x25
Расход питательной воды, [т/ч]	x26

В течение обозначенного выше трехлетнего периода было зарегистрировано шесть сбоев в работе агрегата, сведенных в табл. 2. Информация о сбоях содержит наименование сбоя, а также дату его наступления, при этом момент времени самого отказа определяется дополнительно по признакам останова агрегата (например, путем поиска нулевого значения тока амперток). Каждое из указанных в таблице событий отказа можно представить в виде дискретного флага, срабатывающего на временной оси.

ТАБЛИЦА II. ИСТОРИЧЕСКИЕ ДАННЫЕ ПО СОБЫТИЯМ ОТКАЗОВ

Наименование неисправности	Дата наступления неисправности
Повышенная вибрация	08.01.2021
Повышенная вибрация	14.12.2020
Засорение фильтров	26.04.2020
Заклинивание	03.02.2020
Засорение фильтров	29.10.2019
Нагрев подшипника	25.09.2019

В нейросетевой постановке задачи (рис. 1) выходом системы диагностики являются булевы переменные  $y_m$ , представляющие собой флаги аварийных состояний из табл. 2. Заметим, что в табл. 2 на рассматриваемом

историческом горизонте имеются неисправности с одинаковым наименованием, что позволяет сгруппировать их в типизированные флаги, каждый из которых соответствует отдельному типу отказов. При этом множество булевых переменных  $y_m$ , соответствующих каждому типу отказа, целесообразно заменить одной дискретной переменной  $Y$ :

$$\begin{cases} y_1 = \{0 \ 1\} \\ \vdots \\ y_m = \{0 \ 1\} \end{cases} \rightarrow Y = \{0 \ 1 \dots m\} \quad (1)$$

Любое значение  $Y$ , отличное от нуля, означает наличие признака аварийности  $m$ -го типа в соотв. с табл. 3.

ТАБЛИЦА III. СООТВЕТСТВИЯ ЗНАЧЕНИЯ ВЫХОДНОГО СИГНАЛА НЕЙРОСЕТИ ТИПОВОМУ ОТКАЗУ

Наименование неисправности	Значение $Y$ на выходе нейронной сети
Штатная работа агрегата	0
Повышенная вибрация	1
Засорение фильтров	2
Заклинивание	3
Нагрев подшипника	4

Обучение сети производилось с помощью библиотеки «scikit-learn», в частности классом «MLPClassifier», изменяемые параметры которого можно подробно изучить на официальном сайте [14].

Очевидно, что имеющееся количество срабатываний (6 штук, табл. 2) недостаточно для полноценного обучения нейронной сети, т.к. количество срезов данных в обучающей выборке должно многократно превышать размерность входных данных [15]. При таком количестве элементов выборки ни один алгоритм обучения не применим. Однако, решая задачу обучения, стоит повторно отметить, что все сбои и отказы – это развивающиеся во времени процессы, и определяющие их параметры выходят в зону аварийности не мгновенно. Таким образом, срезы измерений параметров, сделанные за некоторое время до аварии, также являются аномальными и могут быть использованы для обучения нейронной сети:

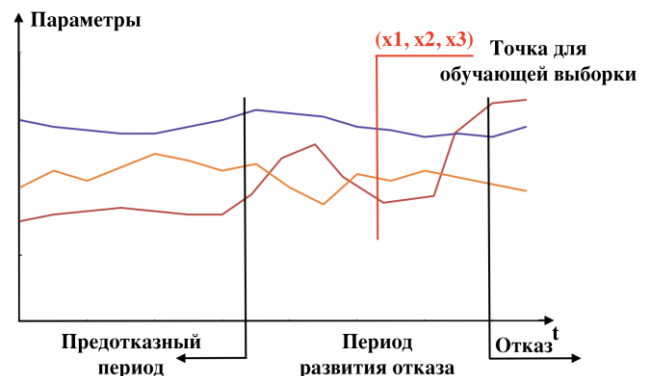


Рис. 3. Использование срезов данных на развивающемся отказе для обучения нейронной сети

В общем случае, длительность развития отказа  $T$  неизвестна и подлежит вычислению опытным путем, либо задается экспертно. В данной работе принято пятисуточное временное окно развития отказа. С учетом

того, что измерения технологических параметров и хранение их в исторической базе осуществляется с посекундным разрешением, общее количество срезов данных для обучения нейронной сети составляет:

$$N_{\text{общ}} = N_{\text{отказов}} \cdot 5_{\text{дней}} \cdot 24_{\text{часа}} \cdot 60_{\text{минут}} \cdot 60_{\text{секунд}} = 2\,592\,000$$

#### IV. ОБУЧЕНИЕ И ТЕСТИРОВАНИЕ НЕЙРОСЕТИ

##### A. Обучение диагностической нейронной сети

Первым шагом к обучению нейронной сети служит предварительная обработка телеметрических входных данных, осуществляемая в два этапа – редукция системы путем исключения взаимно коррелирующих параметров во входной выборке; очистка данных от некачественных периодов (отсутствие измерений, недостоверных участков, выбросов и т. п.). При этом следует заметить, что оба этапа направлены на повышение достоверности получаемой аппроксимации. Очистка позволяет избавиться от так называемых грубых ошибок измерений [16], редукция направлена на избежание захвата шумовых составляющих данных в итоговую нейросетевую модель [17]. Обработка данных по критериям корреляции и качества осуществлена по следующему алгоритму:

- Расчет кросс-корреляционной матрицы, значения в которой вычисляются для пары параметров по формуле Пирсона [18]:

$$r_{n1\ n2} = \frac{\sum_{i=1}^I (x_{n1\ i} - \bar{x}_{n1}) * (x_{n2\ i} - \bar{x}_{n2})}{n \delta_{x_{n1}} \delta_{x_{n2}}}$$

где  $\bar{x}_{n1}, \bar{x}_{n2}$  – среднее арифметическое значение  $n1$ -го и  $n2$ -го параметра входных данных ( $n1=1, \dots, n; n2=1, \dots, n$ );  $\delta_{x_{n1}}, \delta_{x_{n2}}$  – среднеквадратическое отклонение  $n1$ -го и  $n2$ -го параметров входных данных;  $I$  – размерность выборки по каждому из  $n$  входных параметров ( $i=1, \dots, I$ ).

Часть корреляционной матрицы представлена в табл. 4.

ТАБЛИЦА IV. Часть вычисленной корреляционной матрицы

Параметр	x9	x10	x11	x12	x13	x14
x9	1	0,086	0,088	0,083	0,059	0,062
x10	0,086	1	0,999	0,999	0,998	0,001
x11	0,88	0,999	1	0,999	0,999	0,001
x12	0,083	0,999	0,999	1	0,999	0,001
x13	0,059	0,998	0,999	0,999	1	0,001
x14	0,062	0,001	0,001	0,001	0,001	1

- Преобразование корреляционной матрицы в булеву относительно заданного порога корреляции  $p$ :

$$p_{n1\ n2}^{bool} = \begin{cases} 0, & |r_{n1\ n2}| < p \\ 1, & |r_{n1\ n2}| > p \end{cases}; \quad n1, n2 = \overline{1, n}$$

Следует заметить, что закругление элементов матрицы до булевых значений сразу показывает параметры, являющиеся кандидатами на удаление из модели (те параметры, для которых параметр матрицы  $p_{n1\ n2}^{bool}$  равен 1). Часть итоговой булевой матрицы представлена в табл. 5.

ТАБЛИЦА V. Булева матрица относительно порога корреляции параметров

Параметр	x9	x10	x11	x12	x13	x14
x9	1	0	0	0	0	0
x10	0	1	1	1	1	0
x11	0	1	1	1	1	0
x12	0	1	1	1	1	0
x13	0	1	1	1	1	0
x14	0	0	0	0	0	1

- Оценка качества входных данных по каждому  $n$ -му входному параметру на основании разброса, вычисляемого по формуле [16]:

$$R(x_n) = \left| \frac{Me(x_n) - \bar{x}_n}{Me(x_n)} \right|$$

где  $n$  – индекс входной переменной  $x_n$ , для которой осуществляется расчет на основе имеющихся измерений (соответствует индексу столбца в булевой матрице  $p_{n1\ n2}^{bool}$ ),  $\bar{x}_n$  – среднее арифметическое значение ряда измерений  $x_n$ ,  $R(x_n)$  – значение разброса ряда измерений  $x_n$ ,  $Me(x_n)$  – значение медианы ряда измерений  $x_n$ , вычисляемое как срединное значение выборки, упорядоченной по возрастанию. Формула расчета медианы [16]:

$$Me(x_n) = \begin{cases} x_{n, (\frac{I-1}{2}+1)}, & \text{если } \frac{I}{2} \in \mathbb{Z} \\ \frac{x_{n, (\frac{I}{2})} + x_{n, (\frac{I}{2}+1)}}{2}, & \text{если } \frac{I}{2} \notin \mathbb{Z} \end{cases}$$

$I$  – размерность выборки по каждому из  $n$  входных параметров ( $i=1, \dots, I$ ).

Чем меньше вычисленное значение разброса  $R(x_n)$ , тем данные являются более качественными.

- Коррекция булевой матрицы на вычисленное значение разброса  $R$  по формуле:

$$p_{n1\ n2}^R = p_{n1\ n2}^{bool} * R(x_{n2})$$

В итоге получим сводную матрицу, одновременно отображающую коррелирующие параметры и качество данных. Часть указанной матрицы представлена в табл. 6. Среди взаимокоррелирующих параметров выбираются те параметры, качество которых наилучшее (значение в группе наименьшее), такие параметры в табл. 6 отмечены зеленым цветом. Остальные параметры отбрасываются.

ТАБЛИЦА VI. Корреляционная матрица, учитывающая разброс данных

Параметр	x9	x10	x11	x12	x13	x14
x9	0,074	0	0	0	0	0
x10	0	0,442	0,430	0,431	0,563	0
x11	0	0,442	0,430	0,431	0,563	0
x12	0	0,442	0,430	0,431	0,563	0
x13	0	0,442	0,430	0,431	0,563	0
x14	0	0	0	0	0	57,930

При пороге корреляции  $p = 0,8$  из обучающей выборки выбыли 8 параметров (x10, x12, x13, x16, x17, x18, x19, x20) из 26 исходных (см. таблицу 1), в дальнейшем моделировании использованы 18 оставшихся параметров, имеющих наилучшее качество.

Для обучения нейросети используется метод стохастического градиентного спуска [19], поскольку данный метод позволяет эффективно оптимизировать веса узлов нейронной сети и является более быстрым по сравнению с методом обратного распределения ошибок. Особенностью является то, что имеется возможность избежать явления переобучения, задавая параметр  $\alpha$ , регулирующий это явление.

**В. Тестирование нейронной сети на реальных измерениях ПЭН**

Результат работы обученной модели системы диагностики представлен на рис. 4. По горизонтальной расположены предсказанные нейронной сетью события неисправности (красный цвет) и данные, использованные для проверки (синий цвет). Уровень сигнала по вертикальной оси соответствует дискретному признаку аварийности ПЭН, представленному в табл. 3.

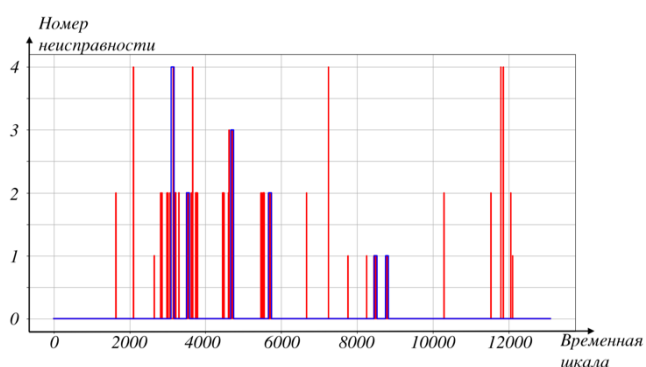


Рис. 4. График, отображающий результат работы системы диагностики

Как видно из графика неисправности хорошо определяются нейронной сетью (совпадение красной и синей линий в левой части рис. 4), однако в работе модели присутствуют ложные тревоги (красная линия в правой части рис. 4). Численные значения представлены в матрице решений в табл. 7.

ТАБЛИЦА VII. МАТРИЦА РЕШЕНИЙ

		Фактическое значение				
		0	1	2	3	4
Предсказанное значение	0	12568	7	56	1	6
	1	109	43	0	0	0
	2	74	1	77	0	0
	3	48	0	0	28	0
	4	68	0	1	0	7

Из табл. 7 вычислим общую точность (ОТ) предсказания каждой неисправности, а также вероятность ложной тревоги (ЛТ).

Вероятность ложной тревоги рассчитывается как отношение суммы чисел неправильно предсказанных состояний к числу правильно предсказанных состояний в столбце с нормальной работой (состояние 0). Расчет:

$$ЛТ = \frac{109 + 74 + 48 + 68}{12568} = 0,0238$$

Общая точность в свою очередь вычисляется как отношение суммы правильно определенных состояний

за вычетом неправильно определенных состояний к сумме правильно определенных состояний. Расчет:

$$ОТ = \frac{(12568 + 43 + 77 + 28 + 7) - 371}{12723} = 0,971$$

Вероятность ЛТ: 0,0238. С таким значением ЛТ система работоспособна, а общая точность предсказания получилась равной, ОТ: 0,971, что является хорошим показателем.

**В. ЗАКЛЮЧЕНИЕ**

Применение нейронной сети в системе диагностики агрегатов и систем на теплоэлектростанциях дало работоспособный результат, о чем свидетельствуют удовлетворительные значения вероятностных параметров точности и ложных срабатываний синтезированной системы, зафиксированные во время тестирования. Настраиваемыми параметрами разработанного нейросетевого алгоритма диагностики являются: окно времени, на котором строится выборка предаварийных состояний при построении нейронной сети, а также порог корреляции для отсеки зашумленных данных. В настоящей статье рассмотрена базовая реализация алгоритма диагностики с удовлетворительными показателями качества. Главным результатом данной работы является преодоление дефицита статистических данных по отказам и обеспечение качественного обучения нейронной сети. Безусловно, при внедрении на реальных объектах модель подлежит уточнению и «тонкой настройке» на имеющихся архивных измерениях.

**СПИСОК ЛИТЕРАТУРЫ**

- [1] Литвак В.И. Автоматическая аварийная защита в системах управления. Москва: Энергия, 1973. 120 с.
- [2] DNO COMMON NETWORK ASSET INDICES METHODOLOGY ([https://www.ofgem.gov.uk/system/files/docs/2017/05/dno\\_common\\_network\\_asset\\_indices\\_methodology\\_v1.1.pdf](https://www.ofgem.gov.uk/system/files/docs/2017/05/dno_common_network_asset_indices_methodology_v1.1.pdf) - дата последнего обращения 07.04.2024).
- [3] B. Mrowca. Emergency diesel generator failure review 1999 - 2001. Technical report, INFORMATION SYSTEMS LABORATORIES, Rockville, September 2011. (<https://www.nrc.gov/docs/ML1125/ML11259A101.pdf> - дата последнего обращения 07.04.2024).
- [4] Прохоров А., Лысачев М. Цифровой двойник. Анализ, тренды, мировой опыт /Под ред. проф. А. Боровкова. Москва: ООО «АльянсПринт», корп. издание Росэнергоатом. 2020. 401 с.
- [5] Некрасов И.В., Правдивец Н.А. Машинное обучение в задачах прогноза отказов оборудования // Математические методы распознавания образов: Тезисы докладов 19-й Всероссийской конференции с международным участием. Москва: Российская академия наук, 2019. 420 с.
- [6] Fu, H., Liu, Y. A deep learning-based approach for electrical equipment remaining useful life prediction. Auton. Intell. Syst. 2, 16 (2022). <https://doi.org/10.1007/s43684-022-00034-2>
- [7] Nekrasov I. Balancing Mathematical Methods and Expert Knowledge in Industrial Equipment Failures Prediction Systems. // “Diagnosing and Predicting Technical State of Electric Power-Plants Equipment” Conference. – All-Russia Thermal Engineering Institute, Moscow, Russia, 2020, pp.55-62. (<https://vti.ru/scientific-event/event-reports/20-maya-2020-goda-v-oao-vti-sostoyalas-on-lajn-mezhdunarodnaya-nauchno-tehnicheskaya-konferenciya-diagnostirovanie-i-prognozirov/soderzhanie-sbornika-dokladov-on-lajn-mezhdunarodnoj-nauchno-tehnicheskoy-konferencii-diagnostirovanie-i-prognozirovanie-tehnich> – in Russian, for full-text papers additional request to organizers on web site is needed)

- [8] Мееров М.В. Системы многосвязного регулирования. Москва: Наука, 1965. 384 с.
- [9] Биленко В.А. Разработка методических основ, исследование и внедрение многосвязанных автоматических систем регулирования теплоэнергетических объектов: дис. в виде научного доклада ... докт. техн. наук / МЭИ, Москва, 2012г.
- [10] ГОСТ Р 53564-2009. Мониторинг состояния оборудования опасных производств. Требования к системам мониторинга.
- [11] I.F. Croall, J.P. Mason (Eds.) Industrial Applications of Neural Networks. ECSC — EEC — EAEC, Brussels — Luxembourg 1992. 312 p.
- [12] Шахтарин Б.И. Оптимальная фильтрация и прогнозирование случайных процессов. Москва: Изд-во МГТУ им. Н.Э. Баумана, 1991. 210 с.
- [13] Кафтаников И.Л., Парасич А.В. Проблемы формирования обучающей выборки в задачах машинного обучения // Вестник ЮУрГУ. Серия «Компьютерные технологии, управление, радиоэлектроника», 2016. Т. 16, № 3. С. 15–24. DOI: 10.14529/ctcr160302
- [14] Scikit-learn reference. Available at: <https://scikit-learn.org/> (accessed 13 March 2024).
- [15] Михайлов А.С., Нехорошкина М.С. Методы определения объема обучающей выборки для искусственной нейронной сети // Информационные технологии и системы 2014 (ИТС 2014): материалы межд. науч. конф., БГУИР, Минск, Беларусь, 29 октября 2014 г. Минск: БГУИР, 2014. С. 106–107.
- [16] Савчук В.П. Обработка результатов измерений. Физическая лаборатория. Ч1: Учеб. пособие для студентов вузов. Одесса: ОНПУ, 2002. 54 с.
- [17] Заенцев И.В. Нейронные сети: основные модели: Учеб. пособие к курсу "Нейронные сети". Воронеж: Воронежский государственный университет. 1999. 76 с.
- [18] Фёрстер Э., Рёнц Б. Методы корреляционного и регрессионного анализа. Руководство для экономистов / пер. с нем. В.М. Ивановой. Москва: «Финансы и статистика», 1983. 303 с.
- [19] Alex Smola and S.V.N. Vishwanathan. Introduction to machine learning. Cambridge University Press, Cambridge, UK, 2008. 234 p. (pp.136-137).

## СОДЕРЖАНИЕ

## Секция 1. Концепции построения гибридного интеллекта

<i>Реализация биоподобных поведенческих функций мобильного робота на сегментной спайковой модели нейрона</i>	
В. В. Иванова (СПбГЭТУ «ЛЭТИ», ЦНИИ РТК), А. М. Корсаков (ЦНИИ РТК), А. А. Демчева (ЦНИИ РТК, СПбПУ), А. В. Бахшиев (СПбПУ) .....	3
<i>Оценка состояния человека на основе данных о его работе на компьютерной клавиатуре</i>	
Даньтин Ма (СПбГЭТУ «ЛЭТИ»).....	7
<i>Транскодирование JPEG-изображений при помощи нейросетевого предсказания коэффициентов ДКП</i>	
И. С. Подцепко, Е. А. Беляев (Университет ИТМО) .....	12
<i>Применение инструментов машинного обучения в системах автоматизированного проектирования технологических процессов (САПР ТП)</i>	
А. А. Лаптев, С. Д. Третьяков (Университет ИТМО) .....	17
<i>Интеллектуальный анализ и прогнозирование динамики промышленного развития регионов России</i>	
Е. С. Митяков, А. И. Ладынин (МИРЭА – Российский технологический университет), С. Н. Митяков, Т. М. Крюкова, Е. А. Назарова (Нижегородский государственный технический университет им. Р.Е. Алексеева) .....	21
<i>Методы машинного обучения для прогнозирования результативности онлайн-курсов</i>	
М. А. Бурнаев, Е. А. Болдырева (Университет ИТМО) .....	25
<i>Генерация трехмерной сцены на основе одного двумерного изображения с помощью глубоких нейронных сетей</i>	
А. С. Сырых (Университет ИТМО), Г. О. Бондаренко (СПбГЭТУ «ЛЭТИ») .....	29
<i>Оценка различных методов машинного обучения для индивидуальной классификации отдельных проб вызванных потенциалов при восприятии правильных и неправильных решений математических примеров</i>	
Ж. В. Нагорнова (ИЭФБ РАН), А. С. Сивцева (Университет ИТМО), Н. В. Шемякина (ИЭФБ РАН) .....	33
<i>Применение больших языковых моделей для выделения намерений в целеориентированных диалоговых системах</i>	
А. Е. Шухман, В. Р. Бадиков, Л. В. Легашев (Оренбургский государственный университет) .....	37
<i>Объединение глубокого обучения и объяснимого ИИ для неинвазивного прогнозирования мутаций EGFR и KRAS в NSCLC: новый радиогеномный подход</i>	
Шариати Фаридоддин, В. А. Павлов, С. В. Федяшина, Н. А. Серебренников (Институт электроники и телекоммуникаций СПбПУ, Санкт-Петербургский научно-исследовательский институт неотложной медицины имени И.И. Джанелидзе) .....	41
<i>Разработка метода выявления дипфейков: применение частотного анализа и уменьшения цветового пространства изображения для повышения точности классификации</i>	
В. Роговой, В. М., Коржук, О. А. Кокорина (Университет ИТМО) .....	45

<i>Распознавание и локализация объектов посредством компьютерного зрения</i>	
В. В. Ковалев, Е. В. Илатовская, Д. М. Филатов, А. О. Федоркова, Д. П. Михайлов (СПбГЭТУ «ЛЭТИ») .....	50
<i>Влияние использования искусственного интеллекта на когнитивные способности школьников: субъективная оценка</i>	
М. А. Сорочинский (Северо-Восточный федеральный университет имени М.К. Аммосова), С. Г. Прохорова (ГАНОУ «Арктическая школа» Республики Саха (Якутия)), К. А. Базанова (Северо-Восточный федеральный университет имени М.К. Аммосова) .....	54
<i>Using an Evolutionary Algorithm such as Genetic Algorithm (Ga) to Optimize Linear Equations</i>	
Abdul sttar Ismail Wdaa (College of Education for pure Science, University of Anbar), Shokhan M. Al-Barzinji (Department of Computer Networks Systems, College of Computer Science and Information Technology, University of Anbar), Hiba A. Tarish (Civil Engineering Departments, University of Technology Baghdad) .....	58
<i>The Impact of Digital Accounting Technologies in Achieving the Quality of Accounting Information</i>	
Ammar Ghazi Ibrahim, Kubra Mohammed Tahir (Mosul University) .....	66
<b>Секция 2. Технологии искусственного интеллекта и их приложения</b>	
<i>Модель компьютерного зрения для анализа загруженности регистратуры КДЦ «НМИЦ им. В.А. Алмазова»</i>	
Г. В. Орлов, А. Н. Калиниченко (СПбГЭТУ «ЛЭТИ») .....	71
<i>Разработка нейросетевого модуля автозаполнения протокола осмотра для модульной медицинской информационной системы</i>	
И. М. Киряков, С. А. Молодяков (СПбПУ Высшая школа программной инженерии) .....	75
<i>Выявление рентгенологической стадии остеоартроза коленного сустава с применением сверточных нейронных сетей</i>	
Н. А. Цыганюк, Е. Ю. Белова (СПбГЭТУ «ЛЭТИ») .....	79
<i>Эффективность выявления лейкемии: сравнительное исследование EfficientNetV3 и EfficientNetV5</i>	
Асиль Алшорайхы, Хусам Хасан Бу Исса, Анагим Ибрагим, Agbonrhienrhien Osazee Osca (СПбГЭТУ «ЛЭТИ»).....	82
<i>Моделирование теплопроводности черного фосфорена методом классической молекулярной динамики с использованием потенциала, полученного на основе глубокого обучения сверточной нейронной сети</i>	
Д. В. Завьялов, Д. Н. Жариков (Волгоградский государственный технический университет), В. И. Конченков (Волгоградский государственный технический университет, Волгоградский государственный социально-педагогический университет), Д. В. Шеин (Волгоградский государственный технический университет) .....	86
<i>Использование искусственного интеллекта и технологии перовскитных материалов для мониторинга светового загрязнения</i>	
Д. Ю. Терехов, А. Р. Попова, Е. В. Патока, В. Д. Бурлака, М. С. Козлов (СПбГЭТУ «ЛЭТИ») .....	91
<i>3D-сегментация опухоли головного мозга с интерполяцией с использованием глубоких нейронных сетей для последующей 3D-визуализации</i>	
А. С. Сырых (Университет ИТМО), Г. О. Бондаренко (СПбГЭТУ «ЛЭТИ») .....	95



<i>Применение модуля внимания в сверточной нейронной сети в задаче анализа ОКТ-снимков</i> Е. Н. Волков (ФИЦ «Информатика и управление» РАН) .....	99
<i>Перспективы использования нейронных сетей для диагностики заболеваний сердечно-сосудистой системы</i> Е. А. Соколова (Лаборатория физиологии биоуправления ОЭФ ФГБНУ «Институт экспериментальной медицины») .....	103
<i>Применение U-Net для выделения рядов виноградника по данным дистанционного зондирования Земли</i> С. А. Шквиро, М. Д. Поляк (ГУАП) .....	106
<i>Глубокое обучение для всестороннего улучшения состояния сетчатки: помимо автоматизированной фундоскопии – классификация, сегментация и реконструкция изображения сетчатки</i> Ali Sultan Maуuа (СПбГЭТУ «ЛЭТИ») .....	110
<i>Проверка наличия музыкального образования по фотографиям лиц с помощью сверточных нейронных сетей и Vision Transformer</i> Я. О. Сениченкова, М. Д. Поляк (ГУАП) .....	115
<i>Особенности нейронных сетей в задаче прогнозирования уровня аварийности на дорогах общего пользования</i> А. Ю. Абалаяев, Л. В. Грунская (Владимирский государственный университет имени Александра Григорьевича и Николая Григорьевича Столетовых) .....	119
<i>Оценка выполнения физических упражнений методами глубокого обучения</i> М. Д. Кораблева, Я. А. Бекенева (СПбГЭТУ «ЛЭТИ») .....	123
<i>Применение гистограммы длин хорд с целью повышения точности распознавания лёгочных узлов на компьютерных томограммах</i> Д. С. Смирнова (Лаборатория анализа биомедицинских изображений и данных СПбПУ, Высшая школа прикладной математики и вычислительной физики СПбПУ), Е. И. Пчицкая (Лаборатория анализа биомедицинских изображений и данных СПбПУ), В. С. Чуканов (Лаборатория анализа биомедицинских изображений и данных СПбПУ, Высшая школа прикладной математики и вычислительной физики СПбПУ) .....	127
<i>Модель оценки качества воды, основанная на AI: данные длительного наблюдения</i> Е. А. Алтай, Л. Базарбай (Satbayev University) .....	132
<i>Measuring the Impact of Strategic Vigilance in Strategic Intelligence</i> Zafar Nasser Hussain (Business Administration Dept. University of Babylon), Riyadh Hussein Obyes, Alaa Hussein Fadhil, Ahmed Abdullah Amanah, Mohammed Faez Hasan (Business Administration Dept. University of Kerbala) .....	136

### **Секция 3. Управление данными и организация вычислений в интеллектуальных системах**

<i>Кластеризация нестационарных временных рядов</i> М. А. Калмыков (СПбГЭТУ «ЛЭТИ») .....	142
<i>Организация процедуры обнаружения действий в видеоданных с использованием парадигмы MLOps</i> М. С. Куприянов, Ю. А. Шичкина, С. Е. Ильин (СПбГЭТУ «ЛЭТИ») .....	146

<i>Personal Document Security Tool for Cloud Application</i> Bdoor Alaa Mahmood (College of Computer Science and Information Technology, University of Kerbala) .....	150
<i>Bayesian Analysis of Zero-Inflated Count Data Using Beta-Lindley Distribution</i> Atheer Ismael Fadhel (University Mustansiriyah), Hossein Jabbari Khamnei (University Tabriz) ....	155
<i>Моделирование дискретизации весов в нейронной сети на основе оптимизированных нанокондитивных мемристоров</i> А. Н. Мацукатова (Национальный исследовательский центр «Курчатовский институт», МГУ), А. Д. Трофимов (Национальный исследовательский центр «Курчатовский институт», НИУ МФТИ), М. А. Рябова (НИУ МФТИ), А. А. Несмелов (Национальный исследовательский центр «Курчатовский институт»), А. В. Емельянов (Национальный исследовательский центр «Курчатовский институт», НИУ МФТИ) .....	160
<i>Отказоустойчивый нейросетевой датчик влажности материала</i> С. В. Артемова, А. И. Ладынин, А. Г. Шмелева, Н. С. Ершов (МИРЭА – Российский технологический университет), М. А. Каменская (Тамбовский государственный технический университет), Д. Н. Лапаев (Нижегородский государственный технический университет им. Р.Е. Алексеева) .....	164
<i>Коллективные классификаторы повторного входа</i> А. Ю. Дорогов (СПбГЭТУ «ЛЭТИ») .....	168
<i>Применение сверточных нейронных сетей для классификации ЭЭГ/ВП сигналов единичных проб объединённой выборки данных испытуемых в ответ на предъявление математических примеров</i> М. О. Петров (СПбГЭТУ «ЛЭТИ»), Ж. В. Нагорнова (ИЭФБ РАН), Н. В. Андреева (СПбГЭТУ «ЛЭТИ»), Н. В. Шемякина (ИЭФБ РАН) .....	172
<i>Прогнозирование неисправностей с использованием данных искусственного интеллекта и технологии перовскитовых материалов</i> Д. Ю. Терехов, А. Р. Попова, Е. В. Патока, В. Д. Бурлака, М. С. Козлов (СПбГЭТУ «ЛЭТИ») .....	178
<i>Оптимизация генеративного ридера в составе RAG системы</i> А. П. Соколов (МГУ), П. А. Замелин (НИУ ВШЭ), Ю. Д. Камелина, П. И. Пластова (Нижегородский государственный университет им. Н.И. Лобачевского) .....	182
<i>Аппаратная реализация АРТ-2 классификатора</i> О. И. Буренева, Я. Б. Ибрагим (СПбГЭТУ «ЛЭТИ»), М. С. Прасад, Ш. Верма (Университет Эмити. Институт космической науки и технологий Эмити) .....	186
<i>Trajectory Analysis of a Swarm of 3U CubeSats in Sun-Synchronous Orbit with Varying RAAN and Non-Overlapping Revisit Time</i> Shivani Verma, M. S. Prasad (Amity Institute of Space Science & Technology, Amity University) .....	190
<i>Performance Evaluation of Ramanujan Transform with OMP and BP Algorithms</i> Shivani Verma, M. S. Prasad (Amity Institute of Space Science & Technology, Amity University), Olga I. Bureneva, Yulia A. Shichkina (Saint Petersburg Electrotechnical University “LETI”) .....	196
<i>Оценка влияния снижения размерности пространства признаков на эффективность классификации движений на основе поверхностной электромиографии</i> И. В. Кузнецов, Ю. В. Пономарчук (Дальневосточный государственный университет путей сообщения) .....	203

<i>Оценка производительности оптимизатора в сверточных нейронных сетях для классификации микрокальцинатов молочной железы: сравнительное исследование</i>	
Э. Ч. Угвуеке, Д. Палоганнидис (СПбГЭТУ «ЛЭТИ») .....	207
<i>Применение алгоритма стекинга для идентификации отклонений в данных газотранспортной сети</i>	
А. К. Петрова, С. Е. Абрамкин (СПбГЭТУ «ЛЭТИ») .....	211
<i>Применение методов машинного обучения для приведения параметров работы ГТУ для диагностики турбокомпрессора</i>	
В. С. Карелин, И. В. Попов, П. В. Соколов (СПбГЭТУ «ЛЭТИ») .....	215
<i>Обнаружение аномалий на медицинских изображениях с использованием трансферного обучения и оптимизации данных</i>	
Я. А. Низамли, В. В. Фадел, А. Ю. Филатов, Ю. А. Шичкина (СПбГЭТУ «ЛЭТИ») .....	219
<i>Предиктивная диагностика отказов энергогенерирующего оборудования на основе нейронной сети</i>	
И. В. Некрасов (ИПУ РАН им. В.А. Трапезникова), Ю. Д. Константиновский (МГТУ им. Н.Э. Баумана), Н. С. Кукин (АНО «Институт инженерной физики») .....	224

V Международная конференция  
по нейронным сетям и нейротехнологиям (NeuroNT'2024)

20 июня 2024  
Санкт-Петербург

---

Подписано в печать 19.06.2024. Формат 60×84 1/8.  
Бумага офсетная. Печать ризограф. Печ. л. 29,25.  
Тираж 25 экз.  
Компьютерная верстка, макет – ИП Яременко С.С.  
Печать ИП Яременко С.С.

***NEURONT.ETU.RU***

197022, Россия,  
Санкт-Петербург,  
ул. Профессора Попова,  
д.5, лит. Ф,  
СПбГЭТУ «ЛЭТИ»  
Тел.: +7 (812) 346-46-37  
E-mail: [IRVC.eltech@mail.ru](mailto:IRVC.eltech@mail.ru)

