

ISSN 2713-3192
DOI 10.15622/ia.2025.24.6
<http://ia.spcras.ru>

ТОМ 24 № 6

ИНФОРМАТИКА И АВТОМАТИЗАЦИЯ

INFORMATICS AND AUTOMATION



СПб ФИЦ РАН

Санкт-Петербург
2025



INFORMATICS AND AUTOMATION

Volume 24 № 6, 2025

Scientific and educational journal primarily specialized in computer science, automation, robotics, applied mathematics, interdisciplinary research

Founded in 2002

Founder and Publisher

St. Petersburg Federal Research Center of the Russian Academy of Sciences (SPC RAS)

Editor-in-Chief

A. L. Ronzhin, Prof., Dr. Sci., St. Petersburg, Russia

Editorial Council

A. A. Ashimov	Prof., Dr. Sci., Academician of the National Academy of Sciences of the Republic of Kazakhstan, Almaty, Kazakhstan
I. A. Kalyaev	Prof., Dr. Sci., Academician of RAS, Taganrog, Russia
Yu. A. Merkuryev	Prof., Dr. Sci., Academician of the Latvian Academy of Sciences, Riga, Latvia
A. I. Rudskoi	Prof., Dr. Sci., Academician of RAS, St. Petersburg, Russia
V. Sgurev	Prof., Dr. Sci., Academician of the Bulgarian Academy of Sciences, Sofia, Bulgaria
B. Ya. Sovetov	Prof., Dr. Sci., Academician of RAE, St. Petersburg, Russia
V. A. Soyfer	Prof., Dr. Sci., Academician of RAS, Samara, Russia

Editorial Board

E. Azarov	Prof., Dr. Sci., Minsk, Belarus
O. Yu. Gusikhin	Ph. D., Dearborn, USA
V. Delic	Prof., Dr. Sci., Novi Sad, Serbia
A. Dolgui	Prof., Dr. Sci., St. Etienne, France
M. N. Favorskaya	Prof., Dr. Sci., Krasnoyarsk, Russia
M. Zelezny	Assoc. Prof., Ph.D., Plzen, Czech Republic
H. Kaya	Assoc. Prof., Ph.D., Utrecht, Netherlands
A. A. Karpov	Assoc. Prof., Dr. Sci., St. Petersburg, Russia
S. V. Kuleshov	Dr. Sci., St. Petersburg, Russia
A. D. Khomonenko	Prof., Dr. Sci., St. Petersburg, Russia
D. A. Ivanov	Prof., Dr. Habil., Berlin, Germany
K. P. Markov	Assoc. Prof., Ph.D., Aizu, Japan
R. V. Meshcheryakov	Prof., Dr. Sci., Moscow, Russia
N. A. Moldovian	Prof., Dr. Sci., St. Petersburg, Russia
V. V. Nikulin	Prof., Ph.D., New York, United States
V. Yu. Osipov	Prof., Dr. Sci., Deputy Editor-in-Chief, St. Petersburg, Russia
V. K. Pshikhopov	Prof., Dr. Sci., Taganrog, Russia
H. Samani	Assoc. Prof., Ph.D., Plymouth, UK
J. Savage	Assoc. Prof., Ph.D., Mexico City, Mexico
M. Secujski	Assoc. Prof., Ph.D., Novi Sad, Serbia
A. V. Smirnov	Prof., Dr. Sci., St. Petersburg, Russia
B. V. Sokolov	Prof., Dr. Sci., St. Petersburg, Russia
L. V. Utkin	Prof., Dr. Sci., St. Petersburg, Russia
L. B. Sheremetov	Assoc. Prof., Dr. Sci., Mexico, Mexico

Editor: A. S. Viktorova

Interpreter: Ya. N. Berezina

Art editor: N. A. Dormidontova

Editorial office address

SPC RAS, 39 litera A , 14-th line V.O., St. Petersburg, 199178, Russia

e-mail: ia@sprcras.ru, web: <http://ia.sprcras.ru>

The journal is indexed in Scopus

The journal is published under the scientific-methodological supervision of Department for Nanotechnologies and Information Technologies of the Russian Academy of Sciences

© St. Petersburg Federal Research Center of the Russian Academy of Sciences, 2025

ИНФОРМАТИКА И АВТОМАТИЗАЦИЯ

Том 24 № 6, 2025

Научный, научно-образовательный журнал с базовой специализацией
в области информатики, автоматизации, робототехники, прикладной математики
и междисциплинарных исследований.

Журнал основан в 2002 году

Учредитель и издатель

Федеральное государственное бюджетное учреждение науки
«Санкт-Петербургский Федеральный исследовательский центр Российской академии наук»
(СПб ФИЦ РАН)

Главный редактор

А. Л. Ронжин, д-р техн. наук, проф., Санкт-Петербург, РФ

Редакционный совет

А. А. Ашимов	академик Национальной академии наук Республики Казахстан, д-р техн. наук, проф., Алматы, Казахстан
И. А. Каляев	академик РАН, д-р техн. наук, проф., Таганрог, РФ
Ю. А. Меркурьев	академик Латвийской академии наук, д-р, проф., Рига, Латвия
А. И. Рудской	академик РАН, д-р техн. наук, проф., Санкт-Петербург, РФ
В. Сгурев	академик Болгарской академии наук, д-р техн. наук, проф., София, Болгария
Б. Я. Советов	академик РАО, д-р техн. наук, проф., Санкт-Петербург, РФ
В. А. Соифер	академик РАН, д-р техн. наук, проф., Самара, РФ

Редакционная коллегия

И. С. Азаров	д-р техн. наук, проф., Минск, Беларусь
О. Ю. Гусихин	д-р наук, Диаборн, США
В. Делич	д-р техн. наук, проф., Нови-Сад, Сербия
А. Б. Долгий	д-р наук, проф. Сент-Этьен, Франция
М. Железны	д-р наук, доцент, Пльзень, Чешская республика
Д. А. Иванов	д-р экон. наук, проф., Берлин, Германия
Х. Кайя	д-р наук, доцент, Утрехт, Нидерланды
А. А. Карпов	д-р техн. наук, доцент, Санкт-Петербург, РФ
С. В. Кулешов	д-р техн. наук, Санкт-Петербург, РФ
К. П. Марков	д-р наук, доцент, Аизу, Япония
Р. В. Мещеряков	д-р техн. наук, проф., Москва, РФ
Н. А. Молдовян	д-р техн. наук, проф., Санкт-Петербург, РФ
В. В. Никулин	д-р наук, проф., Нью-Йорк, США
В. Ю. Осипов	д-р техн. наук, проф., зам. главного редактора, Санкт-Петербург, РФ
В. Х. Пшихопов	д-р техн. наук, проф., Таганрог, РФ
Х. К. Саваж	д-р техн. наук, доцент, Мехико, Мексика
Х. Самани	д-р наук, доцент, Плимут, Соединенное Королевство
М. Сечуиски	д-р техн. наук, доцент, Нови-Сад, Сербия
А. В. Смирнов	д-р техн. наук, проф., Санкт-Петербург, РФ
Б. В. Соколов	д-р техн. наук, проф., Санкт-Петербург, РФ
Л. В. Уткин	д-р техн. наук, проф., Санкт-Петербург, РФ
М. Н. Фаворская	д-р техн. наук, проф., Красноярск, РФ
А. Д. Хомоненко	д-р техн. наук, проф., Санкт-Петербург, РФ
Л. Б. Шереметов	д-р техн. наук, Мехико, Мексика

Выпускающий редактор: А. С. Викторова

Переводчик: Я. Н. Березина

Художественный редактор: Н. А. Дормидонтова

Адрес редакции

14-я линия В.О., д. 39, лит. А, г. Санкт-Петербург, 199178, Россия

e-mail: ia@spcras.ru, сайт: <http://ia.spcras.ru>

Журнал индексируется в международной базе данных Scopus

Журнал входит в «Перечень ведущих рецензируемых научных журналов и изданий, в которых должны быть опубликованы основные научные результаты диссертации на соискание ученой степени доктора и кандидата наук»

Журнал выпускается при научно-методическом руководстве Отделения нанотехнологий и информационных технологий Российской академии наук

© Федеральное государственное бюджетное учреждение науки

«Санкт-Петербургский Федеральный исследовательский центр Российской академии наук», 2025
Разрешается воспроизведение в прессе, а также сообщение в эфир или по кабелю опубликованных в составе печатного периодического издания - журнала «ИНФОРМАТИКА И АВТОМАТИЗАЦИЯ» статей по текущим экономическим, политическим, социальным и религиозным вопросам с обязательным указанием имени автора статьи и печатного периодического издания журнала «ИНФОРМАТИКА И АВТОМАТИЗАЦИЯ»

CONTENTS

Preface. In Memory of Rafael M. Yusupov	1537

Mathematical Modeling and Applied Mathematics

N. Kuznetsov, K. Semenikhin NONLINEAR FILTERING METHOD IN THE PRESENCE OF UNKNOWN NOISE INTENSITY	1544

V. Soifer, Y. Goshin VISUAL ODOMETRY WITH TRAJECTORY CORRECTION BASED ON KALMAN FILTER	1568

Artificial Intelligence, Knowledge and Data Engineering

S. Smagin, Y. Ozaryan, V. Usikov, A. Smagin, K. Tsygulev, R. Sekrieru, S. Orlov, T. Kozhevnikova AUTOMATION OF REMOTE ENVIRONMENTAL MONITORING OF MINERAL RESOURCE COMPLEX FACILITIES IN Khabarovsk Krai	1587

A. Fedorov, I. Datyev, I. Vishnyakov THE METHOD FOR INTEGRATING LARGE LANGUAGE MODELS INTO ALGORITHMS FOR FOCUSED MONITORING OF OPEN SOCIAL MEDIA DATA	1623

V. Osipov STABILITY OF INTELLIGENT SIGNAL PROCESSING BY STREAMING RECURRENT NEURAL NETWORKS WITH CONTINUOUS LEARNING	1649

M. Dolgushin, A. Karpov ANALYTICAL REVIEW OF SPEECH AND MULTIMODAL METHODS FOR COGNITIVE IMPAIRMENTS RECOGNITION IN PEOPLE	1683

S. Karpovich, A. Smirnov, N. Teslya METHODOLOGY FOR CREATING A BENCHMARK TO EVALUATE LLM PERFORMANCE ON NUMERALS	1721

Information Security

I. Kotenko, I. Saenko, O. Lauta, V. Sadovnikov, E. Ichetovkin, W. Li ANALYSIS OF MODERN RESEARCH ON PROTECTION AGAINST ADVERSARIAL ATTACKS IN ENERGY SYSTEMS	1751

A. Moldovyan, D. Moldovyan, A. Kostina RANDOMIZATION IN POST-QUANTUM DIGITAL SIGNATURE ALGORITHMS WITH A SECRET GROUP	1810

Robotics, Automation and Control Systems

A. Saveliev, D. Anikin METHOD OF CALCULATING CAPSULE-SHAPED AIR CORRIDORS OF SAFE ROUTES FOR A GROUP OF UNMANNED AERIAL VEHICLES	1836

D. Levonevskiy, A. Motienko, A. Bogomolov AUTOMATION OF CYBER-PHYSICAL MONITORING OF PATIENTS' CONDITION IN SMART WARDS	1865

СОДЕРЖАНИЕ

Предисловие. Памяти Юсупова Рафаэля Мидхатовича	1537
Математическое моделирование и прикладная математика	
Н.А. Кузнецов, К.В. Семенихин МЕТОД НЕЛИНЕЙНОЙ ФИЛЬТРАЦИИ ПРИ НЕИЗВЕСТНОЙ ИНТЕНСИВНОСТИ ШУМА В НАБЛЮДЕНИЯХ	1544
В.А. Сойфер, Е.В. Гошин ВИЗУАЛЬНАЯ ОДОМЕТРИЯ С КОРРЕКЦИЕЙ ТРАЕКТОРИИ НА ОСНОВЕ ПРИМЕНЕНИЯ ФИЛЬТРА КАЛМАНА	1568
Искусственный интеллект, инженерия данных и знаний	
С.И. Смагин, Ю.А. Озарян, В.И. Усиков, А.С. Смагин, К.С. Цыгулёв, Р.А. Секриеру, С.А. Орлов, Т.В. Кожевникова АВТОМАТИЗАЦИЯ ДИСТАНЦИОННОГО ЭКОЛОГИЧЕСКОГО МОНИТОРИНГА ОБЪЕКТОВ МИНЕРАЛЬНО-СЫРЬЕВОГО КОМПЛЕКСА ХАБАРОВСКОГО КРАЯ	1587
А.М. Федоров, И.О. Датъев, И.Г. Вишняков МЕТОД ИНТЕГРАЦИИ БОЛЬШИХ ЯЗЫКОВЫХ МОДЕЛЕЙ В АЛГОРИТМЫ ФОКУСИРОВАННОГО МОНИТОРИНГА ОТКРЫТЫХ ДАННЫХ СОЦИАЛЬНЫХ МЕДИА	1623
В.Ю. Осипов УСТОЙЧИВОСТЬ ИНТЕЛЛЕКТУАЛЬНОЙ ОБРАБОТКИ СИГНАЛОВ ПОТОКОВЫМИ РЕКУРРЕНТНЫМИ НЕЙРОННЫМИ СЕТЯМИ С НЕПРЕРЫВНЫМ ОБУЧЕНИЕМ	1649
М.Д. Долгушин, А.А. Карпов АНАЛИТИЧЕСКИЙ ОБЗОР РЕЧЕВЫХ И МНОГОМОДАЛЬНЫХ МЕТОДОВ РАСПОЗНАВАНИЯ КОГНИТИВНЫХ НАРУШЕНИЙ ЛЮДЕЙ	1683
С.Н. Карпович, А.В. Смирнов, Н.Н. Тесля МЕТОДОЛОГИЯ ПОСТРОЕНИЯ ЭТАЛОННОГО ТЕСТА ДЛЯ ОЦЕНКИ РАБОТЫ LLM С ЧИСЛИТЕЛЬНЫМИ	1721
Информационная безопасность	
И.В. Котенко, И.Б. Саенко, О.С. Лаута, В.Е. Садовников, Е.А. Ичетовкин, В. Ли АНАЛИЗ СОВРЕМЕННЫХ ИССЛЕДОВАНИЙ ПО ЗАЩИТЕ ОТ СОСТЯЗАТЕЛЬНЫХ АТАК В ЭНЕРГЕТИЧЕСКИХ СИСТЕМАХ	1751
А.А. Молдовян, Д.Н. Молдовян, А.А. Костина РАНДОМИЗАЦИЯ В ПОСТКВАНТОВЫХ АЛГОРИТМАХ ЭЦП С СЕКРЕТНОЙ ГРУППОЙ	1810
Робототехника, автоматизация и системы управления	
А.И. Савельев, Д.А. Аникин МЕТОД РАСЧЕТА КАПСУЛООБРАЗНЫХ ВОЗДУШНЫХ КОРИДОРОВ БЕЗОПАСНЫХ МАРШРУТОВ ГРУППЫ БЕСПИЛОТНЫХ ЛЕТАТЕЛЬНЫХ АППАРАТОВ	1836
Д.К. Левоневский, А.И. Мотиенко, А.В. Богомолов АВТОМАТИЗАЦИЯ КИБЕРФИЗИЧЕСКОГО МОНИТОРИНГА СОСТОЯНИЯ ПАЦИЕНТОВ В УМНЫХ ПАЛАТАХ	1865

ПРЕДИСЛОВИЕ ПАМЯТИ ЮСУПОВА РАФАЭЛЯ МИДХАТОВИЧА



Данный тематический выпуск – Том 24 № 6 2025 г. посвящен памяти выдающегося ученого в области информатики, информационных технологий и теории управления, основателя и руководителя научных школ по теоретическим основам информатизации общества и по теории чувствительности сложных информационно-управляющих систем, руководителя научного направления Санкт-Петербургского института информатики и автоматизации Российской академии наук (СПИИРАН) – СПб ФИЦ РАН, директора СПИИРАН (1991-2018 гг.), доктора технических наук, члена-корреспондента РАН, заслуженного деятеля науки и техники Российской Федерации Юсупова Рафаэля Мидхатовича.

Юсупов Рафаэль Мидхатович родился 17 июля 1934 г. в г. Казани. После окончания в 1952 г. с золотой медалью Казанской спецшколы ВВС Юсупов Р.М. был направлен в Ленинградскую военно-воздушную инженерную академию (ныне Военно-космическая академия имени А.Ф. Можайского), которую окончил с отличием в 1958 г. по специальности «инженер-электрик». В 1964 г. стал выпускником Ленинградского государственного университета по специальности «математика».

С 1958 г. по 1989 г. с перерывом в 1985 г. Юсупов Р.М. проходил военную службу в Военной академии имени А.Ф. Можайского, активно занимаясь организационной, научной и педагогической деятельностью. В 1985 г. исполнял должность начальника Направления моделирования стратегических операций

Центра оперативно-стратегических исследований Генерального Штаба Вооруженных Сил СССР.

В 1989 г. Юсупов Р.М. назначен с оставлением в кадрах Вооруженных Сил заместителем директора по научной работе Ленинградского института информатики и автоматизации АН СССР (с 1991 г. – Санкт-Петербургский институт информатики и автоматизации РАН), а в 1991 г. избран директором этого института и проработал в этой должности до 2018 г. С 2018 г. Юсупов Р.М. являлся руководителем научного направления СПИИРАН.

Юсуповым Р.М. подготовлено 11 докторов наук и 46 кандидатов наук, он автор более 500 научных трудов, в том числе более 35 монографий и 19 изобретений.

Важной отличительной особенностью Юсупова Р.М. выступало внимательное и доброжелательное отношение к людям.

Ушел из жизни Юсупов Р.М. 7 ноября 2024 г. Могила Юсупова Р.М. находится на Смоленском православном кладбище в Санкт-Петербурге.

Опубликованные в данном выпуске статьи содержат теплые слова, воспоминания коллег, учеников и сотрудников, вдохновленных идеями и напутствиями Юсупова Р.М.

Доктор технических наук, профессор, советник РАН, академик РАН Кузнецов Николай Александрович замечает: «Идея описанного метода фильтрации возникла у одного из авторов в те годы, когда ему посчастливилось сотрудничать с Юсуповым Рафаэлем Мидхатовичем в области разработки алгоритмов автоматического управления для технических систем общего и специального назначения. Памяти этого выдающегося ученого и неординарного человека посвящается данная статья».

Доктор технических наук, профессор, президент Самарского национального исследовательского университета имени академика С.П. Королёва, лауреат Государственной премии Российской Федерации в области науки и техники, академик РАН Сойфер Виктор Александрович размышляет: «Для авторов большая честь участвовать данной статьёй в выпуске журнала, посвящённом памяти выдающегося учёного, члена-корреспондента РАН Юсупова Рафаэля Мидхатовича. Тематика статьи относится к научным интересам Юсупова Р.М., который проводил исследования в области теории управления полётом летательных аппаратов, теории адаптивных систем.

С Самарой и Самарским университетом имени академика С.П. Королева Рафаэля Мидхатовича связывали многолетние

творческие и дружеские отношения. Будучи волжанином, Юсупов Р.М. часто бывал в Самаре, принимал участие в научных конференциях, общался с профессорско-преподавательским составом, научными работниками и студентами. Творческое наследие Юсупова Р.М. способствует развитию научных исследований, а обаяние его личности остаётся в нашей памяти».

Доктор физико-математических наук, профессор, главный научный сотрудник Вычислительного центра Дальневосточного отделения Российской академии наук, академик РАН Смагин Сергей Иванович замечает: «Работа посвящена памяти ушедшего из жизни Юсупова Рафаэля Мидхатовича – выдающегося ученого в области информатики, информационных технологий и теории управления, основателя и руководителя научных школ по теоретическим основам информатизации общества и по теории чувствительности сложных информационно-управляющих систем, доктора технических наук, профессора, член-корреспондента РАН, заслуженного деятеля науки и техники Российской Федерации, руководителя научного направления СПИИРАН Санкт-Петербургского Федерального исследовательского центра Российской академии наук, директора СПИИРАН (1991-2018 гг.), главного редактора журнала «Информатика и автоматизация».

Доктор технических наук, профессор, директор СПИИРАН Осипов Василий Юрьевич отмечает: «Следуя принципам научной школы члена-корреспондента РАН Юсупова Р.М., проведен анализ возможных решений по обеспечению устойчивой интеллектуальной обработки сигналов потоковыми импульсными РНС с непрерывным обучением. К наиболее перспективным подходам по обеспечению этой устойчивости отнесены использование неперекрывающихся и ортогональных представлений обрабатываемых в РНС сигналов, а также другие методы на основе регуляризации».

Доктор технических наук, профессор, руководитель лаборатории интегрированных систем автоматизации Смирнов Александр Викторович отмечает: «Авторы посвящают настоящую статью памяти доктора технических наук, члена-корреспондента Российской академии наук, заслуженного деятеля науки и техники Российской Федерации Юсупова Рафаэля Мидхатовича. Его выдающиеся результаты внесли значительный вклад в развитие отечественной научной школы и формирование современного научного мышления во многих областях информатики, информационных технологий и теории управления. Научное наследие Юсупова Р.М. продолжает служить основой для дальнейших

исследований и остаётся важной частью достижений отечественной и мировой науки и техники».

Доктор технических наук, профессор, главный научный сотрудник лаборатории проблем компьютерной безопасности Молдовян Александр Андреевич отмечает: «Становление нового направления в области постквантовой криптографии, связанного с разработкой алгебраических алгоритмов цифровой подписи, стало возможным благодаря вниманию и поддержке выдающегося ученого в области информатики, информационных технологий и теории управления, члена-корреспондента РАН, заслуженного деятеля науки и техники РФ Юсупова Рафаэля Мидхатовича, а также многих других инициативных начинаний сотрудников СПИИРАН, работающих в области криптографии. В сердцах авторов и его ближайших коллег благодарная память об этом останется навсегда».

Доктор технических наук, профессор, руководитель лаборатории речевых и многомодальных интерфейсов Карпов Алексей Анатольевич выразил глубокую признательность к позитивному качеству Рафаэля Мидхатовича поддерживать новые научные направления, развитие и инициативы молодых ученых и отмечает: «Группа речевой информатики, основанная доктором технических наук Косаревым Ю.А., при поддержке Юсупова Р.М. в 2008 году стала полноценной лабораторией речевых и многомодальных интерфейсов, из которой впоследствии выделилась новая лаборатория автономных робототехнических систем. Сейчас обе лаборатории продолжают академические традиции преемственности и эволюционного развития, заложенные Рафаэлем Мидхатовичем в СПИИРАН.

Мы хотели бы посвятить эту работу памяти Юсупова Рафаэля Мидхатовича, который всегда активно поддерживал направления исследований и разработок лаборатории речевых и многомодальных интерфейсов СПИИРАН – СПб ФИЦ РАН. Рафаэль Мидхатович также являлся одним из основоположников становления и развития в России новой научной области – интеллектуальных многомодальных человеко-машинных интерфейсов и систем. Светлая ему память».

Доктор технических наук, профессор, заслуженный деятель науки и техники Российской Федерации Соколов Б.В. отмечает: «Первый раз я встретился с Юсуповым Р.М. в 1972 г., когда проходил обучение в ВКА им. А.Ф. Можайского и был курсантом-слушателем данной академии. В то время д.т.н., профессор, инженер-полковник Юсупов Р.М. руководил кафедрой «Прикладной математики» и проводил занятия с будущими военными инженерами-баллистиками. Его лекции отличались большой наглядностью и широтой, что

позволило уже в те годы нам понять необходимость рассмотрения всех существующих и перспективных задач управления космическими средствами с междисциплинарных системно-кибернетических позиций. Второй раз я активно взаимодействовал с уже генерал-майором д.т.н., профессором Юсуповым Р.М. в 80-х гг. XX века, когда мы в ВКА им. А.Ф. Можайского на межфакультетском уровне осуществляли исследования проблем создания и внедрения интегрированных распределенных экспертных систем поддержки принятия решений и технологий искусственного интеллекта (ИИ) в вооруженных силах СССР. Рафаэль Мидхатович отвечал тогда за общую координацию данных работ в масштабе всей академии. В это время мне, начальнику научно-исследовательской лаборатории, несколько раз пришлось встречаться с Рафаэлем Мидхатовичем и обсуждать наши насущные научные и организационные проблемы. Следует отметить, что несмотря на его высокое воинское и ученое звание, большой авторитет, он всегда очень просто и доступно общался с сослуживцами и стремился в доброжелательной форме ставить перед нами трудные научные задачи, одновременно подсказывая перспективные пути их решения. Самый интересный период моей совместной работы с Юсуповым Р.М. приходится на то время, когда я выполнял обязанности заместителя директора СПИИРАН по научной работе (2006-2017 г.).

Будучи его непосредственным подчиненным на протяжении более 10 лет я участвовал во многих интересных и важных мероприятиях (сессиях РАН, международных и российских научно-технических конференциях и семинарах, диссертационных советах и т.п.) и был свидетелем того, как директор СПИИРАН чл.-корр. РАН Юсупов Р.М. профессионально и талантливо организовывал их. Успех его научной и организационной деятельности базировался на том, что любая сложная проблема, стоящая перед институтом, обсуждалась коллегиально, в горячих спорах и дискуссиях командой единомышленников, целенаправленно и умело сформированной им в течение долгих лет. При этом в любых сложных конфликтных ситуациях он всегда старался найти такое компромиссное решение, при котором бы эффективно решались задачи, поставленные перед институтом, и одновременно обеспечивалась нормальная морально-нравственная атмосфера в трудовом коллективе. В целом благодаря высокому междисциплинарному научному кругозору Юсупова Р.М., его научной прозорливости и предвидению, а также выдающимся организаторским способностям, в сложных условиях социально-экономических реформ удалось не только сохранить, но и развить

и увеличить интеллектуальный потенциал СПИИРАН, а также закрепить за институтом роль одного из ведущих научных учреждений в области информатики и автоматизации не только в Российской академии наук и в Санкт-Петербурге, но и в целом в России и за ее пределами. Все эти факторы позволяют нам уже в рамках СПб ФИЦ РАН продолжать и дальше развивать научное наследие моего учителя, выдающего отечественного ученого в области кибернетики и информатики члена-корреспондента РАН Юсупова Р.М.».

Доктор технических наук, профессор Искандеров Ю.М. замечает, что каждый человек, в особенности проявляющий себя в научной сфере, определяет для себя тех людей, которых можно считать своими учителями: «Рафаэль Мидхатович для меня – Учитель с большой буквы, обладающий уникальным сочетанием профессиональных и личностных качеств. Его опыт и умение найти оптимальное решение в самых разных ситуациях, будь то организация, проведение и реализация научно-исследовательских проектов, или поиск новых форм взаимодействия с различными организациями, или преодоление естественным образом возникающих повседневных будничных проблем, всегда вызывали восхищение и служили примером. Независимо от обсуждаемой темы, причем всегда с определенной долей юмора, общение с Рафаэлем Мидхатовичем доставляло удовольствие и побуждало работать с новой энергией, с новой интенсивностью. Надеемся, что, следуя достойному примеру Рафаэля Мидхатовича, современное поколение исследователей внесет свой весомый вклад в развитие отечественной науки».

Кандидат технических наук, руководитель лаборатории технологий больших данных социкиберфизических систем Левоневский Дмитрий Константинович и доктор технических наук, профессор, главный научный сотрудник лаборатории технологий больших данных социкиберфизических систем Богомолов Алексей Валерьевич отмечают: «Авторский коллектив считает важным почтить память выдающегося учёного в области информатики, информационных технологий и теории управления, доктора технических наук, члена-корреспондента РАН, заслуженного деятеля науки и техники РФ Юсупова Рафаэля Мидхатовича. Опыт личного общения с ним, оставивший глубокий и запоминающийся след, стал для авторов важной частью профессионального становления. Научная школа Юсупова Р.М., отличавшаяся особой системностью и вниманием к строгому обоснованию решений, в значительной степени повлияла на подходы, использованные при подготовке представленного исследования. Во многом благодаря этим

методологическим ориентирам удалось обеспечить целостность и структурированность представленных результатов.

Кандидат технических наук, руководитель лаборатории автономных робототехнических средств Савельев Антон Игоревич отмечает: «Юсупов Рафаэль Мидхатович навсегда останется в памяти как выдающийся учёный, мудрый наставник и человек высокой ответственности. Его путь – от военного инженера-исследователя до директора Санкт-Петербургского института информатики и автоматизации РАН – стал образцом преданности делу и служения стране. Более четверти века, возглавляя институт, он укреплял его научный потенциал и развивал новые направления в области информатики и теории управления. Обладая редкой глубиной мышления и организаторским даром, он создал научную школу, подготовил десятки учеников и внёс весомый вклад в развитие информационного общества. Светлая память о Рафаэле Мидхатовиче – как о талантливом руководителе и наставнике – навсегда сохранится в сердцах коллег, учеников и всех, кому посчастливилось работать с ним».

Доктор технических наук, профессор РАН Ронжин А.Л. вспоминает Юсупова Рафаэля Мидхатовича, как требовательного учителя и одновременно демократичного руководителя, умеющего в сложных сплетениях исследовательских, социальных, экономических аспектов научной деятельности находить верные компромиссы для системного целостного развития большой организации и отдельных ученых. До последних дней Рафаэль Мидхатович вел активную научную деятельность, на должности главного редактора журнала Информатика и автоматизация вычитывал каждую статью, формировал политику редакции в сложных условиях наукометрических показателей и обязательных мер соблюдения научной этики.

Вся его жизнь, работа, отношение к людям, тяга к новому, интуиция и высокая работоспособность останутся примером служения науке, душевной щедрости и порядочности.

Светлая память о Рафаэле Мидхатовиче Юсупове навсегда останется в сердцах тех, кому посчастливилось его знать, работать и общаться вместе с ним.

Н.А. КУЗНЕЦОВ, К.В. СЕМЕНИХИН
**МЕТОД НЕЛИНЕЙНОЙ ФИЛЬТРАЦИИ ПРИ НЕИЗВЕСТНОЙ
ИНТЕНСИВНОСТИ ШУМА В НАБЛЮДЕНИЯХ**

Кузнецов Н.А., Семенихин К.В. Метод нелинейной фильтрации при неизвестной интенсивности шума в наблюдениях.

Аннотация. В работе изложен метод обратной статистической линеаризации – метод нелинейной фильтрации для оценивания состояний линейно-гауссовских дифференциальных систем с неизвестной интенсивностью шума в наблюдениях. Предложенный метод основан на нелинейном преобразовании разностной ошибки с сохранением коэффициента передачи, используемого в фильтре Калмана-Бьюси. В результате нелинейный фильтр описывается системой дифференциальных уравнений того же порядка, что и вектор состояния без использования уравнений на ковариационную матрицу ошибки. Уравнения нелинейного фильтра найдены в аналитическом виде для модели одномерного движения, в которой только на старшую производную действует возмущение в виде белого шума, а наблюдению доступно лишь положение с белым шумом аддитивной помехой неизвестной интенсивности. Проведен анализ соответствующей разностной схемы нелинейной фильтрации: обоснована несмещенность оценок и получено уравнение на ковариационную матрицу ее ошибок в стационарном режиме. Теоретические результаты подтверждены численным экспериментом, в котором сравнивалась точность оценок оптимального и нелинейного фильтров.

Ключевые слова: нелинейная фильтрация, неизвестная интенсивность шума, линейная стохастическая дифференциальная система, фильтр Калмана-Бьюси.

1. Введение. Методы нелинейной фильтрации востребованы в современных исследованиях, нацеленных на решение разнообразных задач оценивания параметров и процессов частично наблюдаемых стохастических систем: оптимальная фильтрация в скрытой марковской модели при наличии шума, зависящего от оцениваемого состояния [1]; расширение метода нормальной аппроксимации на неявные стохастические дифференциальные системы [2]; адаптация калмановской фильтрации для идентификации отказов на основе метода наименьших модулей [3]; разработка устойчивых алгоритмов расширенной калмановской фильтрации без использования производных [4]; модификация расширенного фильтра Калмана для обработки навигационных наблюдений [5]; развитие метода информационных множеств для гарантированного оценивания фазовых координат многошаговых систем [6]; обоснование робастности нелинейных методов фильтрации с привлечением вероятностных и информационных метрик [7]; методика сравнительного анализа алгоритмов нелинейной фильтрации [8].

За последние два десятилетия наметился тренд на разработку методов обработки измерительной информации в сетевых или распределенных системах, в которых объект представляет собой группу взаимодействующих агентов, а датчики образуют сеть с определенным протоколом обмена информации [9–12]. Практическая реализация таких алгоритмов наталкивается на теоретически обоснованный барьер скорости передачи информации между различными узлами распределенной измерительной системы [13], что приводит к необходимости разработки алгоритмов оценивания и фильтрации с учетом ограниченности или случайности коммуникационного взаимодействия [14–16].

Последнее делает актуальной задачу разработки упрощенных алгоритмов фильтрации, рассчитанных на передачу ограниченного объема информации за один такт обработки. Для многомерных моделей серьезной вычислительной проблемой является численное обновление ковариационной матрицы ошибки оценивания в рекуррентных алгоритмах фильтрации. Эта проблема еще более усугубляется в ситуации, когда у шумов в наблюдениях неизвестна ковариационная матрица (в случае дискретного времени) или интенсивность (в случае непрерывного времени).

Для учета неточности задания характеристик случайных шумов используют *робастный подход*, в русле которого лежат методы минимаксной фильтрации: они ориентированные на применение оптимального линейного фильтра при наихудшем выборе неопределенной интенсивности шума [17, 18], его ковариационной матрицы [19, 20] или распределения внутри вероятностного множества неопределенности [21, 22].

При использовании *адаптивного подхода* синтезируется совместная схема фильтрации ненаблюдаемого состояния стохастической системы и оценивания ее неизвестных параметров: в [23] применены градиентный метод и приближенный метод Ньютона для адаптивной подстройки матриц ковариаций возмущений в объекте и шумов измерения; в [24, 25] реализованы схемы фильтра частиц с включением байесовских оценок неизвестных дисперсий (или ковариационных матриц) на основе априорных моделей обратного гамма-распределения (или обратного распределения Уишарта); в [26] использован ЕМ-алгоритм для синтеза адаптивного фильтра Калмана в линейно-гауссовской системе с неизвестными ковариациями шумов; в [27] разработана методика адаптивной сигма-точечной фильтрации для отслеживания навигационных параметров GPS-сигналов по измерениям с неизвестной интенсивностью шума.

Тем не менее, указанные два подхода (робастный и адаптивный) обладают своими недостатками. С одной стороны, робастные методы оценивания зачастую порождают достаточно консервативные решения. С другой стороны, адаптивные методы в случае многопараметрической неопределенности приводят к сложным и подчас неустойчивым численным схемам. Для изучаемой в настоящей статье задачи фильтрации состояния линейно-гауссовской дифференциальной системы с неизвестной интенсивностью шума в наблюдениях можно предложить альтернативный подход. Он основан на нелинейном преобразовании разностной ошибки с сохранением коэффициента передачи, используемого в фильтре Калмана-Бьюси. Тем самым оптимальный линейный фильтр с неопределенными коэффициентами заменяется нелинейным фильтром с теми же мгновенными моментными характеристиками без использования дополнительной схемы оценивания неизвестной интенсивности. Поэтому этот метод можно назвать *методом обратной статистической линеаризации*.

Его идея была впервые предложена более 50 лет тому назад [28] и развита нашим коллегой Ф.Н. Григорьевым в его кандидатской диссертации [29]. Более того, данный метод нелинейной фильтрации был использован для построения оценок параметров движения крупнотоннажных транспортных судов и разработки информационной системы предупреждения их столкновений в условиях, когда величины измерительных ошибок неизвестны [30]. Соответствующие алгоритмы были внедрены в автоматизированную систему управления судовождением на морском флоте (Государственная премия СССР в области науки и техники за 1985 год). К сожалению, в научной литературе данный метод нелинейной стохастической фильтрации до сих пор не получил должного отражения. Поэтому целью настоящей статьи является его аккуратное изложение и анализ свойств соответствующих оценок.

Статья имеет следующую структуру. В разделе 2 приведены описание рассматриваемого класса систем и постановка задачи. В разделе 3 предьявлено аналитическое решение для модели частного вида, а также приведены соображения, позволяющие распространить данный метод на более общий случай. Раздел 4 содержит синтез разностной схемы реализации искомого нелинейного фильтра, обоснование несмещенности оценок и вывод рекуррентного уравнения для ковариационной матрицы ошибки. В разделе 5 представлены результаты численного эксперимента для сравнения точности оценок построенного нелинейного фильтра с эталонным качеством оценивания

оптимального фильтра при разном уровне помех. Раздел 6 представляет собой заключение статьи с кратким описанием полученных результатов.

2. Постановка задачи. Предположим, что состояние объекта $X(t) \in \mathbb{R}^m$ и вектор наблюдений $Y(t) \in \mathbb{R}^n$ описываются линейной стохастической дифференциальной системой

$$\dot{X} = AX + u(t) + Bv(t), \quad \dot{Y} = CX + \sqrt{\Sigma}\varepsilon(t), \quad (1)$$

с заданными постоянными матрицами $A \in \mathbb{R}^{m \times m}$, $B \in \mathbb{R}^{m \times q}$, $C \in \mathbb{R}^{n \times m}$, входом $u(t) = U(t, \{Y(s): s < t\})$, определяемым известным законом управления $U(\cdot)$, и неизвестной матрицей интенсивности $\Sigma \in \mathbb{R}_+^{n \times n}$ ($\mathbb{R}_+^{n \times n}$ обозначает конус неотрицательно определенных матриц размера $n \times n$), где $v(t) \in \mathbb{R}^q$, $\varepsilon(t) \in \mathbb{R}^n$ – независимые стандартные гауссовские белые шумы.

Допустим, что матрица интенсивности шума в наблюдениях положительно определена: $\Sigma \succ O$, пара (A, C) удовлетворяет условию наблюдаемости

$$\ker \begin{bmatrix} C \\ CA \\ \dots \\ CA^{m-1} \end{bmatrix} = \{0\}, \quad (2)$$

а пара (A, B) – условию управляемости

$$\text{im} [B \quad AB \quad \dots \quad A^{m-1}B] = \mathbb{R}^m, \quad (3)$$

($\ker$ и im обозначают ядро и образ соответствующего линейного преобразования).

При этих условиях определен стационарный фильтр Калмана-Бьюси [31], т.е. оценка $\hat{X}(t) = M\{X(t) | Y(s): s < t\}$ описывается линейным стохастическим дифференциальным уравнением

$$\dot{\hat{X}} = A\hat{X} + u(t) + K[\dot{Y} - C\hat{X}], \quad (4)$$

с постоянным коэффициентом усиления

$$K = QC^T \Sigma^{-1} \in \mathbb{R}^{m \times n} \quad (5)$$

и постоянной ковариационной матрицей ошибки фильтрации

$$Q = \text{cov}\{\hat{X}(t) - X(t)\} \in \mathbb{R}_+^{m \times m}, \quad (6)$$

удовлетворяющей алгебраическому уравнению Риккати

$$AQ + QA^T + BB^T - QC^T \Sigma^{-1} CQ = O, \quad (7)$$

решение которого в классе положительно определенных матриц существует и единственно [32].

Кроме того, разностная ошибка $\delta(t) = \dot{Y}(t) - C\hat{X}(t)$ представляет собой гауссовский белый шум с матрицей интенсивности Σ . Тогда преобразованная разностная ошибка $K\delta(t)$ тоже будет гауссовским белым шумом с матрицей интенсивности

$$S = K\Sigma K^T = QC^T \Sigma^{-1} CQ. \quad (8)$$

Ясно, что каждая из трех матриц (5), (6) и (8) является функцией неизвестной матрицы Σ , поэтому далее (если это потребуется) будем использовать запись: $K = K(\Sigma)$, $Q = Q(\Sigma)$, $S = S(\Sigma)$ и т.д.

Предлагается заменить линейный фильтр (4), содержащий неопределенный коэффициент K , нелинейным фильтром

$$\dot{\check{X}} = A\check{X} + u(t) + \varphi(\dot{Y} - C\check{X}), \quad (9)$$

в котором нелинейное преобразование $\varphi: \mathbb{R}^n \rightarrow \mathbb{R}^m$ выбирается из условия соответствия моментных характеристик преобразованной разностной ошибки в фильтре Калмана-Бьюси. Последнее означает, что при любом выборе матрицы $\Sigma \succ O$ будет верно

$$M\{\varphi(\xi)\} = 0, \quad \text{cov}\{\varphi(\xi)\} = \text{cov}\{K\xi\}, \quad \xi \sim \mathcal{N}(0, \Sigma), \quad (10)$$

где $K = K(\Sigma)$ определяется из (5), (6). Тем самым искомое нелинейное преобразование $\varphi(\xi)$ центрированного гауссовского вектора ξ с положительно определенной ковариационной матрицей Σ должно иметь те же математическое ожидание и ковариацию, что и линейное преобразование $K(\Sigma)\xi$ с коэффициентом из стационарного фильтра Калмана-Бьюси.

Если совпадение ковариационных матриц в (10) является невозможным, то будем накладывать менее строгое ограничение – совпадение дисперсий для каждой компоненты $\varphi_i(\xi)$ искомого преобразования

$$D\{\varphi_i(\xi)\} = (K\Sigma K^T)_{i,i} \quad \text{при} \quad i = 1, \dots, m, \quad \xi \sim \mathcal{N}(0, \Sigma). \quad (11)$$

Требуется: предъявить класс систем вида (1), для которых можно реализовать предложенный метод обратной статистической линеаризации, дать корректное определение нелинейной стохастической системе (9) и изучить качество соответствующих оценок $\hat{X}(t)$ в сравнении с тем, что дает оптимальный фильтр $\hat{X}(t)$, рассчитанный при истинной матрице интенсивности шума Σ .

3. Описание метода фильтрации. Начнем с важного частного случая, в котором искомый нелинейным фильтр допускает аналитическое описание [28, 29].

Рассмотрим модель, в которой вектор состояния $X(t) \in \mathbb{R}^m$ составлен из значений скалярного процесса $x(t)$ и его производных $x^{(i)}(t)$, $i = 1, \dots, m - 1$, т.е.

$$X = \text{col}[X_1, \dots, X_m] = \text{col}[x, \dot{x}, \ddot{x}, \dots, x^{(m-1)}], \quad (12)$$

($\text{col}[\dots]$ обозначает столбец), возмущение присутствует явно только в уравнении старшей производной $x^{(m-1)}(t)$, а измерению доступен процесс $x(t)$ с аддитивной помехой:

$$\dot{X}_1 = X_2, \dots, \dot{X}_{m-1} = X_m, \dot{X}_m = \beta v(t), \quad \dot{Y} = X_1 + \sigma \varepsilon(t), \quad (13)$$

где β, σ – положительные коэффициенты, из которых β задано, а интенсивность шума σ неизвестна. Возмущение $v(t)$ и помеха $\varepsilon(t)$ предполагаются независимыми стандартными гауссовскими белыми шумами.

Ясно, что (13) – частный случай модели (1) с матрицами

$$A = \begin{bmatrix} 0 & 1 & & \\ \vdots & & \ddots & \\ 0 & & & 1 \\ 0 & 0 & \dots & 0 \end{bmatrix}, \quad B = \beta E, \quad E = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix}, \quad (14)$$

$$C = \begin{bmatrix} 1 & 0 & \dots & 0 \end{bmatrix}, \quad \Sigma = \sigma^2. \quad (15)$$

В этом случае известно [33], что уравнение относительно ковариационной матрицы ошибки стационарного фильтра Калмана (7) имеет аналитическое решение:

$$Q = \{Q_{i,j}\}_{i,j=1,\dots,m}: \quad Q_{i,j} = G_{i,j} \sigma^2 (\beta/\sigma)^{(i+j-1)/m}, \quad (16)$$

где коэффициенты $\{G_{i,j}\}$ не зависят от параметров системы (13) и образуют симметричную матрицу $G \in \mathbb{R}^{m \times m}$, которая представляет собой единственное положительно определенное решение уравнения Риккати

$$AG + GA^T + EE^T - GC^T CG = O. \quad (17)$$

Данное уравнение полностью аналогично (7) применительно к системе (13) с параметрами $\beta = \sigma = 1$. Существование и единственность решения $G \succ O$ вытекает из условий наблюдаемости и управляемости, выполненных для пар (A, C) и (A, E) соответственно.

Таким образом, коэффициент усиления (5) принимает вид

$$K = QC^T(\sigma^2)^{-1} = \text{col}[K_1, \dots, K_m], \quad K_i = G_{i,1}(\beta/\sigma)^{i/m}. \quad (18)$$

Тогда соответствующий оптимальный фильтр (4) допускает запись

$$\dot{\hat{X}}_i = \hat{X}_{i+1} + G_{i,1}(\beta/\sigma)^{i/m}\delta(t), \quad i = 1, \dots, m-1, \quad (19)$$

$$\dot{\hat{X}}_m = G_{m,1}(\beta/\sigma)\delta(t), \quad (20)$$

где $\delta(t) = \dot{Y}(t) - \dot{\hat{X}}_1(t)$ – разностная ошибка, образующая гауссовский белый шум интенсивности σ^2 , а коэффициент усиления при нем является (с точностью до множителя) степенной функцией

$$k(\sigma) = \sigma^{-\alpha}, \quad \text{где } 0 < \alpha \leq 1. \quad (21)$$

Тогда согласно методу обратной статистической линейаризации необходимо заменить линейное преобразование $\eta = k(\sigma)\xi$ гауссовской случайной величины $\xi \sim \mathcal{N}(0, \sigma^2)$ с неизвестной дисперсией σ^2 ее нелинейным преобразованием $\phi(x)$ с сохранением первых двух моментов, т.е.:

$$M_\sigma \phi(\xi) = 0, \quad D_\sigma \phi(\xi) = (k(\sigma)\sigma)^2 \quad \forall \sigma > 0. \quad (22)$$

Для устойчивости замкнутой системы и сохранения симметрии искомую функцию $\phi(x)$ следует выбрать монотонной, нечетной и не более чем линейного роста. Если взять степенную функцию, то в случае (21) получаем

$$\phi(x) = \gamma_p^{-1/2} \text{sign}(x)|x|^{1-\alpha}, \quad (23)$$

где γ_p обозначает абсолютный момент порядка $p = 2(1 - \alpha)$ стандартной нормальной величины $\nu \sim \mathcal{N}(0, 1)$, т.е.:

$$\gamma_p = M|\nu|^p = \frac{2^{p/2}}{\sqrt{\pi}} \Gamma\left(\frac{p+1}{2}\right).$$

Действительно, первое из условий (22) имеет место в силу нечетности функции $\phi(x)$, а второе следует из цепочки равенств

$$M_\sigma(\phi(\xi))^2 = \gamma_p^{-1} M_\sigma |\xi|^p = \gamma_p^{-1} M(\sigma|\nu|)^p = \sigma^p = (k(\sigma)\sigma)^2. \quad (24)$$

Теперь в соответствии с (9) получаем уравнения нелинейного фильтра

$$\dot{\check{X}}_i = \check{X}_{i+1} + G_{i,1} \gamma_{2(1-i/m)}^{-1/2} \beta^{i/m} \text{sign}(\delta(t)) |\delta(t)|^{1-i/m} \quad (25)$$

при $i = 1, \dots, m-1$,

$$\dot{\check{X}}_m = G_{m,1} \beta \text{sign}(\delta(t)), \quad (26)$$

с учетом разностной ошибки $\delta(t) = \dot{Y}(t) - \check{X}_1(t)$.

Структурная схема построенного нелинейного фильтра приведена на рисунке 1.

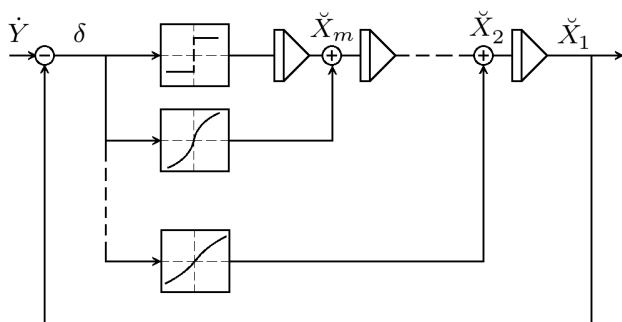


Рис. 1. Структурная схема нелинейного фильтра (25), (26)

Отметим, что построенное нелинейное преобразование $\varphi: \mathbb{R} \rightarrow \mathbb{R}^m$ подчиняется более слабому требованию (11). Нетрудно проверить, что отличие между двумя ковариационными матрицами в исходном условии (10) определяется коэффициентами

$$\frac{\gamma(p_i+p_j)/2}{\sqrt{\gamma p_i \gamma p_j}} \quad \text{для } i, j = 1, \dots, m \quad \text{и} \quad p_i = 2(1 - i/m). \quad (27)$$

Теперь рассмотрим более общую ситуацию, в которой по-прежнему наблюдения скалярны (т.е. $n = 1$) и имеют неизвестную интенсивность шума σ^2 , но матрицы A, B, C теперь будут произвольными – лишь бы были выполнены условия наблюдаемости (2) и управляемости (3).

Коэффициент усиления (5) представляет собой вектор-столбец с компонентами $K_i(\sigma^2)$, $i = 1, \dots, m$, зависящими от неизвестной интенсивности σ^2 . Предположим, что эта зависимость определена численно или аналитически. Тогда для назначения функции $\varphi: \mathbb{R} \rightarrow \mathbb{R}^m$, соответствующей условию (11), предположим, что ее компоненты $\varphi_i(x)$ выбраны нечетными функциями, такими что

$$D_\sigma \varphi_i(\xi) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} \varphi_i^2(x) e^{-x^2/(2\sigma^2)} dx = \sigma^2 K_i^2(\sigma^2). \quad (28)$$

После замены переменной $y = x^2/2$ и $\theta = 1/\sigma^2$ получаем, что искомая функция определяется в результате обратного преобразования Лапласа

$$\int_0^\infty \psi_i(y) e^{-\theta y} dy = \frac{K_i^2(1/\theta)}{\theta^{3/2}} \sqrt{\pi/2}, \quad \varphi_i(x) = \sqrt{x\psi_i(x^2/2)}, \quad x \geq 0. \quad (29)$$

В случае векторных наблюдений предлагается привести их к одному масштабу, воспользоваться оценкой сверху на неизвестную матрицу интенсивности шума в виде $\Sigma \preceq \sigma^2 I$, задать нелинейное преобразование $\varphi(x)$ через функции одной переменной $\varphi_i(x_i)$ и искать последние из того же условия (28), но с тем уточнением, что K_i^2 обозначает сумму квадратов i -й строки матричного коэффициента усиления (5), в котором надо положить $\Sigma = \sigma^2 I$.

Таким образом, описан общий метод построения нелинейного преобразования $\varphi(\cdot)$ в искомом фильтре (9). Однако остается вопрос численной реализации соответствующего фильтра вместе с проблемой

применимости нелинейного звена к случайному процессу типа белого шума.

Синтез соответствующей численной схемы и обоснование ее устойчивости проведены в следующем разделе.

4. Численная схема. Рассмотрим численную реализацию методов фильтрации на основе метода Эйлера [34]. Для линейного фильтра (4) соответствующая численная схема имеет вид

$$\begin{aligned}\hat{X}(t_{k+1}) = & \hat{X}(t_k) + h(A\hat{X}(t_k) + u(t_k)) + \\ & + K[Y(t_{k+1}) - Y(t_k) - hC\hat{X}(t_k)],\end{aligned}\quad (30)$$

где $\{t_{k+1} = t_k + h, k = 0, 1, 2, \dots\}$ – временная сетка с достаточно малым шагом $h > 0$.

Важно отметить, что выражение в квадратных скобках представляет собой аппроксимацию интеграла

$$\int_{t_k}^{t_{k+1}} \delta(t) dt = Y(t_{k+1}) - Y(t_k) - C \int_{t_k}^{t_{k+1}} \hat{X}(t) dt \quad (31)$$

от разностной ошибки $\delta(t) = \dot{Y}(t) - C\hat{X}(t)$, которая образует гауссовский белый шум интенсивности Σ . Тогда по свойству стохастического интеграла вектор (31) является гауссовским, имеет нулевое математическое ожидание, а его ковариационная матрица равна $h\Sigma$. После умножения (31) на коэффициент $1/\sqrt{h}$ получаем случайный вектор $\xi \sim \mathcal{N}(0, \Sigma)$, к которому можно применить нелинейное преобразование, сохраняющее математическое ожидание и ковариационную матрицу (10).

Поэтому искомая численная схема реализации нелинейного фильтра (9) принимает вид

$$\begin{aligned}\check{X}(t_{k+1}) = & \check{X}(t_k) + h(A\check{X}(t_k) + u(t_k)) + \\ & + \sqrt{h} \varphi\left(\frac{Y(t_{k+1}) - Y(t_k)}{\sqrt{h}} - \sqrt{h} C\check{X}(t_k)\right).\end{aligned}\quad (32)$$

Если для состояния и наблюдаемого процесса воспользоваться аппроксимацией, основанной на методе Эйлера

$$X(t_{k+1}) = X(t_k) + h(A X(t_k) + u(t_k)) + B \int_{t_k}^{t_{k+1}} v(t) dt, \quad (33)$$

$$Y(t_{k+1}) = Y(t_k) + h C X(t_k) + \sqrt{\Sigma} \int_{t_k}^{t_{k+1}} \varepsilon(t) dt, \quad (34)$$

то для ошибки оценки нелинейного фильтра $e(t_k) = \check{X}(t_k) - X(t_k)$ получается следующее рекуррентное уравнение:

$$e(t_{k+1}) = [I + hA]e(t_k) + \sqrt{h} \left\{ -Bv_{k+1} + \varphi \left(-\sqrt{h} C e(t_k) + \sqrt{\Sigma} \varepsilon_{k+1} \right) \right\}, \quad (35)$$

где случайные векторы $v_1, v_2, \dots \in \mathbb{R}^q$ и $\varepsilon_1, \varepsilon_2, \dots \in \mathbb{R}^n$, взаимно независимы и подчиняются стандартному нормальному распределению $\mathcal{N}(0, I)$.

Если функция $\varphi(x)$ является нечетной, т.е.

$$\varphi(-x) = -\varphi(x) \quad \forall x \in \mathbb{R}^n, \quad (36)$$

и начальная ошибка фильтрации $e(t_0)$ имеет симметричное относительно нуля распределение, то тем же свойством будет обладать и распределение ошибки фильтрации $e(t_k)$ на всех последующих шагах. Поэтому оценка определяемая рекуррентными уравнениями (32) в дискретизованной модели (33), (34) является несмещенной:

$$M\{\check{X}(t_k) - X(t_k)\} = 0. \quad (37)$$

Для анализа точности нелинейного фильтра рассмотрим ковариационную матрицу ошибки оценивания $\Delta(t_k) = \text{cov}\{e(t_k)\}$, которая в силу (37) может быть определена как $\Delta(t_k) = M\{e(t_k)e^T(t_k)\}$, а также условную матрицу вторых моментов ошибки $\Delta(t_k|x) = M\{e(t_k)e^T(t_k) | e(t_{k-1}) = x\}$. Указанные матрицы связаны равенством

$$\Delta(t_k) = M\{\Delta(t_k|e(t_{k-1}))\}. \quad (38)$$

Тогда из (35) следует

$$\Delta(t_{k+1}|x) = hBB^T + M\{([I + hA]x + \sqrt{h}\varphi(-\sqrt{h}Cx + \xi))(\star)^T\}, \quad (39)$$

где $\xi \sim \mathcal{N}(0, \Sigma)$, а символ $\star$ означает дублирование выражения, стоящего слева в круглых скобках.

Дальнейшие вычисления проведем в предположении $h \downarrow 0$.

Заметим, что при $y \rightarrow 0$

$$M\varphi(y + \xi) = L\Sigma^{-1}y + o(y), \quad L = L(\Sigma) = M\{\varphi(\xi)\xi^T\} \in \mathbb{R}^{m \times n}. \quad (40)$$

Это следует из равенства $M\varphi(\xi) = 0$, представления

$$M\varphi(y + \xi) = \int_{\mathbb{R}^n} \varphi(z)g(z - y)dz, \quad g(v) = \frac{\exp\{-v^T\Sigma^{-1}v/2\}}{(2\pi)^{n/2}\sqrt{\det\Sigma}}$$

и вида градиента $\nabla g(v) = -g(v)\Sigma^{-1}v$. Тем самым (40) влечет

$$\sqrt{h}M\{\varphi(-\sqrt{h}Cx + \xi)\} = -hL\Sigma^{-1}Cx + o(h). \quad (41)$$

Аналогично получаем:

$$hM\{\varphi(-\sqrt{h}Cx + \xi)\varphi^T(-\sqrt{h}Cx + \xi)\} = h\Phi + o(h), \quad (42)$$

где ковариационная матрица

$$\Phi = \Phi(\Sigma) = M\{\varphi(\xi)\varphi^T(\xi)\} \quad (43)$$

совпадает с матрицей $K\Sigma K^T$ в случае (10) или имеет с последней одинаковые диагональные элементы в случае (11).

Теперь из (39) и полученных соотношений следует

$$\begin{aligned} \Delta(t_{k+1}|x) = & hBB^T + xx^T + h\{Axx^T + x(Ax)^T\} - \\ & - h\{L\Sigma^{-1}Cxx^T + x(L\Sigma^{-1}Cx)^T\} + h\Phi + o(h). \end{aligned} \quad (44)$$

Применяя (38) и исключая бесконечно малые порядка $o(h)$, получаем приближенную рекуррентную схему для определения ковариационной матрицы ошибки нелинейного фильтра:

$$\begin{aligned} \Delta(t_{k+1}) = \Delta(t_k) + h \{ & (A - L\Sigma^{-1}C)\Delta(t_k) + \Delta(t_k)(A - L\Sigma^{-1}C)^T + \\ & + BB^T + \Phi \}. \end{aligned} \quad (45)$$

Данная схема будет устойчивой, если все собственные значения λ_i матрицы

$$\Lambda = \Lambda(\Sigma) = A - L\Sigma^{-1}C \quad (46)$$

имеют отрицательные вещественные части $\operatorname{Re} \lambda_i$, а шаг дискретизации h настолько мал, что числа $1 + h \operatorname{Re}[\lambda_i + \lambda_j]$ при любых $i, j = 1, \dots, m$ лежат внутри единичного интервала $(-1, 1)$ (лемма И.1 из [32]). Последнее имеет место при

$$h < -1/\operatorname{Re} \lambda_i \quad \forall \lambda_i \in \sigma[\Lambda]. \quad (47)$$

Если это выполнено, то при любом начальном условии решение уравнения (45), которое представляет собой рекуррентный вариант уравнения Ляпунова, сходится при $k \rightarrow \infty$ к решению алгебраического уравнения Ляпунова

$$\Lambda\Delta + \Delta\Lambda^T + BB^T + \Phi = O, \quad (48)$$

причем его решение Δ существует и единственно в классе $\mathbb{R}_+^{m \times m}$ (а также будет положительно определенной матрицей, если $BB^T + \Phi \succ O$).

В ситуации неизвестной матрицы интенсивности шума Σ уравнение (48) дает лишь качественную оценку точности нелинейного фильтра. Однако условие (47) вводит количественные ограничения на параметры модели и численной схемы, т.е. на интенсивность шума и размер шага, при которых разработанный алгоритм нелинейной фильтрации является работоспособным.

Приведем пример, в котором можно дать явные оценки качества предложенной нелинейной схемы фильтрации.

Пример 1. Рассмотрим одномерный вариант модели (13), т.е. случай $m = 1$. Тогда для скалярного состояния $X(t)$ имеем

$$\dot{X}(t) = \beta v(t), \quad \dot{Y}(t) = X(t) + \sigma \varepsilon(t). \quad (49)$$

Согласно (16) дисперсия стационарного фильтра Калмана-Бьюси равна $Q = \sigma^2(\beta/\sigma) = \beta\sigma$, так как $G_{1,1} = 1$.

В силу (26) нелинейный фильтр определяется уравнениями

$$\dot{\check{X}}(t) = \varphi(\dot{Y}(t) - \check{X}(t)), \quad \varphi(x) = \beta \operatorname{sign} x. \quad (50)$$

С учетом $\xi \sim \mathcal{N}(0, \sigma^2)$ коэффициент (40) принимает вид

$$L = M\{\varphi(\xi)\xi\} = \beta M|\xi| = \beta\sigma\gamma_1, \quad \text{где } \gamma_1 = \sqrt{2/\pi}. \quad (51)$$

Тогда $\Lambda = -L(\sigma^2)^{-1} = -\gamma_1\beta/\sigma$ в силу (46).

Теперь условие (47) означает, что размер шага h надо выбирать так чтобы $h < \gamma_1^{-1}\sigma/\beta$ (иначе говоря h должен быть, как максимум, того же порядка что интенсивность помех).

Если учесть (8), то $K^2\sigma^2 = Q^2(\sigma^2)^{-1} = \beta^2$, откуда в силу (48) получаем

$$\Delta = \frac{\beta^2 + K^2\sigma^2}{-2\Lambda} = \gamma_1^{-1}\beta\sigma = \gamma_1^{-1}Q, \quad (52)$$

где $\gamma_1^{-1} = \sqrt{\pi/2} \approx 1,25$. Тем самым дисперсия ошибки нелинейного фильтра Δ лишь на четверть больше дисперсии ошибки оптимального фильтра Q , что в исходной шкале составляет около 11%.

Таким образом, синтезирована разностная схема, реализующая искомый нелинейный фильтр; доказана несмещенность его оценок; выведено рекуррентное уравнение для ковариационной матрицы ошибки оценивания и получено условие устойчивости численной реализации нелинейного фильтра.

5. Численный пример. Для иллюстрации предложенного подхода рассмотрим одномерную модель движения маневрирующей цели, положение которой описывается процессом $x(t)$, а ускорение $\ddot{x}(t)$ представляет собой процесс броуновского движения, т.е. третья производная $\ddot{\ddot{x}}(t)$ образует гауссовский белый шум с известной интенсивностью β^2 . Допустим, что наблюдаемый процесс $Y(t)$ представляет собой результат измерения положения $x(t)$ с гауссовской белозумной помехой неизвестной интенсивности σ^2 . Указанные предположения записываются аналогично (13) при $m = 3$:

$$\dot{X}_1 = X_2, \quad \dot{X}_2 = X_3, \quad \dot{X}_3 = \beta v(t), \quad \dot{Y} = X_1 + \sigma \varepsilon(t),$$

где $X_1(t) = x(t)$, $X_2(t) = \dot{x}(t)$, $X_3(t) = \ddot{x}(t)$, а $v(t)$ и $\varepsilon(t)$ – независимые стандартные гауссовские белые шумы.

Тогда уравнения оптимального фильтра (19), (20) принимают вид

$$\begin{cases} \dot{\hat{X}}_1 = \hat{X}_2 + 2(\beta/\sigma)^{1/3} [\dot{Y}(t) - \hat{X}_1], \\ \dot{\hat{X}}_2 = \hat{X}_3 + 2(\beta/\sigma)^{2/3} [\dot{Y}(t) - \hat{X}_1], \\ \dot{\hat{X}}_3 = (\beta/\sigma) [\dot{Y}(t) - \hat{X}_1], \end{cases} \quad (53)$$

с учетом того, что уравнение Риккати (17) имеет следующее решение:

$$G = \begin{bmatrix} 2 & \star & \star \\ 2 & 3 & \star \\ 1 & 2 & 2 \end{bmatrix},$$

(здесь и далее символ $\star$ используется для дублирования элементов симметричной матрицы).

Отсюда ковариационная матрица ошибки оптимальной оценки равна

$$Q = \sigma^2 \begin{bmatrix} 2\rho^{1/3} & \star & \star \\ 2\rho^{2/3} & 3\rho & \star \\ \rho & 2\rho^{4/3} & 2\rho^{5/3} \end{bmatrix}, \quad \rho = \beta/\sigma.$$

В силу (25), (26) получаем уравнения нелинейного фильтра, которые не зависят от неизвестной интенсивности σ^2 :

$$\begin{cases} \dot{\check{X}}_1 = \check{X}_2 + \varphi_1(\delta(t)), & \varphi_1(x) = 2\beta^{1/3}\gamma_{4/3}^{-1/2}\text{sign}(x)|x|^{2/3}, \\ \dot{\check{X}}_2 = \check{X}_3 + \varphi_2(\delta(t)), & \varphi_2(x) = 2\beta^{2/3}\gamma_{2/3}^{-1/2}\text{sign}(x)|x|^{1/3}, \\ \dot{\check{X}}_3 = \varphi_3(\delta(t)), & \varphi_3(x) = \beta\text{sign}(x), \end{cases} \quad (54)$$

где $\delta(t) = \dot{Y}(t) - \dot{\check{X}}_1(t)$, причем $\gamma_{4/3} \approx 0,831$ и $\gamma_{2/3} \approx 0,802$.

Чтобы убедиться, что расхождение между ковариационной матрицей используемого нелинейного преобразования и требуемой ковариацией в (10) незначительно, приведем значения коэффициентов (27):

$$\begin{bmatrix} 1 & \star & \star \\ \gamma_1(\gamma_{2/3}\gamma_{4/3})^{-1/2} & 1 & \star \\ \gamma_{2/3}(\gamma_{4/3})^{-1/2} & \gamma_{1/3}(\gamma_{2/3})^{-1/2} & 1 \end{bmatrix} = \begin{bmatrix} 1 & \star & \star \\ 0,977\dots & 1 & \star \\ 0,880\dots & 0,957\dots & 1 \end{bmatrix}.$$

На рисунке 2 приведены траектории исходного сигнала и его оценок, полученных с помощью уравнений оптимального (53) и нелинейного (54) фильтров. При этом нижний график в сравнении с верхним описывает ситуацию, когда σ^2 , т.е. интенсивность шума в наблюдениях, увеличена в 100 раз, а именно, σ было взято равным 5 и 50, соответственно. Параметр β , определяющий интенсивность возмущений β^2 , был равен 3.

На рисунке приведены также границы, построенные по правилу трех «сигм» на основе среднеквадратичной (с.к.) ошибки оптимальной фильтрации (здесь одна «сигма» равна $\sqrt{Q_{1,1}} = \sigma\sqrt{2\rho^{1/3}} = \sqrt{2\beta^{1/6}\sigma^{5/6}}$). При вариации параметра σ в 10 раз соответствующая с.к. ошибка $\sqrt{Q_{1,1}}$ меняется примерно в 7 раз: от 6,5 до 44,2 (ср. ширину коридора на верхнем и нижнем графиках).

Таким образом, нелинейный фильтр, не использующий информацию об интенсивности помех, демонстрирует поведение, почти идентичное оптимальному фильтру, который основан на точном знании характеристик измерительных ошибок.

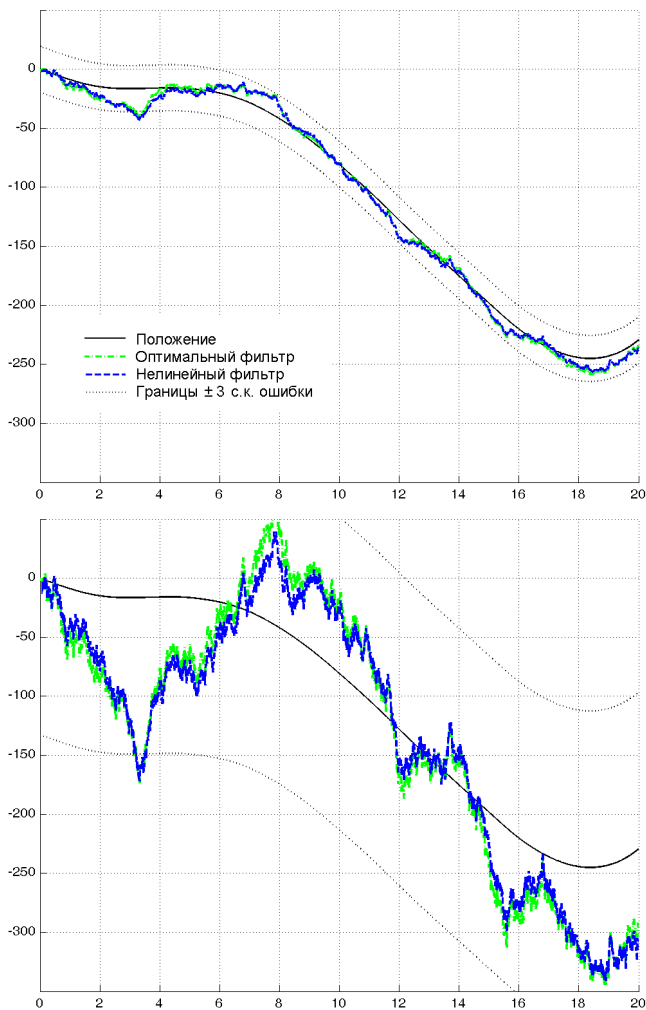


Рис. 2. Траектория положения цели $X_1(t)$, оптимальная оценка $\hat{X}_1(t)$, нелинейная оценка $\check{X}_1(t)$ и границы радиуса три с.к. ошибки оптимального фильтра (в двух экспериментах: на нижнем графике интенсивность шума σ^2 в 100 раз больше, чем на верхнем)

6. Заключение. В работе описан метод нелинейной фильтрации (названный методом обратной статистической линеаризации), предназначенный для оценивания состояния линейной стохастической дифференциальной системы при наличии неизвестной интенсивности гауссовских белых шумов. Предложенный метод представляет собой альтернативу двум подходам – робастному и адаптивному, традиционно используемым в задачах управления и наблюдения применительно к стохастическим системам с неопределенными параметрами.

В статье описана многомерная система со скалярными возмущениями и шумами, в которой предложенный метод приводит к явному описанию нелинейного фильтра системой дифференциальных уравнений (того же порядка, что и размерность вектора состояния). Представлено обоснование разработанного метода нелинейной фильтрации на основе асимптотического анализа соответствующей разностной схемы для ковариационной матрицы ошибки. Результаты численного эксперимента показывают, что погрешность оценивания нелинейного фильтра остается на том же уровне, что и для оптимального фильтра при вариации интенсивности шумов на два порядка.

Идея изложенного выше метода фильтрации возникла у одного из авторов в те годы, когда ему посчастливилось сотрудничать с Юсуповым Рафаэлем Мидхатовичем в области разработки алгоритмов автоматического управления для технических систем общего и специального назначения. Памяти этого выдающегося ученого и неординарного человека посвящается данная статья.

Литература

1. Borisov A., Sokolov I. Optimal filtering of Markov jump processes given observations with state-dependent noises: Exact solution and stable numerical schemes // *Mathematics*. 2020. vol. 8. no. 4.
2. Синицын И.Н. Методы нормальной субоптимальной фильтрации в наблюдаемых неявных гауссовских стохастических системах // *Системы и средства информатики*. 2025. Т. 35. № 1. С. 41–58.
3. Миллер Б.М., Колосов К.С. Робастное оценивание на основе метода наименьших модулей и фильтра Калмана // *Автоматика и телемеханика*. 2020. № 11. С. 72–92.
4. Kulikova M.V., Kulikov G.Y. On derivative-free extended Kalman filtering and its Matlab-oriented square-root implementations for state estimation in continuous-discrete nonlinear stochastic systems // *European J. Control*. 2023. vol. 73.
5. Босов А.В., Урюпин И.В. Модифицированный вариант расширенного фильтра Калмана по методу линейных псевдонаблюдений // *Информ. и ее примен.* 2025. Т. 19. № 2. С. 17–26.
6. Ананьев Б.И. Об оценивании динамических систем при неточных ограничениях на помехи // *Тр. Ин-та математики и механики УрО РАН*. 2025. Т. 31. № 2. С. 15–29.

7. Wang S. Distributionally robust state estimation for nonlinear systems // *IEEE Trans. Signal Process.* 2022. vol. 70. pp. 4408–4423.
8. Степанов О.А., Исаев А.М. Методика сравнительного анализа рекуррентных алгоритмов нелинейной фильтрации в задачах обработки навигационной информации на основе предсказательного моделирования // *Гирроскопия и навигация.* 2023. Т. 31. № 3(122). С. 48–65.
9. Olfati-Saber R. Kalman-consensus filter: Optimality, stability, and performance // *Proc. 48th IEEE Conf. Decision and Control (CDC'2009) held jointly with 28th Chinese Control Conf.* 2009. pp. 7036–7042. DOI: 10.1109/CDC.2009.5399678.
10. Sharma R., Beard R.W., Taylor C.N., Quebe S. Graph-based observability analysis of bearing-only cooperative localization // *IEEE Trans. Robotics.* 2011. vol. 28. no. 2. pp. 522–529.
11. Wang L., Liu J., Morse A.S. A distributed observer for a continuous-time linear system with time-varying network // *arXiv:2003.02134.* 2020.
12. Talebi P.S., Mandic D. On the dynamics of multi agent nonlinear filtering and learning // *arXiv:2309.03557v2.* 2023.
13. Андриевский Б.Р., Матвеев А.С., Фрадков А.Л. Управление и оценивание при информационных ограничениях: к единой теории управления, вычислений и связи // *Автоматика и телемеханика.* 2010. № 4. С. 34–99.
14. Song W., Wang Z., Li Z., Wang J., Han Q.-L. Nonlinear filtering with sample-based approximation under constrained communication: Progress, insights and trends // *IEEE/CAA J. Autom. Sinica.* 2024. vol. 11. no. 7. pp. 1539–1556.
15. Voortman Q., Efimov D., Pogromsky A., Richard J.-P., Nijmeijer H. Remote state estimation of steered systems with limited communications: An event-triggered approach // *IEEE Trans. Automat. Control.* 2024. vol. 69. no. 7. pp. 4199–4214.
16. Tanwani A. Suboptimal filtering over sensor networks with random communication // *IEEE Trans. Automat. Control.* 2022. vol. 67. no. 10. pp. 5456–5463.
17. Ананьев Б.И. Минимаксные среднеквадратические оценки в статистически неопределенных системах // *Дифференц. уравнения.* 1984. Т. 20. № 8. С. 1291–1297.
18. Лебедев М.В., Семенихин К.В. Минимаксная фильтрация в стохастической дифференциальной системе с нестационарными возмущениями неизвестной интенсивности // *Изв. РАН. Теория и системы управления.* 2007. № 2. С. 45–56.
19. Панков А.Р., Платонов Е.Н., Семенихин К.В. Робастная фильтрация процесса в стационарной разностной стохастической системе // *Автоматика и телемеханика.* 2011. № 2. С. 167–182.
20. Коган М.М. Робастное оценивание и фильтрация в неопределенных линейных системах при неизвестных ковариациях // *Автоматика и телемеханика.* 2015. № 10. С. 50–66.
21. Shafieezadeh Abadeh S., Nguyen V.A., Kuhn D., Mohajerin Esfahani P.M. Wasserstein distributionally robust Kalman filtering // *Advances in Neural Information Processing Systems.* 2018. vol. 31.
22. Kargin T., Hajar J., Malik V., Hassibi B. Distributionally robust Kalman filtering over finite and infinite horizon // *arXiv:2407.18837v1.* 2024.
23. Барабанов А.Е. Линейная фильтрация с адаптивной подстройкой матриц ковариаций возмущений в объекте и шумов измерения // *Автоматика и телемеханика.* 2016. № 1. С. 30–49.
24. Li K., Zhao S., Liu F. Joint state estimation for nonlinear state-space model with unknown time-variant noise statistics // *Internat. J. Adaptive Control and Signal Processing.* 2021. vol. 35. no. 4. pp. 498–512.

25. Huang Y., Zhang Y., Wu Z., Li N., Chambers J. A novel adaptive Kalman filter with inaccurate process and measurement noise covariance matrices // IEEE Trans. Automat. Control. 2018. vol. 63. no. 2. pp. 594–601.
26. Chen Y., Li W., Wang Y. Online adaptive Kalman filter for target tracking with unknown noise statistics // IEEE Sensors Letters. 2021. vol. 5. no. 3. pp. 1–4.
27. Кануж М.М., Клоков А.В. Адаптивный ансамбльный фильтр Калмана для слежения за GPS-сигналами при неизвестной и изменяющейся во времени ковариации шума // Гирскопия и навигация. 2021. Т. 29. № 3(114). С. 34–51.
28. Григорьев Ф.Н., Кузнецов Н.А. Квазиоптимальная фильтрация при неизвестной интенсивности шума в наблюдениях // Доклады II-го Всесоюзного совещания по статистическим методам теории управления. М: Наука, 1970. С. 112–118.
29. Григорьев Ф.Н. Методы управления процессом наблюдения в непрерывных системах: Дис. канд. техн. наук: 05.13.02 / Моск. физ.-техн. ин-т. Москва, 1975. 104 с.
30. Трапезников В.А. Теоретические вопросы построения АСУ крупнотоннажными транспортными судами: Сб. статей. М.: Наука, 1978. 211 с.
31. Liptser R.S., Shirayev A.N. Statistics of random processes. 3rd edition. New York: Springer, 2005.
32. Поляк Б.Т., Хлебников М.В., Рапопорт Л.Б. Математическая теория автоматического управления. М.: ЛЕНАНД, 2019. 500 с.
33. Калачев М.Г. Один метод многократного дифференцирования сигнала в системах автоматического регулирования // Автоматика и телемеханика. 1970. № 6. С. 29–36.
34. Миллер Б.М., Панков А.Р. Теория случайных процессов в примерах и задачах. М.: Физматлит, 2002. 317 с.

Кузнецов Николай Александрович — д-р техн. наук, профессор, академик РАН, советник РАН, заведующий лабораторией, лаборатория № 351, Институт радиотехники и электроники им. В.А. Котельникова Российской академии наук (ИРЭ им. В.А. Котельникова РАН); заведующий кафедрой, кафедра «инфокоммуникационные системы и сети», Московский физико-технический институт. Область научных интересов: управление инфокоммуникационными системами и сетями, теория рассинхронизованных систем, методы управления наблюдениями в стохастических системах, разработка алгоритмов идентификации для систем управления движущимися объектами, оптимизация транспортных потоков, обработка биомедицинской информации. Число научных публикаций — 310. kuznetsov@cplire.ru; улица Моховая, 11-7, 125009, Москва, Россия; р.т.: +7(495)629-3005; факс: +7(495)629-3005.

Семенхин Константин Владимирович — д-р физ.-мат. наук, доцент, профессор, образовательный центр института № 8 «компьютерные науки и прикладная математика», Национальный исследовательский университет «Московский авиационный институт» (МАИ). Область научных интересов: -методы робастной стохастической оптимизации, управление и оценивание в стохастических системах, оптимизация сетей массового обслуживания. Число научных публикаций — 125. siemenkv@mail.ru; Волоколамское шоссе, 4, 125993, Москва, Россия; р.т.: +7(499)158-4560; факс: +7(499)158-4560.

Поддержка исследований. Работа выполнена по государственному заданию: ИРЭ им. В.А. Котельникова РАН (FFWZ-2025-0015).

N. KUZNETSOV, K. SEMENIKHIN
**NONLINEAR FILTERING METHOD IN THE PRESENCE
OF UNKNOWN NOISE INTENSITY**

Kuznetsov N., Semenikhin K. Nonlinear Filtering Method in the Presence of Unknown Noise Intensity.

Abstract. The paper describes the method of inverse statistical linearization – the method of nonlinear filtering for state estimation in linear Gaussian differential systems with observation noise of unknown intensity. The proposed technique is based on finding a nonlinear perturbation of the innovation process while keeping the gain coefficient in the Kalman-Bucy filter. As a result, the nonlinear filter is defined by a system of differential equations of the same order as the state vector without using any additional equations for the error covariance matrix. The nonlinear filter is found in an analytical form for a one-dimensional motion model in which only the highest order derivative is affected by a white-noise stochastic disturbance, and only one output is available for observing the position with additive noise of unknown intensity. The recurrent nonlinear filtering scheme is examined to establish the unbiasedness of the estimates and to obtain the steady-state equation for their error variances and covariances. The theoretical results are confirmed on the basis of computer simulations carried out to compare the estimation accuracy of the optimal filter and the proposed nonlinear filtering scheme.

Keywords: nonlinear filtering, unknown noise intensity, linear stochastic differential system, Kalman-Bucy filter.

References

1. Borisov A., Sokolov I. Optimal filtering of Markov jump processes given observations with state-dependent noises: Exact solution and stable numerical schemes. *Mathematics*. 2020. vol. 8. no. 4.
2. Sinitsyn I.N. Normal suboptimal filtering methods in implicit observable Gaussian stochastic systems. *Sistemy i sredstva informatiki – Systems and Means of Informatics*. 2025. vol. 35. no. 1. pp. 41–58. (In Russ.).
3. Miller B., Kolosov K. [Robust estimation based on the least absolute deviations method and the Kalman filter]. *Automation and Remote Control*. 2020. vol. 81. no. 11. pp. 1994–2010.
4. Kulikova M.V., Kulikov G.Y. On derivative-free extended Kalman filtering and its Matlab-oriented square-root implementations for state estimation in continuous-discrete nonlinear stochastic systems. *European J. Control*. 2023. vol. 73.
5. Bosov A.V., Uryupin I.V. [A modified extended Kalman filter by the linear pseudomeasurement method]. *Inform. i ee primeneniye – Informatics and Applications*. 2025. vol. 19. no. 2. pp. 17–26. (In Russ.).
6. Ananyev B.I. [On estimation of dynamical systems under inexact constraints on parameters]. *Trudy Instituta Matematiki i Mekhaniki UrO RAN – Tr. In-ta matematiki i mekhaniki UrO RAN*. 2025. vol. 31. no. 2. pp. 15–29. (In Russ.).
7. Wang S. Distributionally robust state estimation for nonlinear systems. *IEEE Trans. Signal Process*. 2022. vol. 70. pp. 4408–4423.
8. Stepanov O.A., Isaev A.M. [A procedure of comparative analysis of recursive nonlinear filtering algorithms in navigation data processing based on predictive modeling]. *GyroscoPy Navig*. 2023. vol. 14. pp. 213–224.

9. Olfati-Saber R. Kalman-consensus filter: Optimality, stability, and performance. Proc. 48th IEEE Conf. Decision and Control (CDC'2009) held jointly with 28th Chinese Control Conf. 2009. pp. 7036–7042. DOI: 10.1109/CDC.2009.5399678.
10. Sharma R., Beard R.W., Taylor C.N., Quebe S. Graph-based observability analysis of bearing-only cooperative localization. IEEE Trans. Robotics. 2011. vol. 28. no. 2. pp. 522–529.
11. Wang L., Liu J., Morse A.S. A distributed observer for a continuous-time linear system with time-varying network. arXiv:2003.02134. 2020.
12. Talebi P.S., Mandic D. On the dynamics of multi agent nonlinear filtering and learning. arXiv:2309.03557v2. 2023.
13. Andrievsky B.R., Matveev A.S. Fradkov A.L. [Control and estimation under information constraints: Toward a unified theory of control, computation and communications]. Automation and Remote Control. 2010. vol. 71. no. 4. pp. 572–633.
14. Song W., Wang Z., Li Z., Wang J., Han Q.-L. Nonlinear filtering with sample-based approximation under constrained communication: Progress, insights and trends. IEEE/CAA J. Autom. Sinica. 2024. vol. 11. no. 7. pp. 1539–1556.
15. Voortman Q., Efimov D., Pogromsky A., Richard J.-P., Nijmeijer H. Remote state estimation of steered systems with limited communications: An event-triggered approach. IEEE Trans. Automat. Control. 2024. vol. 69. no. 7. pp. 4199–4214.
16. Tanwani A. Suboptimal filtering over sensor networks with random communication. IEEE Trans. Automat. Control. 2022. vol. 67. no. 10. pp. 5456–5463.
17. Anan'ev B.I. [Minimax mean-square estimates in statistically indeterminate systems]. Differenc. uravnenija – Differential equations. 1984. vol. 20. no. 8. pp. 1291–1297. (In Russ.).
18. Lebedev M.V., Semenikhin K.V. [Minimax filtering in a stochastic differential system with non-stationary perturbations of unknown intensity]. Journal of Computer and System Sciences International. 2007. vol. 46. no. 2. pp. 206–217.
19. Pankov A.R., Platonov E.N., Semenikhin K.V. Robust filtering of a process in the stationary difference stochastic system. Automation and Remote Control. 2011. vol. 72. no. 2. pp. 377–392.
20. Kogan M.M. [Robust estimation and filtering in uncertain linear systems under unknown covariations]. Automation and Remote Control. 2015. vol. 76. no. 10. pp. 1751–1764.
21. Shafieezadeh Abadeh S., Nguyen V.A., Kuhn D., Mohajerin Esfahani P.M. Wasserstein distributionally robust Kalman filtering. Advances in Neural Information Processing Systems. 2018. vol. 31.
22. Kargin T., Hajar J., Malik V., Hassibi B. Distributionally robust Kalman filtering over finite and infinite horizon. arXiv:2407.18837v1. 2024.
23. Barabanov A.E. [Linear filtering with adaptive adjustment of the disturbance covariation matrices in the plant and measurement noise]. Automation and Remote Control. 2016. vol. 77. no. 1. pp. 21–36.
24. Li K., Zhao S., Liu F. Joint state estimation for nonlinear state-space model with unknown time-variant noise statistics. Internat. J. Adaptive Control and Signal Processing. 2021. vol. 35. no. 4. pp. 498–512.
25. Huang Y., Zhang Y., Wu Z., Li N., Chambers J. A novel adaptive Kalman filter with inaccurate process and measurement noise covariance matrices. IEEE Trans. Automat. Control. 2018. vol. 63. no. 2. pp. 594–601.
26. Chen Y., Li W., Wang Y. Online adaptive Kalman filter for target tracking with unknown noise statistics. IEEE Sensors Letters. 2021. vol. 5. no. 3. pp. 1–4.

27. Kanouj M.M., Klovok A.V. [Adaptive unscented Kalman filter for tracking GPS signals in the case of an unknown and time-varying noise covariance]. *Gyroscopy Navig.* 2021. vol. 12. pp. 224–235.
28. Grigor'ev F.N., Kuznetsov N.A. [Quasi-optimal filtering in the presence of unknown noise intensity]. Conference Proceedings: Doklady II-go Vsesoyuznogo soveschaniya po statisticheskim metodam teorii upravleniya [Reports of the II All-Union Meeting on Statistical Methods of Management Theory]. Moscow: Nauka, 1970. pp. 112–118. (In Russ.).
29. Grigor'ev F.N. Metody upravleniya protsessom nablyudeniya v nepreryvnykh sistemakh [Methods for control of observation processes in continuous-time systems]: Ph.D. Diss. kand. tekhn. nauk: 05.13.02. Moscow: Moscow Institute of Physics and Technology, 1975. 104 p. (In Russ.).
30. Teoreticheskiye voprosy postroyeniya ASU krupnotonnazhnymi transportnymi sudami: Sbornik statey [Theoretical issues of constructing ACSs (Automated Control Systems) for large-capacity transport vessels: Collection of articles]. Moscow: Nauka, 1978. 211 p. (In Russ.).
31. Liptser R.S., Shirayev A.N. Statistics of random processes. 3rd edition. New York: Springer, 2005.
32. Polyak B.T., Khlebnikov M.V., Rapoport L.B. Matematicheskaya teoriya avtomaticheskogo upravleniya [Mathematical theory of automatic control]. M.: LENAND, 2019. 500 p. (In Russ.).
33. Kalatchev M.G. [A method of multiple differentiation of signal in automatic control systems]. *Avtomatika i Telemekhanika – Automation and telemechanics.* 1970. no. 6. pp. 29–36. (In Russ.).
34. Miller B.M., Pankov A.R. Teoriya sluchaynykh protsessov v primerakh i zadachakh [Theory of Random Processes: Examples and Problems]. Moscow: Fizmatlit, 2002. 317 p. (In Russ.).

Kuznetsov Nikolay — Ph.D., Dr.Sci., Professor, Academician of the Russian Academy of Sciences, Advisor to the Russian Academy of Sciences, Head of the laboratory, Laboratory no. 351, Kotelnikov Institute of Radioengineering and Electronics of Russian Academy of Sciences (IRE RAS); Head of the department, Department «infocommunication systems and networks», Moscow Institute of Physics and Technology. Research interests: control of infocommunication systems and networks, theory of non-synchronous systems, methods of observation control in stochastic systems, development of identification algorithms for control systems of moving objects, optimization of traffic flows, biomedical data processing. The number of publications — 310. kuznetsov@cplire.ru; 11-7, Mokhovaya St., 125009, Moscow, Russia; office phone: +7(495)629-3005; fax: +7(495)629-3005.

Semenikhin Konstantin — Ph.D., Dr.Sci., Associate Professor, Professor, Educational center of department no. 8 «Computer Science and Applied Mathematics», Moscow Aviation Institute (National Research University) (MAI). Research interests: methods of distributionally robust optimization, control and estimation in stochastic systems, optimization of queuing networks. The number of publications — 125. siemenkv@mail.ru; 4, Volokolamskoye Hwy, 125993, Moscow, Russia; office phone: +7(499)158-4560; fax: +7(499)158-4560.

Acknowledgements. This research is supported by the State assignment: Kotelnikov Institute of Radioengineering and Electronics, RAS (FFWZ-2025-0015).

В.А. Сойфер, Е.В. Гошин
**ВИЗУАЛЬНАЯ ОДОМЕТРИЯ С КОРРЕКЦИЕЙ ТРАЕКТОРИИ
НА ОСНОВЕ ПРИМЕНЕНИЯ ФИЛЬТРА КАЛМАНА**

Сойфер В.А., Гошин Е.В. Визуальная одометрия с коррекцией траектории на основе применения фильтра Калмана.

Аннотация. В настоящей работе предложен метод коррекции траектории камеры в задаче монокулярной визуальной одометрии на основе фильтра Калмана. Рассмотрен подход, в котором неоднозначность масштаба разрешается не через восстановление трёхмерной структуры сцены, а через последовательное использование относительных смещений между текущим и несколькими предыдущими положениями камеры. Такой подход снижает зависимость от долговременного отслеживания одних и тех же ключевых точек и делает метод применимым в условиях ограниченного параллакса, повторяющихся текстур и частичных окклюзий. Подобное представление повышает чувствительность метода к шумам. Для компенсации этого недостатка применяется фильтр Калмана, который предотвращает накопление ошибок при последовательных оценках и обеспечивает корректность работы даже при отсутствии сведений о движении камеры, а также позволяет учитывать качество измерений и их достоверность. В рамках работы представлены модель состояния и модель наблюдений, позволяющие уточнять положение камеры, используя только данные последовательных наблюдений. Состояние формируется в виде скользящего окна из текущего и нескольких предыдущих положений камеры, а наблюдения строятся из направленных относительных смещений, оцениваемых по паре или набору кадров, без восстановления трёхмерной структуры сцены. В отличие от распространённых фильтров Калмана с ограничениями на множественные состояния системы (MSCKF), которые обычно интегрируют данные от инерциальных датчиков, предложенный метод использует исключительно информацию, извлекаемую из последовательности кадров. Эффективность метода подтверждена на синтетических данных в различных сценариях движения.

Ключевые слова: визуальная одометрия, фильтр Калмана, монокулярное зрение, оценка положения, робототехника.

1. Введение. Визуальная одометрия [1 – 3] представляет собой задачу оценки собственного движения камеры или робота на основе последовательности изображений. Точность таких оценок критически важна для автономной навигации [4 – 6], включая управление беспилотными летательными аппаратами [7, 8] и мобильными платформами. Традиционные методы, такие как восстановление структуры из движения (Structure from Motion, SfM) [9] и одновременная локализация и построение карты (Simultaneous Localization and Mapping, SLAM) [10], активно исследуются в последние десятилетия [11]. Известные методы оптимизации траектории, такие как групповая корректировка (bundle adjustment) [12], позволяют минимизировать ошибки репроекции, однако высокая вычислительная сложность ограничивает их применение в реальном времени [13].

Визуальная одометрия исторически развивалась в русле двух парадигм – прямых и признаков методов – с последующим появлением гибридных подходов. Прямые методы минимизируют фотометрическую ошибку между кадрами, что обеспечивает высокую точность при стабильных условиях освещения и достаточном параллаксе. Признаковые методы опираются на детекторы и дескрипторы (например, SIFT/ORB/AKAZE), что делает их устойчивыми к умеренным фотометрическим и геометрическим искажениям, но чувствительными к деградации текстуры. Гибридные методы стремятся объединить преимущества обеих групп, дополняя геометрию фотометрией и наоборот, что особенно важно в динамических сценах и при малом параллаксе.

Задачи SfM/SLAM и одометрии находятся в тесной связи, но отличаются постановкой: SLAM нацелен на совместную оценку траектории и карты, тогда как одометрия обычно ограничивается локальной последовательной оценкой положения и ориентации камеры без явного построения глобальной карты. Это различие приводит к компромиссу между точностью и вычислительной сложностью: в то время как глобальные оптимизации (bundle adjustment) обеспечивают минимизацию ошибок репроекции на больших окнах, они труднее масштабируются для строго реального времени на ограниченных вычислительных ресурсах. В этой связи модели со скользящим окном и рекурсивной фильтрацией выглядят практически привлекательными.

В связи с этим активно исследуются подходы, позволяющие повысить точность и устойчивость при меньших вычислительных затратах. Одним из таких подходов является использование фильтра Калмана [14 – 16], который обеспечивает рекурсивную оценку состояния системы с учётом неопределённостей процесса и измерений. Как правило, фильтр Калмана применяется в составе визуально-инерциальных систем [17, 18], а также в сочетании с данными лидара [19], что обеспечивает высокую точность и устойчивость, но делает систему более сложной. Наибольшее распространение получили фильтры Калмана с ограничениями на множественные состояния системы (multi-state constraint Kalman filter, MSCKF) [20], где объединяются данные от нескольких последовательных положений камеры и инерциальных датчиков. Эти методы демонстрируют высокую точность, но требуют наличия и калибровки инерциальных датчиков, что повышает требования к аппаратной части.

С практической точки зрения, зависимость от инерциальных датчиков приводит к дополнительным этапам межсенсорной

калибровки, синхронизации и подавления смещений и дрейфа акселерометров и гироскопов. Более того, для лёгких и энергоограниченных платформ (микро-БПЛА, наземные мобильные микророботы) минимизация числа сенсоров часто является ключевым требованием. Поэтому варианты, использующие исключительно визуальные данные при сохранении устойчивости оценок, представляют особую ценность для реальных применений.

В настоящей работе рассматривается метод, использующий только визуальную информацию. Он сохраняет идею MSCKF – учитывать несколько последовательных положений камеры, – но устраняет зависимость от дополнительных сенсоров. Такой подход позволяет:

- работать исключительно с изображениями;
- адаптивно учитывать качество измерений (например, число верно сопоставленных точек после применения метода RANSAC [21], ошибку репроекции, величину параллакса);
- обеспечивать устойчивость даже в условиях сложных сцен, когда отслеживание одних и тех же ключевых точек на последовательных кадрах затруднено.

Работа демонстрирует применение фильтра Калмана для задачи коррекции траектории, полученной в результате визуальной одометрии. Предлагается модель состояния в виде скользящего окна из текущего и нескольких предыдущих положений камеры, что позволяет учитывать краткосрочную динамику без явного ввода скоростей и ускорений. Формируется модель наблюдений, основанная на согласовании направлений относительных смещений, оцениваемых по набору кадров, что устраняет необходимость восстановления трёхмерной структуры сцены и уменьшает зависимость от долговременного отслеживания одних и тех же точек. Вводится схема адаптивного взвешивания наблюдений по метрикам качества сопоставления, позволяющая повышать робастность к выбросам и шумам в визуальных данных. Метод ориентирован на работу в реальном времени и не требует инерциальных сенсоров, что снижает аппаратную сложность системы.

Помимо методологической значимости, подход имеет практическую ценность для задач автономной навигации в условиях отсутствия GPS/GNSS (склады, шахты, помещения со сложной геометрией), а также при ограниченном параллаксе (движение вдоль оси оптического луча, повторяющиеся или однородные текстуры). В таких сценариях отказ от восстановления плотной карты и использование направлений смещений в сочетании с фильтрацией по Калману создают устойчивую альтернативу тяжёлым глобальным

оптимизациям. В настоящей работе метод демонстрируется на задаче уточнения положения; при этом общая постановка распространяется и на оценку ориентации камеры.

2. Описание задачи. Целью исследования является разработка метода коррекции траектории камеры на основе последовательности монокулярных изображений с использованием фильтра Калмана. Входными данными служат оценки относительного движения, полученные стандартными методами определения положения и ориентации камеры по паре кадров (например, через фундаментальную или существенную матрицу). Эти оценки, как правило, надёжно задают направление смещения между положениями камеры, но не определяют масштаб (величину смещения).

Предполагается, что камера жёстко закреплена на движущемся носителе (рисунок 1). Для корректной работы монокулярной визуальной одометрии признаки желательны к наблюдению не менее чем в трёх последовательных кадрах. На практике это требование регулярно нарушается (динамические сцены, слабоструктурированные поверхности, изменения освещения), из-за чего традиционная схема [3], предполагающая восстановление трёхмерных точек, быстро теряет устойчивость.

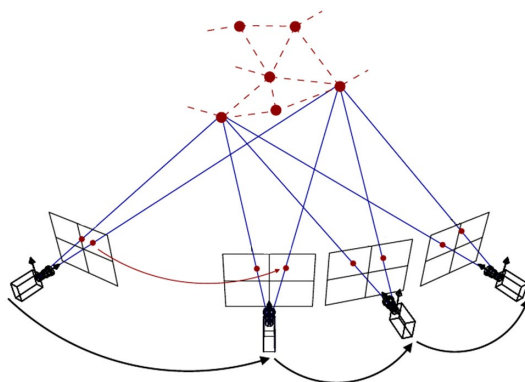


Рис. 1. Модель съёмки

Как правило, процедура оценки траектории движения состоит из следующих шагов:

1. Извлечение признаков на первом и втором кадрах.
2. Сопоставление признаков между кадрами и оценка относительного поворота и смещения (например, 5-точечным алгоритмом). Так как смещение задано с точностью до масштаба, на

этом этапе вводится произвольный масштаб между первыми двумя положениями.

3. Извлечение признаков на следующем (третьем) кадре и сопоставление с ранее отобранными признаками.

4. Оценка относительного движения между текущим положением и несколькими предыдущими (обычно двумя и более), получая направления смещений из каждого предыдущего положения к текущему (рисунок 2).

5. Коррекция текущих координат камеры по информации, накопленной к текущему моменту.

6. Переход к следующему кадру и повторение процедуры, начиная с шага 4.

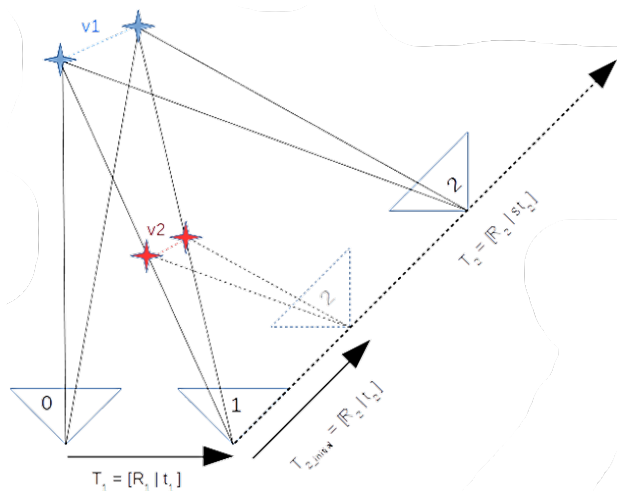


Рис. 2. Коррекция положения камеры

Отличие настоящей работы заключается в том, что на шаге 5 производится согласование направлений из нескольких предыдущих положений. Масштаб между первыми двумя кадрами задаётся произвольно. Далее для каждого последующего кадра рассматриваются несколько направлений смещения, оценённых относительно нескольких предыдущих положений (например, из положения на двух предыдущих к текущему). Геометрически это соответствует согласованию нескольких лучей: искомое положение для текущего положения выбирается как точка, наиболее согласующаяся с лучами, выходящими из предыдущих положений вдоль оценённых направлений. Такой подход не требует

восстановления трёхмерных точек сцены и не опирается на сохранение одних и тех же ключевых точек в соседних кадрах, что делает его применимым в трудных для отслеживания сценах (низкая текстура, повторяющиеся паттерны, быстро меняющиеся детали).

Как правило, в литературе этот способ редко описывается и применяется, поскольку чувствителен к угловым ошибкам направлений и к флуктуациям оценок относительных смещений. В настоящей работе этот недостаток компенсируется фильтром Калмана, который сглаживает случайные отклонения и препятствует накоплению ошибок в траектории, работает исключительно по визуальным данным, не требуя наличия инерциальных датчиков или других дополнительных сенсоров, а также учёт качества измерений, например, по доле выбросов после RANSAC.

3. Описание метода

Уравнение состояния. Для описания движения камеры будем использовать вектор состояния, включающий несколько предыдущих положений. Пусть положение камеры в момент времени t задаётся вектором $\mathbf{p}_t \in \mathbb{R}^3$.

Тогда вектор состояния определяется как

$$\mathbf{x}_t = \begin{bmatrix} \mathbf{p}_t \\ \mathbf{p}_{t-1} \\ \mathbf{p}_{t-2} \\ \vdots \\ \mathbf{p}_{t-k} \end{bmatrix} = \begin{bmatrix} \mathbf{x}_t \\ \mathbf{y}_t \\ \mathbf{z}_t \\ \mathbf{x}_{t-1} \\ \mathbf{y}_{t-1} \\ \mathbf{z}_{t-1} \\ \vdots \\ \mathbf{x}_{t-k} \\ \mathbf{y}_{t-k} \\ \mathbf{z}_{t-k} \end{bmatrix} \in \mathbb{R}^{3(k+1)}.$$

В отличие от традиционных схем, в нашем случае скорость и ускорение не включаются в явном виде в состояние системы, а при необходимости извлекается из последовательности положений камеры. Так, например, скорость может быть извлечена через конечные разности как:

$$\hat{\mathbf{v}}_t = \mathbf{p}_t - \mathbf{p}_{t-1}.$$

Соответственно, прогноз текущего положения выполняется по формуле линейной экстраполяции

$$\mathbf{p}_{t+1} \approx \mathbf{p}_t + \hat{\mathbf{v}}_t = 2\mathbf{p}_t - \mathbf{p}_{t-1}.$$

Уравнение состояния описывает эволюцию системы во времени и записывается в виде:

$$\mathbf{x}_t = \mathbf{F}\mathbf{x}_{t-1} + \mathbf{w}_t,$$

где $\mathbf{F}$ – матрица перехода, экстраполирующая новое состояние по предыдущим данным, $\mathbf{w}_k$ – процессный шум, моделирующий неопределённость прогноза (например, из-за изменения скорости или манёвров).

Поскольку состояние $\mathbf{x}_t$ состоит из текущей позиции и k предыдущих, структура $\mathbf{F}$ естественным образом блочная. Очевидно, что для нового состояния $\mathbf{x}_{t+1}$ положения камеры в предыдущие моменты времени $\mathbf{p}_{t+1-i}$, $i = \overline{1, k}$ переносятся из предыдущего шага без изменений, то есть нижняя часть матрицы содержит единичные 3×3 блока на поддиагонали, обеспечивая сдвиг состояний.

Текущее положение экстраполируется с использованием двух последних точек траектории $\mathbf{p}_t \approx 2\mathbf{p}_{t-1} - \mathbf{p}_{t-2}$. Соответственно, верхняя блок-строка $\mathbf{F}$ имеет ненулевые коэффициенты при блоках, соответствующих $\mathbf{p}_{t-1}$ и $\mathbf{p}_{t-2}$ (например, $2\mathbf{I}$ и $-\mathbf{I}$) тогда как остальные блоки в этой строке равны нулю. Таким образом, матрица $\mathbf{F}$ приобретает вид

$$\mathbf{F} = \begin{bmatrix} 2\mathbf{I} & -\mathbf{I} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} \\ \mathbf{I} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{I} & \mathbf{0} \end{bmatrix},$$

где $\mathbf{I}$ – единичная матрица размерности 3×3 , $\mathbf{0}$ – нулевая матрица размерности 3×3 .

Такое построение матрицы перехода одновременно остаётся простым вычислительно и позволяет учитывать краткосрочную динамику движения без явного введения скоростей и ускорений в состояние. За счёт процессного шума $\mathbf{w}$ модель остаётся гибкой и устойчивой к отклонениям от идеальной модели постоянной скорости.

При необходимости точность прогноза можно повысить введением дополнительных слагаемых, учитывающих ускорение или повороты камеры. Следует добавить, что предложенная схема

позволяет гибко увеличивать окно учитываемых позиций в зависимости от доступной памяти и вычислительных возможностей. Например, при наличии аппаратного ускорения (параллельной обработки или GPU) можно расширить окно до 6–8 позиций, что позволяет более надёжно сглаживать шум и предсказывать траекторию в условиях низкого параллакса. На практике размер окна следует выбирать как компромисс между точностью и допустимой задержкой в конкретной системе реального времени.

Ковариационная матрица процесса. Ковариационная матрица ошибки состояния $\mathbf{P}_t$ отражает нашу уверенность в прогнозе и на каждом шаге пересчитывается с учётом динамики модели. В рамках фильтра Калмана это делается по стандартной формуле

$$\mathbf{P}_t = \mathbf{F}\mathbf{P}_{t-1}\mathbf{F}^T + \mathbf{Q}_t,$$

где первый член переносит накопленную неопределённость через матрицу перехода $\mathbf{F}$, а $\mathbf{Q}_t$ моделирует неточность прогноза, связанную с неожиданными изменениями движения.

Учитывая структуру состояния (текущее и k предыдущих положений), естественно задавать $\mathbf{Q}_t$ так, чтобы шум процесса влиял прежде всего на текущую компоненту, а прошлые позиции лишь переносились со своей уже накопленной неопределённостью. Практически это выражается в блочно-диагональной форме

$$\mathbf{Q}_t = \begin{bmatrix} q\mathbf{I} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} \end{bmatrix},$$

где $\mathbf{I}$ – единичная матрица размерности 3×3 , а $q > 0$ – коэффициент, отвечающий за погрешность прогнозирования.

Чем более изменчивым ожидается движение (резкие ускорения, манёвры), тем большим следует выбирать q ; при плавной, близкой к равномерной траектории q можно уменьшать. Так, если движение камеры предсказуемо и плавно (например, при съёмке с робота на колёсах), величина дополнительного шумового члена может быть уменьшена для повышения точности оценки. В противоположность этому, при резких манёврах разумно увеличить ковариацию процесса, чтобы фильтр быстрее реагировал на новые измерения.

Такой выбор $\mathbf{Q}_t$ позволяет явно учитывать неопределённость в экстраполяции текущего положения, не увеличивая ковариацию для уже наблюждённых поз прошлого окна, а также сохранять простоту модели и стабильность оценки, поскольку дополнительные корреляции между блоками вносит сама операция $\mathbf{F}\mathbf{P}_{t-1}\mathbf{F}^T$.

При наличии априорных сведений о направлении большей вариабельности (например, ограничения по высоте или доминирующее движение в плоскости) матрицу $\mathbf{Q}_t$ можно сделать анизотропной, заменив $\mathbf{Q}_t = q\mathbf{I}$ на $\mathbf{Q}_t = \text{diag}(q_x, q_y, q_z)$ с различными дисперсиями по осям. Если шум естественнее задавать в системе камеры, допустимо учитывать ориентацию и задавать $\mathbf{Q}_t = \mathbf{R}\text{diag}(q_x, q_y, q_z)\mathbf{R}^T$ для текущего блока. В любом случае ключевая идея сохраняется: на каждом шаге ковариация процесса добавляется прежде всего к компоненте, отвечающей за текущее положение, что позволяет фильтру гибко реагировать на изменения характера движения.

Уравнение наблюдения. Уравнение наблюдения связывает состояние системы с измерениями:

$$\mathbf{z}_t = \mathbf{H}_t\mathbf{x}_t + \mathbf{v}_t,$$

где:

$\mathbf{x}_t$ – вектор измерений (проекция точек сцены на изображение);

$\mathbf{H}_t$ – матрица наблюдения;

$\mathbf{v}_t$ – шум измерений.

Пусть в каждый новый момент времени t методами визуальной одометрии выполняется оценка относительного смещения текущего положения камеры относительно нескольких предыдущих. На практике визуальная одометрия надёжно определяет направление смещения $\mathbf{d}$ (как правило, нормированное, $\|\mathbf{d}\| = 1$) от положения $\mathbf{p}_i, i \in \{t-1, t-k\}$ к новому положению $\mathbf{p}_t$, но абсолютное значение (масштаб) этого смещения неизвестно.

Таким образом, задана прямая, на которой расположены новые координаты, через предыдущие известные координаты $\mathbf{p}_{t-i}$ и направление $\mathbf{d}_{t,i}$ с точностью до неизвестного коэффициента α :

$$\mathbf{p}_t = \mathbf{p}_{t-i} + \alpha\mathbf{d}_{t,i},$$

Рассмотрим проекцию

$$proj_{\mathbf{d}_{t,i}}(\mathbf{p}_t - \mathbf{p}_{t-i}) = \frac{(\mathbf{p}_t - \mathbf{p}_{t-i}) \cdot \mathbf{d}_{t,i}}{\mathbf{d}_{t,i} \cdot \mathbf{d}_{t,i}} \mathbf{d}_{t,i}$$

вектора смещения $\mathbf{p}_t - \mathbf{p}_{t-i}$ на направление $\mathbf{d}_{t,i}$. В случае, если направление $\mathbf{d}_{t,i}$ и векторы смещения вычислены точно, эта проекция равна:

$$proj_{\mathbf{d}_{t,i}}(\mathbf{p}_t - \mathbf{p}_{t-i}) = \frac{(\alpha \mathbf{d}_{t,i}) \cdot \mathbf{d}_{t,i}}{\mathbf{d}_{t,i} \cdot \mathbf{d}_{t,i}} \mathbf{d}_{t,i} = \alpha \mathbf{d}_{t,i} = \mathbf{p}_t - \mathbf{p}_{t-i}.$$

Таким образом, для каждой пары $\mathbf{p}_t, \mathbf{p}_{t-i}$ можно задать наблюдение в момент времени t в виде:

$$\begin{aligned} \mathbf{h}_{t,i} &= proj_{\mathbf{d}_{t,i}}(\mathbf{p}_t - \mathbf{p}_{t-i}) - (\mathbf{p}_t - \mathbf{p}_{t-i}) = (\mathbf{p}_t - \mathbf{p}_{t-i}) \cdot \mathbf{d}_{t,i} \cdot \mathbf{d}_{t,i} - \\ &(\mathbf{p}_t - \mathbf{p}_{t-i}) = (\mathbf{d}_{t,1} \mathbf{d}_{t,1}^T - \mathbf{I}) \mathbf{p}_t - (\mathbf{d}_{t,1} \mathbf{d}_{t,1}^T - \mathbf{I}) \mathbf{p}_{t-i} = \mathbf{0}. \end{aligned}$$

Для каждого предыдущего состояния формируется одно наблюдение такого рода, в результате чего итоговая матрица наблюдения приобретает вид:

$$\mathbf{H}_k = \begin{bmatrix} \mathbf{d}_{t,1} \mathbf{d}_{t,1}^T - \mathbf{I} & -\mathbf{d}_{k-1} \mathbf{d}_{k-1}^T + \mathbf{I} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{d}_{t,2} \mathbf{d}_{t,2}^T - \mathbf{I} & \mathbf{0} & -\mathbf{d}_{k-2} \mathbf{d}_{k-2}^T + \mathbf{I} & \dots & \mathbf{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{d}_{t,k} \mathbf{d}_{t,k}^T - \mathbf{I} & \mathbf{0} & \mathbf{0} & \dots & -\mathbf{d}_{t,k} \mathbf{d}_{t,k}^T + \mathbf{I} \end{bmatrix}.$$

Такая модель наблюдений удобна для интеграции с существующими блоками визуальной одометрии, поскольку не требует модификации внутренней логики сопоставления признаков. Фактически она позволяет использовать уже имеющиеся результаты относительных оценок движения и лишь дополнять их статистическим уточнением через фильтр Калмана.

Ковариационная матрица шума измерений. Поскольку наблюдения формируются на основе результатов визуальной одометрии и неизбежно содержат ошибки (неточности сопоставления, угловые ошибки направлений, неопределённость масштаба), их достоверность необходимо учесть. Ковариационная матрица шума измерений $\mathbf{V}_t$ задаётся как

$$\mathbf{R}_t = \begin{bmatrix} \sigma_{t-1}^2 \mathbf{I} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \sigma_{t-2}^2 \mathbf{I} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \sigma_{t-3}^2 \mathbf{I} & \cdots & \mathbf{0} & \mathbf{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \sigma_{t-k}^2 \mathbf{I} \end{bmatrix},$$

где σ_{t-i}^2 – дисперсия i -го скалярного измерения z_i . Тем самым каждое наблюдение получает собственный вес в зависимости от качества оценки.

Практически σ_{t-i}^2 могут быть одинаковыми при отсутствии информации о качестве визуальной одометрии, либо выбираться на основе доступных метрик, например:

- количество сопоставленных ключевых точек и доля выбросов (после RANSAC);
- средняя или медианная ошибка репроекции;
- оценка неопределённости существенной или фундаментальной матрицы;
- величина параллакса (большой параллакс – меньшая неопределённость).

Нормировка направлений $\mathbf{d}_{t,i}$ делает наблюдения z_i сопоставимыми по шкале, что упрощает калибровку $\mathbf{R}_t$. При наличии априорной информации допускается адаптивная схема: $\mathbf{R}_t$ пересчитывается на каждом шаге по выбранным метрикам качества, благодаря чему фильтр автоматически уменьшает влияние ненадёжных наблюдений и усиливает вклад более точных. Такой механизм повышает устойчивость оценки в условиях переменного уровня шума и наличия выбросов.

В итоге совокупность представленных уравнений состояния, наблюдений и ковариаций образует полную схему фильтра Калмана, предназначенного для визуальной одометрии без инерциальных данных. Такое построение обеспечивает баланс между простотой реализации и достаточной точностью, позволяя применять метод как на мощных вычислительных платформах, так и на встраиваемых устройствах.

4. Экспериментальное исследование на синтетических данных. Для демонстрации принципиальной работоспособности подхода использовались генерированные данные. Такой формат позволяет контролировать траекторию движения, уровень шума и параметры наблюдений, а значит – корректно сравнивать методы в одинаковых условиях.

Моделирование движения аппарата выполнялось в трёх сценариях:

- линейная траектория с незначительными отклонениями;
- круговая траектория;
- произвольная траектория со случайным изменением скорости и направления.

Для каждой из этих траекторий моделирование движения аппарата производилось следующим образом.

Линейная траектория с незначительными отклонениями:

$$\mathbf{p}_{i+1} = \mathbf{p}_i + \Delta \mathbf{p} \cdot v_i + \mathcal{N}(0, \sigma_p^2),$$

$$v_{i+1} = v_i + \mathcal{N}(0, \sigma_v^2),$$

где $\mathbf{p}_i$ – точка траектории, $\Delta \mathbf{p}$ – направление движения, v_i – абсолютное значение скорости камеры, σ_p^2 – дисперсия отклонения от траектории, σ_v^2 – дисперсия скорости. Изначальная скорость v_i задаётся при моделировании. Начальное положение точки задано как $(0, 0, 0)$. Такая траектория позволяет оценить качество траектории для случаев, когда соответствующие лучи практически коллинеарны.

Круговая или эллиптическая траектория:

$$\mathbf{p}_i = c + (\cos(\psi_i) \cdot R_x + \sin(\psi_i) \cdot R_y) + \mathcal{N}(0, \sigma_p^2),$$

$$\psi_i = \psi_{i+1} + v_i,$$

$$v_{i+1} = v_i + \mathcal{N}(0, \sigma_v^2),$$

где $\mathbf{p}_i$ – точка траектории, c – центр круга или эллипса, R_x, R_y – полуоси эллипса, ψ_i – угловое положение камеры, v_i – угловая скорость камеры, σ_p^2 – дисперсия отклонения от траектории, σ_v^2 – дисперсия скорости. Изначальная скорость v_i задаётся при моделировании.

Произвольная траектория со случайным изменением направления и абсолютного значения скорости:

$$\mathbf{p}_{i+1} = \mathbf{p}_i + \mathbf{v}_i + \mathcal{N}(0, \sigma_p^2),$$

$$\mathbf{v}_{i+1} = \mathbf{v}_i + \mathcal{N}(0, \sigma_v^2),$$

где $\mathbf{p}_i$ – точка траектории, $\mathbf{v}_i$ – скорость камеры, σ_p^2 – дисперсия отклонения от траектории, σ_v^2 – дисперсия скорости. Изначальная скорость v_i задаётся при моделировании. Начальное положение точки

задано как $(0, 0, 0)$. Такая траектория позволяет оценить качество траектории для случаев, когда скорость значительно меняется во времени как по амплитуде, так и по направлению, вследствие чего погрешность прогноза нового состояния системы велика.

С использованием известных положений камеры были вычислены точные значения направлений $\mathbf{d}_{t,i}$, после чего они зашумлялись аддитивным гауссовым шумом, моделирующим погрешности визуальной одометрии (ошибки сопоставления). Интенсивность шума варьировалась через параметр SNR.

Было рассмотрено два подхода к определению следующего положения камеры:

1. Метод минимизации расстояний. Используется k предыдущих положений аппарата и направление смещения относительно каждого. Минимизируется сумма расстояний от соответствующих прямых.

$$\mathbf{p}_t = \min \|\sum dist(l_{t,k}, \mathbf{p})\|,$$

$l_{t,k}$ – прямая, заданная точкой $\mathbf{p}_{t-k}$ и направлением $\mathbf{d}_{t,k}$, $dist(l_{t,k}, \mathbf{p})$ – расстояние от точки до прямой.

2. Предложенный метод (фильтр Калмана). Используется k предыдущих положений аппарата и направление смещение относительно каждого. Применяется фильтр Калмана для уточнения положения.

Также иногда применяется проход, при котором используется только одно предыдущее положение аппарата и направление смещения. Поскольку при изменении скорости такой подход демонстрирует нестабильность и существенные ошибки, в ходе эксперимента он не использовался.

В качестве метрики качества далее используется отклонение по всем точкам траектории; анализ проводится для всех трёх типов движения и различных уровней шума. Далее будут представлены результаты сравнения традиционного и предложенного методов в условиях неизвестных свойств шума.

Оценка качества проводилась на синтетических траекториях для трёх сценариев движения: линейного с незначительными отклонениями, кругового и произвольного с изменяющейся скоростью и направлением. Для каждого сценария рассматривались различные уровни шума наблюдений ($\text{SNR} = 50, 60, 70$). В качестве метрики использовалось среднее отклонение по всем точкам траектории.

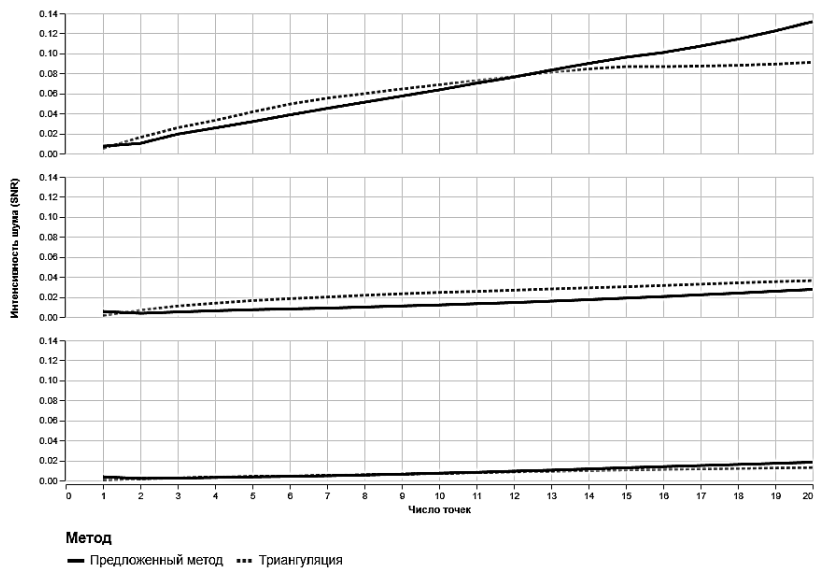


Рис. 3. Линейная траектория с незначительными отклонениями, зависимость среднего отклонения от номера кадра (сверху вниз, SNR: 50, 60, 70)

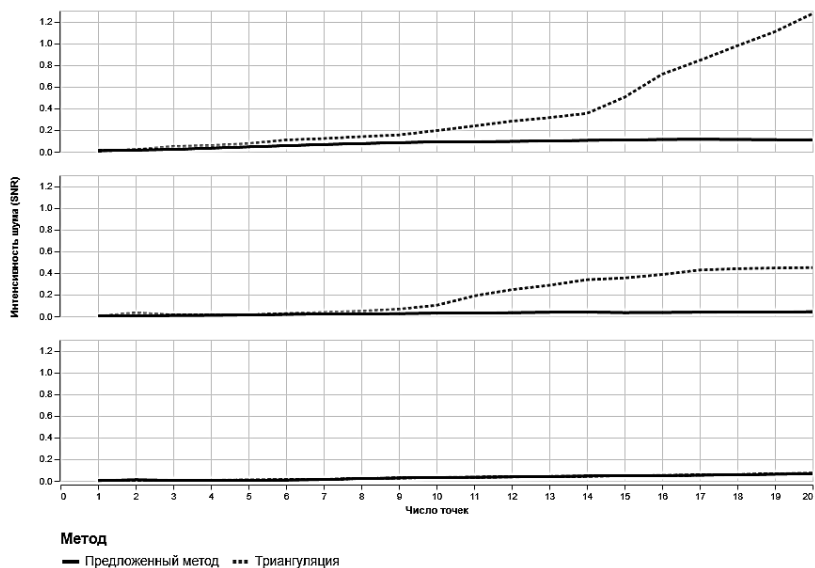


Рис. 4. Круговая траектория, зависимость среднего отклонения от номера кадра (сверху вниз, SNR: 50, 60, 70)

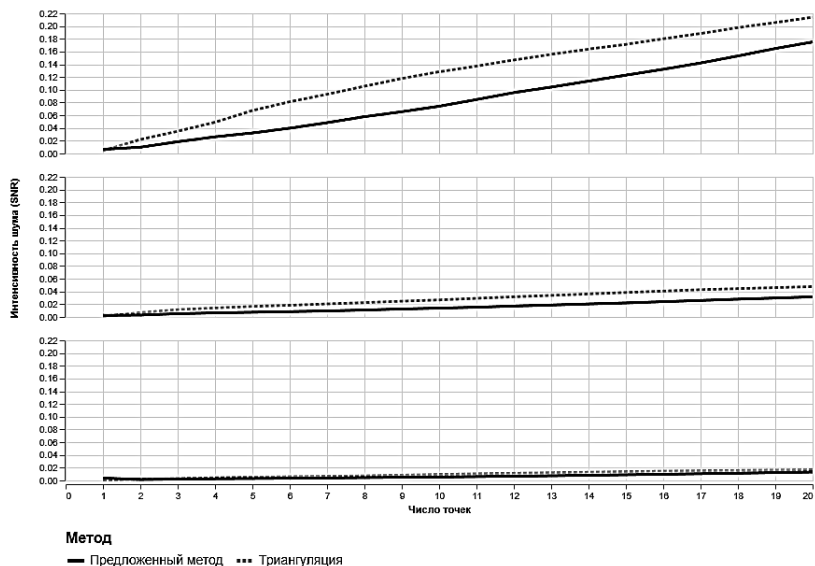


Рис. 5. Произвольная траектория со случайным изменением скорости, зависимость среднего отклонения от номера кадра (сверху вниз, SNR: 50, 60, 70)

Экспериментальное исследование показало, что предложенный метод на основе фильтра Калмана обеспечивает более устойчивую и точную оценку положения по сравнению с традиционной триангуляцией. Наиболее заметное преимущество наблюдается на круговой траектории и при высоких уровнях шума, что указывает на робастность предлагаемого решения к неблагоприятным условиям.

5. Заключение. В данной работе предложен метод коррекции траектории камеры в задаче визуальной одометрии с использованием фильтра Калмана. Предложены модель состояний, матрица перехода и модель наблюдений, обеспечивающие решение поставленной задачи.

Экспериментально показано, что предложенный метод обеспечивает более высокую точность оценки положения по сравнению с базовыми подходами, особенно в условиях переменного шума и наличия выбросов в данных. Это делает его перспективным для применения в реальных системах визуальной одометрии, включая автономные транспортные средства, БПЛА и мобильных роботов.

Основными достоинствами предложенного метода являются:

- Возможность работы только с визуальными данными, без необходимости использования дополнительных сенсоров.

- Адаптивная настройка параметров фильтра на основе оценки достоверности измерений.
- Высокая устойчивость к шумам и выбросам в данных.
- Возможность работы в реальном времени благодаря низкой вычислительной сложности.

В дальнейшем планируется провести тестирование в реальных условиях для оценки применимости метода в практических сценариях.

Для авторов большая честь участвовать данной статьёй в журнале, посвящённом памяти выдающегося учёного, члена-корреспондента РАН Рафаэля Мидхатовича Юсупова.

Тематика статьи относится к научным интересам Р.М. Юсупова, который проводил исследования в области теории управления полётом летательных аппаратов, теории адаптивных систем.

С Самарой и Самарским университетом имени академика С.П. Королева Рафаэль Мидхатовича связывали многолетние творческие и дружеские отношения. Будучи волжанином, Р.М. Юсупов часто бывал в Самаре, принимал участие в научных конференциях, общался с профессорско-преподавательским составом, научными работниками и студентами.

Творческое наследие Р.М. Юсупова способствует развитию научных исследований, а обаяние его личности остаётся в нашей памяти.

Литература

1. Neyestani A., Picariello F., Ahmed I., Daponte P., De Vito L. From pixels to precision: A survey of monocular visual odometry in digital twin applications // *Sensors*. 2024. vol. 24. no. 4.
2. Yuan S., Zhang J., Lin Y., Yang L. Hybrid self-supervised monocular visual odometry system based on spatio-temporal features // *Electronic Research Archive*. 2024. vol. 32. no. 5. pp. 3543–3568.
3. Scaramuzza D., Fraundorfer F. Visual odometry [tutorial] // *IEEE Robotics & Automation Magazine*. 2011. vol. 18. no. 4. p. 80–92.
4. Ou Y., Cai Y., Sun Y., Qin T. Autonomous navigation by mobile robot with sensor fusion based on deep reinforcement learning // *Sensors*. 2024. vol. 24. no. 12.
5. Лебедев А.О., Васильев В.В., Паулиш А.Г. Алгоритм управления полетом БПЛА вдоль железной дороги с помощью технического зрения // *Компьютерная оптика*. 2025. Т. 49. № 2. С. 320–326. DOI: 10.18287/2412-6179-CO-1532.
6. Belkin I.V., Abramenko A.A., Bezuglyi V.D., Yudin D.A. Localization of mobile robot in prior 3D LiDAR maps using stereo image sequence // *Computer Optics*. 2024. vol. 48. no. 3. pp. 406–417.
7. Горбачев В.А., Калугин В.Ф. Разработка алгоритма многообъектного трекинга с необучаемыми признаками сопоставления объектов // *Компьютерная оптика*. 2023. Т. 47. № 6. С. 1002–1010. DOI: 10.18287/2412-6179-CO-1275.
8. Ляхов П.А., Оразаев А.Р. Анализ видеоданных беспилотного летательного аппарата на основе индекса структурного сходства // *Компьютерная оптика*. 2025. Т. 49. № 4. С. 624–633. DOI: 10.18287/2412-6179-CO-1569.

9. Jiang S., Jiang C., Jiang W. Efficient structure from motion for large-scale UAV images: A review and a comparison of SfM tools // *ISPRS Journal of Photogrammetry and Remote Sensing*. 2020. vol. 167. pp. 230–251.
10. Le V.H. Visual SLAM and Visual Odometry Based on RGB-D Images Using Deep Learning: A Survey // *Journal of Robotics and Control (JRC)*. 2024. vol. 5. no. 4. pp. 1050–1079.
11. Yan F., Li Z., Zhou Z. Robust and efficient edge-based visual odometry // *Computational Visual Media*. 2022. vol. 8. no. 3. pp. 467–481.
12. Weber S., Demmel N., Chan T.C., Cremers D. Power bundle adjustment for large-scale 3D reconstruction // *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 2023. pp. 281–289.
13. Triggs B., McLauchlan P., Hartley R., Fitzgibbon A. Bundle adjustment – A modern synthesis // *Vision Algorithms: Theory and Practice. Lecture Notes in Computer Science*. 2000. vol. 1883. pp. 298–372. DOI: 10.1007/3-540-44480-7_21.
14. Сойфер В.А., Фурсов В.А., Харитонов С.И. Калмановская фильтрация одного класса изображений динамических объектов // *Информатика и автоматизация*. 2024. Т. 23. № 4. С. 953–968. DOI: 10.15622/ia.23.4.1.
15. Tavares Jr A.J.A., Oliveira N.M.F. A Novel Approach for Kalman Filter Tuning for Direct and Indirect Inertial Navigation System/Global Navigation Satellite System Integration // *Sensors*. 2024. vol. 24. no. 22. DOI: 10.3390/s24227331.
16. Yuan Y., Li F., Chen J., Wang Y., Liu K. An improved Kalman filter algorithm for tightly GNSS/INS integrated navigation system // *Mathematical Biosciences and Engineering*. 2024. vol. 21. pp. 963–983.
17. Wang Z., Pang B., Song Y., Yuan X., Xu Q., Li Y. Robust visual-inertial odometry based on a Kalman filter and factor graph // *IEEE Transactions on Intelligent Transportation Systems*. 2023. vol. 24. no. 7. pp. 7048–7060.
18. Nguyen K.D., Tran D.T., Nguyen D.T., Inoue K., Lee J.H., Nguyen A.Q. Learning Visual-Inertial Odometry with Robocentric Iterated Extended Kalman Filter // *IEEE Access*. 2024. vol. 12. pp. 109943–109956.
19. Lee D., Jung M., Yang W., Kim A. Lidar odometry survey: recent advancements and remaining challenges // *Intelligent Service Robotics*. 2024. vol. 17. no. 2. pp. 95–118.
20. Xue C., Huang Y., Zhao C., Li X., Mihaylova L., Li Y., Chambers J.A. A Gaussian-generalized-inverse-Gaussian joint-distribution-based adaptive MSCKF for visual-inertial odometry navigation // *IEEE Transactions on Aerospace and Electronic Systems*. 2022. vol. 59. no. 3. pp. 2307–2328.
21. Fischler M.A., Bolles R.C. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography // *Communications of the ACM*. 1981. vol. 24. no. 6. pp. 381–395.

Сойфер Виктор Александрович — д-р техн. наук, профессор, академик РАН, лауреат Государственной премии Российской Федерации в области науки и техники, президент, Самарский национальный исследовательский университет имени академика С.П. Королёва. Область научных интересов: компьютерная оптика, обработка изображений, распознавание образов. Число научных публикаций — 871. soifer@ssau.ru; Московское шоссе, 34, 443086, Самара, Россия; р.т.: +7(846)332-2604.

Гошин Егор Вячеславович — канд. техн. наук, доцент, заведующий кафедрой, кафедра киберфотоники, Самарский национальный исследовательский университет имени академика С.П. Королёва. Область научных интересов: обработка изображений. Число научных публикаций — 78. goshine@ssau.ru; Молодогвардейская улица, 151, 443001, Самара, Россия; р.т.: +7(846)332-2606.

Поддержка исследований. Работа выполнена в рамках государственного задания по теме FSSS-2023-0006.

V. SOIFER, Y. GOSHIN
**VISUAL ODOMETRY WITH TRAJECTORY CORRECTION
BASED ON KALMAN FILTER**

Soifer V., Goshin Y. Visual Odometry with Trajectory Correction Based on Kalman Filter.

Abstract. This paper proposes a Kalman-filter-based method for camera trajectory correction in monocular visual odometry. The approach resolves the scale ambiguity not by reconstructing the 3D structure of the scene, but by sequentially using relative displacements between the current and several previous camera positions. This mechanism reduces reliance on long-term tracking of the same keypoints and makes the method applicable under limited parallax, repetitive textures, and partial occlusions. However, this formulation increases the method's sensitivity to noise. To compensate for this drawback, a Kalman filter is employed, which prevents error accumulation during consecutive estimates, ensures correct operation even in the absence of information about camera motion, and allows measurement quality and reliability to be taken into account. A state model and an observation model are proposed, which allow for refining the camera pose using only data from consecutive observations. The state is represented as a sliding window comprising the current and several preceding camera poses, while the observations are constructed from directions of relative displacements estimated from an image pair or a short set of frames, without reconstructing the three-dimensional structure of the scene. Unlike common Multi-State Constraint Kalman Filters (MSCKF), which typically integrate inertial measurements, the proposed method relies exclusively on information extracted from the image sequence. The effectiveness of the proposed method is confirmed with synthetic data in various motion scenarios.

Keywords: visual odometry, Kalman filter, monocular vision, position estimation, robotics.

References

1. Neyestani A., Picariello F., Ahmed I., Daponte P., De Vito L. From pixels to precision: A survey of monocular visual odometry in digital twin applications. *Sensors*. 2024. vol. 24. no. 4.
2. Yuan S., Zhang J., Lin Y., Yang L. Hybrid self-supervised monocular visual odometry system based on spatio-temporal features. *Electronic Research Archive*. 2024. vol. 32. no. 5. pp. 3543–3568.
3. Scaramuzza D., Fraundorfer F. Visual odometry [tutorial]. *IEEE Robotics & Automation Magazine*. 2011. vol. 18. no. 4. p. 80–92.
4. Ou Y., Cai Y., Sun Y., Qin T. Autonomous navigation by mobile robot with sensor fusion based on deep reinforcement learning. *Sensors*. 2024. vol. 24. no. 12.
5. Lebedev A.O., Vasil'ev V.V., Paulish A.G. [Algorithm for UAV flight controlling along a railway using technical vision]. *Komp'yuternaya optika – Computer Optics*. 2025. vol. 49. no. 2. pp. 320–326. DOI: 10.18287/2412-6179-CO-1532. (In Russ.).
6. Belkin I.V., Abramenko A.A., Bezuglyi V.D., Yudin D.A. Localization of mobile robot in prior 3D LiDAR maps using stereo image sequence. *Computer Optics*. 2024. vol. 48. no. 3. pp. 406–417.
7. Gorbachev V.A., Kalugin V.F. [Development of a multi-object tracking algorithm with untrained features of object matching]. *Komp'yuternaya optika – Computer Optics*. 2023. vol. 47. no. 6. pp. 1002–1010. DOI: 10.18287/2412-6179-CO-1275. (In Russ.).
8. Ljahov P.A., Orazzaev A.R. [Analysis of video data from an unmanned aerial vehicle based on a structural similarity index]. *Komp'yuternaya optika – Computer Optics*. 2025. vol. 49. no. 4. pp. 624–633. DOI: 10.18287/2412-6179-CO-1569. (In Russ.).

9. Jiang S., Jiang C., Jiang W. Efficient structure from motion for large-scale UAV images: A review and a comparison of SfM tools. *ISPRS Journal of Photogrammetry and Remote Sensing*. 2020. vol. 167. pp. 230–251.
10. Le V.H. Visual SLAM and Visual Odometry Based on RGB-D Images Using Deep Learning: A Survey. *Journal of Robotics and Control (JRC)*. 2024. vol. 5. no. 4. pp. 1050–1079.
11. Yan F., Li Z., Zhou Z. Robust and efficient edge-based visual odometry. *Computational Visual Media*. 2022. vol. 8. no. 3. pp. 467–481.
12. Weber S., Demmel N., Chan T.C., Cremers D. Power bundle adjustment for large-scale 3D reconstruction. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 2023. pp. 281–289.
13. Triggs B., McLauchlan P., Hartley R., Fitzgibbon A. Bundle adjustment – A modern synthesis. *Vision Algorithms: Theory and Practice. Lecture Notes in Computer Science*. 2000. vol. 1883. pp. 298–372. DOI: 10.1007/3-540-44480-7_21.
14. Soifer V.A., Fursov V.A., Kharitonov S.I. [Kalman filtering of one class of dynamic object images]. *Informatics and Automation*. 2024. vol. 23. no. 4. pp. 953–968. DOI: 10.15622/ia.23.4.1. (In Russ.).
15. Tavares Jr A.J.A., Oliveira N.M.F. A Novel Approach for Kalman Filter Tuning for Direct and Indirect Inertial Navigation System/Global Navigation Satellite System Integration. *Sensors*. 2024. vol. 24. no. 22. DOI: 10.3390/s24227331.
16. Yuan Y., Li F., Chen J., Wang Y., Liu K. An improved Kalman filter algorithm for tightly GNSS/INS integrated navigation system. *Mathematical Biosciences and Engineering*. 2024. vol. 21. pp. 963–983.
17. Wang Z., Pang B., Song Y., Yuan X., Xu Q., Li Y. Robust visual-inertial odometry based on a Kalman filter and factor graph. *IEEE Transactions on Intelligent Transportation Systems*. 2023. vol. 24. no. 7. pp. 7048–7060.
18. Nguyen K.D., Tran D.T., Nguyen D.T., Inoue K., Lee J.H., Nguyen A.Q. Learning Visual-Inertial Odometry with Robocentric Iterated Extended Kalman Filter. *IEEE Access*. 2024. vol. 12. pp. 109943–109956.
19. Lee D., Jung M., Yang W., Kim A. Lidar odometry survey: recent advancements and remaining challenges. *Intelligent Service Robotics*. 2024. vol. 17. no. 2. pp. 95–118.
20. Xue C., Huang Y., Zhao C., Li X., Mihaylova L., Li Y., Chambers J.A. A Gaussian-generalized-inverse-Gaussian joint-distribution-based adaptive MSCKF for visual-inertial odometry navigation. *IEEE Transactions on Aerospace and Electronic Systems*. 2022. vol. 59. no. 3. pp. 2307–2328.
21. Fischler M.A., Bolles R.C. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*. 1981. vol. 24. no. 6. pp. 381–395.

Soifer Victor — Ph.D., Dr.Sci., Professor, Academician of the RAS, Recipient of the Russian Federation State Prize in science and technology, President, Samara National Research University. Research interests: computer optics, image processing, pattern recognition. The number of publications — 871. soifer@ssau.ru; 34, Moscow Hwy, 443086, Samara, Russia; office phone: +7(846)332-2604.

Goshin Yegor — Ph.D., Associate Professor, Head of the department, Department of cyberphotonics, Samara National Research University. Research interests: image processing. The number of publications — 78. goshine@ssau.ru; 151, Molodogvardeyskaya St., 443001, Samara, Russia; office phone: +7(846)332-2606.

Acknowledgements. The research was carried out within the state assignment theme FSSS-2023-0006.

С.И. СМАГИН, Ю.А. ОЗАРЯН, В.И. УСИКОВ, А.С. СМАГИН,
К.С. ЦЫГУЛЁВ, Р.А. СЕКРИЕРУ, С.А. ОРЛОВ, Т.В. КОЖЕВНИКОВА
**АВТОМАТИЗАЦИЯ ДИСТАНЦИОННОГО ЭКОЛОГИЧЕСКОГО
МОНИТОРИНГА ОБЪЕКТОВ МИНЕРАЛЬНО-СЫРЬЕВОГО
КОМПЛЕКСА ХАБАРОВСКОГО КРАЯ**

Смагин С.И., Озарян Ю.А., Усиков В.И., Смагин А.С., Цыгулёв К.С., Секриеру Р.А., Орлов С.А., Кожевникова Т.В. Автоматизация дистанционного экологического мониторинга объектов минерально-сырьевого комплекса Хабаровского края.

Аннотация. В работе рассматривается проблема автоматизации горно-экологического мониторинга объектов минерально-сырьевого комплекса с использованием данных дистанционного зондирования Земли. Обработка и анализ этих данных выполняется с целью оценки влияния процессов добычи на состояние природной среды в местах размещения горнодобывающих предприятий и прилегающих территорий. Представлен макет информационной системы, включающий модули управления данными, обработки и анализа спутниковых снимков. Для автоматизации выделения интересующих объектов горнопромышленного комплекса применяются сверточные и трансформерные модели нейронных сетей. Модели сегментации дражных отвалов, карьеров, хвостохранилищ и угольного пылевого загрязнения были обучены на подготовленных выборках, включающих размеченные спутниковые снимки исследуемых объектов в семи субъектах Российской Федерации. Предложены методы оценки влияния угольного пылевого загрязнения на растительность на основе вегетационных индексов; картирования типов леса с использованием полносвязной нейронной сети; расчета объемов дражных отвалов на основе цифровых моделей рельефа, центральной линии полигона и алгоритма построения диаграммы Вороного; и содержания тяжелых металлов в почве на основе анализа проб, отобранных за многолетний период, методами математической статистики. Сделан корреляционный анализ данных о содержании элементов в образцах почвогрунтов и воде в зоне воздействия добычи полиметаллических руд. Проведена апробация предлагаемых методов на территориях горнодобывающих предприятий Хабаровского края. Совместный анализ накопленных результатов многолетних наблюдений и актуальных данных о современном состоянии объектов природной среды позволяет прогнозировать развитие исследуемых природно-технических систем на среднесрочную перспективу.

Ключевые слова: информационная система, мониторинг, горнопромышленный комплекс, дистанционное зондирование, месторождение, экологическая безопасность, компьютерное зрение, машинное обучение.

1. Введение. Горнодобывающая промышленность оказывает комплексное воздействие на окружающую среду, затрагивая различные компоненты экосистемы: воздух, воду, землю, биоразнообразие и т.д. Эти воздействия имеют глобальный характер, но особенно заметны в странах с развитой горнодобывающей промышленностью. В процессе добычи и переработки полезных ископаемых в атмосферу выделяются различные вещества, такие как углекислый и сернистый газ, оксиды азота, частицы пыли и др. [1 – 4]. Горнодобывающая промышленность также играет важную роль в

загрязнении почвы и водоемов тяжелыми металлами [7 – 11]. В совокупности техногенные загрязнения приводят к разрушению растительного покрова и почти полному уничтожению биоты, превращая территории в «индустриальные пустыни» с крайне низкой биологической продуктивностью [5, 6]. Это делает рекультивацию нарушенных земель одной из приоритетных природоохранных задач, требующей значительных финансовых и трудовых ресурсов.

Помимо непосредственно горнодобывающей деятельности, экологический эффект оказывают и отходы, остающиеся после её завершения. При этом, их объёмы, измеряемые тысячами тонн, представляют определенную ценность и являются потенциальными техногенными месторождениями. Их повторное использование способствует значительной трансформации сформировавшихся в процессе первичной переработки ландшафтно-геохимических условий, обусловленной изъятием и перемещением больших масс горных пород и природных вод в пределах горного отвода.

Исходя из этого, высокую значимость имеет оценка состояния природно-технических систем и изучения участков, вовлеченных во вторичную переработку, и разработки рекомендаций по минимизации ущерба.

Традиционно горно-экологический мониторинг осуществляется специалистами в этой области непосредственно на горнопромышленных объектах. Удаленность некоторых месторождений и отсутствие транспортной инфраструктуры определяет их малую доступность для изучения экологической ситуации с участием квалифицированных специалистов и хорошей приборной базой. В связи с этим активно развиваются методы и технологии дистанционного мониторинга районов добычи полезных ископаемых, использующие спутниковые данные об изучаемых объектах. На основе этих данных проводится мониторинг экологического состояния территории месторождения, оцениваются масштабы проведения горнорудных работ, определяются места размещения и объёмы горнопромышленных отходов, оцениваются перспективы их повторной переработки, планируются мероприятия по предотвращению техногенных аварий.

Для решения вышеупомянутых проблем необходим эффективный мониторинг и методики, позволяющие понять масштабы проблемы и разработать необходимые стратегии по сохранению и восстановлению природных систем.

Современные технологии дистанционного зондирования Земли (ДЗЗ) позволяют получать мультиспектральные изображения

исследуемых территорий с высоким разрешением за длительный временной период. Это дает возможность в реальном масштабе времени следить за объектами горнопромышленного комплекса, существенно упрощая процесс мониторинга их влияния на близлежащие территории.

Дистанционный мониторинг горнопромышленных объектов сопряжен с необходимостью получения, обработки и анализа больших объемов данных об исследуемом объекте. Такую информацию трудно анализировать без применения современных средств высокопроизводительных вычислений.

Применение машинного обучения в ДЗЗ является важным направлением, позволяющим извлекать нужную информацию из спутниковых данных. Модели машинного обучения используются при решении различных задач дистанционного мониторинга. Их применение обеспечивает новые возможности обработки больших объемов спутниковых данных, создает условия для улучшения и автоматизации мониторинга окружающей среды и управления ресурсами. Они позволяют классифицировать объекты на спутниковых снимках, создавать модели для экстраполяции найденных закономерностей на большие территории, анализировать неструктурированные массивы информации и, тем самым, влиять на управление процессом мониторинга и планирования использования данных ДЗЗ. Создание информационной системы, которая в автоматизированном режиме будет выполнять все этапы получения и обработки массивов данных дистанционного зондирования может существенно улучшить оперативность и качество комплексного экологического мониторинга территорий проведения горнорудных работ.

Научная новизна работы заключается в комплексном подходе к экологическому мониторингу горнопромышленных территорий с использованием данных ДЗЗ и его автоматизации, реализуемом в виде разрабатываемой информационной системы. В системе реализованы: модели нейронных сетей и ансамблевых алгоритмов для решения задачи сегментации объектов горнопромышленного комплекса и классификации типов леса; алгоритм оценки влияния угольного пылевого загрязнения на растительность на основе временных рядов спутниковых данных и вегетационных индексов; расчёт объёмов отвалов по спутниковым данным с модификацией алгоритма построения диаграммы Вороного для более точного расчёта.

2. Состояние вопроса и методические подходы. Разработка технологических решений, позволяющих снизить техногенное

воздействие, является актуальной проблемой как в России, так и за рубежом. Интерес к техногенным образованиям возник достаточно давно и в настоящее время переживает период очередной активизации. Появляются многочисленные публикации, посвященные изучению строения, принципам типизации, методам оценки и способам отработки техногенных месторождений. Наиболее эффективно осваиваются техногенные месторождения, сформированные в процессе отработки собственно золотых месторождений. Из них добывается около 18% общего объема добычи золота в нашей стране [12, 13].

Проблема влияния объектов горнодобывающей промышленности на окружающую среду активно исследуется в современной научной литературе [14]. Отвалы представляют экологическую угрозу из-за содержания токсичных компонентов и тяжелых металлов [15]. Переработка отвалов может повысить экологическую безопасность и обеспечить дополнительные сырьевые ресурсы [16, 17].

Для переработки отвалов необходимо оценить их параметры. В обзоре [18] проанализированы современные методы оценки объемов отвалов. Традиционно оценка объемов отвалов проводится с помощью ручных измерений. Они являются трудоемкими, требуют значительных временных затрат и подвержены человеческим ошибкам, особенно при работе с большими или нерегулярными по форме отвалами. Спутниковые снимки часто не обеспечивают достаточной точности для сложных или динамически изменяющихся запасов, где форма и размер могут варьироваться, хотя и позволяют охватывать большие территории. Точнее оценить объемы отвалов можно с помощью БПЛА с фотограмметрическими системами и лидаров [19].

Проблема негативного влияния угольного пылевого загрязнения является объектом пристального внимания со стороны исследователей. Проведено множество работ по оценке этого влияния на различные компоненты окружающей среды. Часто встречающимся способом количественной оценки угольного загрязнения является анализ снежного покрова [20, 21]. Для качественной оценки проводится анализ самой пыли [22].

Также исследователи рассматривали вопрос использования методов ДЗЗ. Корреляция спектральных характеристик каналов спутниковых снимков с уровнем запыленности снега в местах отбора проб показала взаимосвязь между уровнем запыленности снега в точке отбора и его спектральными характеристиками на снимке [23 – 28].

Похожую взаимосвязь определили между спектральными характеристиками снежного покрова, выраженными в виде индексов снега, и его физико-химическими свойствами с данными о пылевой нагрузке в нём [29].

Помимо исследования корреляции, также решалась проблема определения области угольного пылевого загрязнения. Для этого разрабатывались соответствующие индексы: ICDI [30], BCI [31], ECDI [32] и др. В работе [33] предложен метод отслеживания распространения загрязнения угольной пылью, основанный на данных об оптической глубине аэрозолей (AOD) [34].

При анализе состояния земной поверхности и почвенного покрова, как правило, используются вегетационные индексы. Применение вегетационных индексов является традиционным способом оценки процессов самореабилитации месторождений [35]. Также они используются для экологического мониторинга угольных разрезов на основе спутниковых данных [36].

В зарубежной литературе часто встречаются работы с использованием моделей машинного обучения в исследованиях, так или иначе касающихся горной промышленности и землепользования [37, 38]. Их применение позволяет ускорить решение не сложных, но затратных по времени задач. В работах российских ученых также применяют нейронные сети, но чаще в задачах, непосредственно не связанных со спутниковыми данными ДЗЗ [39].

3. Структура информационной системы. Информационная система имеет модульную структуру: менеджер данных, модули обработки данных, анализа данных, а также программный интерфейс приложения и оболочка пользователя. Структурная схема информационной системы представлены на рисунке 1.

В качестве основных входных данных система использует спутниковые снимки Sentinel-2. Для расширения временного периода мониторинга также используются данные с семейства спутников Landsat, что позволяет выполнять более детальный анализ исследуемых объектов. Помимо спутниковых данных, в системе могут быть использованы данные полевых исследований и информация о состоянии исследуемой территории из открытых источников.

Спутниковые системы имеют различия в предоставляемых данных. Пространственное разрешение спутниковых снимков и спектральные профили каналов – то есть кривые зависимости интенсивности света от длины волны, отличаются. Кроме того, данные спутниковых систем различаются по пространственному

позиционированию, что может в последующем усложнить их анализ. Поэтому, в модуле обработки спутниковых данных проводится унификация спутниковых данных путем их приведения к одной системе пространственного позиционирования, к одному пространственному разрешению и проведения спектральной коррекции. В обработку также входит выделение объектов и классификация типов растительности.

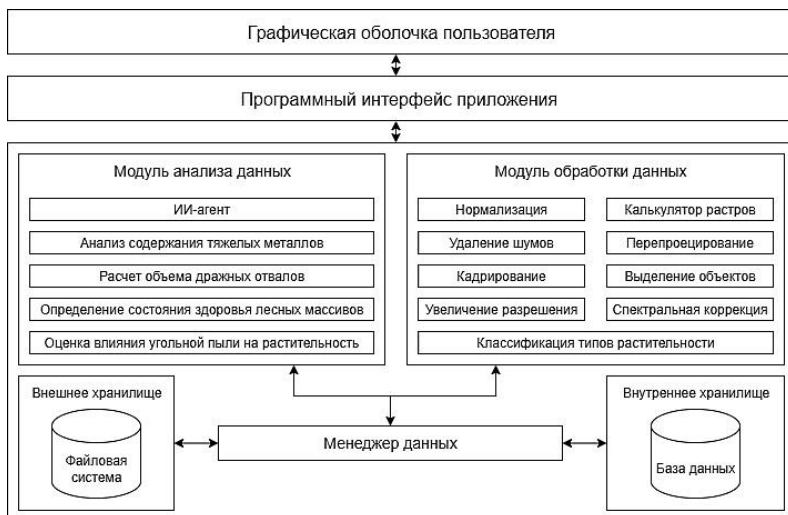


Рис. 1. Структура информационной системы

После обработки полученных данных модулем анализа проводится геоэкологический анализ с привлечением ИИ-агента.

Хранилище информации имеет три логических блока, хранящих информацию о необработанных данных, обработанных данных и объектах наблюдения. Исходной точкой базы данных служит блок необработанных спутниковых снимков. В нём хранится детальная информация о поступающих в систему первичных изображениях, полученных непосредственно от организаций-владельцев спутниковых систем. Данный блок содержит как метаданные изображений, такие как места съемки, даты съемки, системы координат, так и сведения о спутниках-источниках этих снимков, включая их спектральные и пространственные характеристики. Данные этого блока представляют собой первичный материал для дальнейшей обработки изображений Земли.

Информация о результатах первичной обработки изображений хранится в блоке обработанных спутниковых снимков. Он содержит метаданные предобработки изображений и ссылки на снимки, прошедшие различные этапы предварительной подготовки. К таким этапам относятся удаление шумов, нормализация, спектральная коррекция, перепроецирование, увеличение разрешения, а также возможное формирование композитов и другие. Для каждого обработанного снимка фиксируется использованный метод предобработки и устанавливается связь с исходным необработанным изображением, обеспечивая прослеживаемость данных.

Блок объектов наблюдения предназначен для хранения информации об областях интереса и о самих объектах, расположенных в их пределах. Области интереса представляют из себя географически локализованные прямоугольные участки земной поверхности, представляющие особый интерес для исследователей. Каждая область интереса определяется своими географическими координатами и привязывается к конкретным обработанным спутниковым снимкам, покрывающим данную территорию. Объектами выступают наземные сущности, являющиеся предметом изучения. Среди них могут быть как участки растительности и водные объекты, так и техногенные объекты или иные элементы. Каждый объект характеризуется своим типом, пространственными атрибутами и связан с определенной областью интереса.

Совокупность информации, хранимой в данных блоках, обеспечивает эффективное управление всем жизненным циклом спутниковых снимков в системе – от добавления первичной информации до извлечения нужных сведений об объектах на земной поверхности.

3.1. Модуль обработки данных. Для выделения объектов предлагается использование сверточных и трансформерных нейронных сетей, которые обучены на заранее размеченных объектах. Так как для нейронных сетей требуется достаточно большой объем данных, то в качестве примеров взяты объекты как Хабаровского края, так и других субъектов РФ.

Исходными данными послужили спутниковые снимки Sentinel-2 (таблица 1), так как они имеют высокое пространственное разрешение, и, соответственно, более детальную информацию о спектральных характеристиках поверхности. Это способствует более точному мониторингу экологического состояния территорий.

Таблица 1. Используемые каналы снимков Sentinel-2

Канал	Название	Длина волны, нм	Разрешение, м
B02	BLUE	458 – 523	10
B03	GREEN	543 – 578	
B04	RED	650 – 680	
B08	NIR	785 – 900	
B05	VRE	698 – 713	20
B06		733 – 748	
B07		773 – 793	
B11	SWIR1	1565 – 1655	
B12	SWIR2	2100 – 2280	

Использовались данные о местах проведения горных работ для следующих субъектов РФ: Амурская область, Забайкальский край, Хабаровский край, Иркутская область, Республика Саха (Якутия), Республика Хакасия, Кемеровская область.

Для каждого типа объектов было получено следующее количество снимков:

- Дrajные отвалы – 186 снимков;
- Угольное пылевое загрязнение – 52 снимка;
- Карьеры и хвостохранилища – 43 снимка.

Снимки проходили обработку, которая включает в себя нормализацию данных в интервал от 0 до 1, разметку объектов путём создания маски и разделение снимков и масок на отдельные части (тайлы).

Исходные спутниковые снимки Sentinel-2 имеют размер 10980x10980 пикселей. Так как искомые объекты занимают существенно меньшую площадь от всего снимка, то определяются прямоугольные области интереса, содержащие объекты. Исходные снимки обрезаются по этим областям.

После этого на полученных областях снимков проводится разметка искомым объектов. В случае карьеров и хвостохранилищ разметка проводилась вручную. Для разметки территорий с дражными отвалами использовалась ручная разметка с использованием индексов MNDWI и NDBSI:

$$MNDWI = \frac{GREEN - SWIR1}{GREEN + SWIR1},$$

$$NDBSI = \frac{(RED + SWIR1) - (NIR + BLUE)}{(RED + SWIR1) + (NIR + BLUE)}.$$

Угольное пылевое загрязнение размечалось с помощью индекса ECDI с последующей ручной доработкой. Формула выглядит следующим образом:

$$ECDI = \frac{SWIR1 - NIR + SWIR2}{SWIR1 + NIR - SWIR2}.$$

Обработанные снимки и их маски делятся на части заданного размера и распределяются по обучающим и валидационным выборкам. Части, не содержащие искомые объекты, не входят ни в одну из выборок. Для предотвращения утечки информации и обеспечения независимости выборок объекты одного типа не перемешивались между выборками. Например, если конкретный карьер был включен в валидационную выборку, то все данные, относящиеся к этому карьере, не входили в обучающую выборку. Таким образом, размеры полученных обучающих и валидационных выборок для каждого типа объектов составил: для дражных отвалов – 1355 и 461; для угольного пылевого загрязнения – 272 и 77; для карьеров и хвостохранилищ – 725 и 242 частей.

Для детектирования каждого типа объектов проводилось обучение нескольких моделей нейронных сетей: UNet [40], DeepLabV3 [41], Segformer [42], SWIN-UNet [43], градиентный бустинг [44], случайный лес [45].

UNet представляет собой полностью сверточную нейронную сеть с U-образной архитектурой, состоящую из энкодера для извлечения признаков и декодера с пропускными соединениями для восстановления пространственного разрешения. Её выбор обусловлен производительностью на малых выборках и высокой точностью определения границ объектов.

DeepLabV3 использует расширенные свертки и модуль Atrous Spatial Pyramid Pooling для учета мультимасштабного контекста, что важно для выделения объектов разного размера на спутниковых снимках. Также эта модель эффективна при работе с нечетко определенными границами объектов, что актуально для рассматриваемых в данной работе карьеров, отвалов, хвостохранилищ и угольного пылевого загрязнения.

Segformer объединяет преимущества трансформеров и сверточных сетей, используя иерархический трансформерный энкодер и многоуровневый перцептронный декодер. Механизм внимания позволяет учитывать зависимость между удаленными участками изображения. Это необходимо для правильной классификации протяженных горнопромышленных комплексов.

SWIN-UNet адаптирует архитектуру Swin Transformer для задач сегментации, используя сдвинутые окна внимания и иерархическую структуру признаков, что обеспечивает эффективное моделирование дальних зависимостей. Механизм сдвинутых окон позволяет обрабатывать полученные части малого размера без потери контекстуальной информации. Он позволяет выделять мелкие элементы дражных отвалов и лучше определять нечеткие границы искоемых объектов.

Градиентный бустинг и случайный лес представляют ансамблевые методы машинного обучения, работающие с извлеченными из изображений признаками и обеспечивающие высокую устойчивость к переобучению. Они тестировались для сегментации поверхности, покрытой частицами угля (угольный пылевой след), так как её спектральные характеристики имеют ярко-выраженные отличия в сравнении с другими типами поверхности.

Расчёт площади проводится исходя из результатов выделения границ. При рассмотрении влияния объектов горной промышленности на экологию близлежащих зон оценивается здоровье растительности на исследуемой территории.

По результатам работы моделей выделялись 4 варианта присвоения класса объекту, обозначаемые True Positive (TP), False Negative (FN), True Negative (TN), False Positive (FP). TP – это правильно предсказанный положительный результат, истинно-отрицательная TN – это правильно предсказанный отрицательный результат, ложноположительная FP – это отрицательный результат, предсказанный как положительный, и ложноотрицательная FN – это положительный результат, предсказанный как отрицательный.

Оценка точности выбранных моделей производилась по метрикам Precision, Recall, F1 и Intersection over Union (IoU).

Precision – это метрика, показывающая долю истинно-положительного результата относительно всех предсказанных положительных меток.

Recall – это метрика, которая показывает долю положительных результатов среди всех положительных меток.

F1-score – это метрика, составленная из Precision и Recall. Она может быть эффективна для оценки точности в задачах, где классы в обучающей выборке несбалансированные.

IoU – это метрика, часто применяемая в задаче сегментации. Она показывает перекрытие между предсказанием модели и маской.

Представленные метрики вычисляются по формулам:

$$Precision = \frac{TP}{TP + FP}, Recall = \frac{TP}{TP + FN},$$

$$F1 = \frac{TP}{TP + 0.5(FP + FN)}, IoU = \frac{TP}{TP + FP + FN}.$$

Метрика IoU считается основной в рамках поставленной задачи при оценке точности моделей, так как учитывает пространственное отношение между предсказанием модели и маской.

3.2. Модуль анализа данных. Для оценки объёма отходов, размещенных в дражных отвалах, использовались цифровые модели рельефа (DEM), которые представляют числовые значения высотных характеристик земной поверхности и подразделяются на цифровые модели местности (DTM) и цифровые модели поверхности (DSM). DTM является моделью, отображающей непосредственно значения высоты без учёта искусственных объектов и растительности. DSM включает в себя высоты всех объектов, расположенных на земной поверхности (здания, сооружения, растительность).

Для расчета объемов использовались цифровые модели поверхности ASTER, ALOS, Copernicus и цифровые модели местности FABDEM и GEDTM [46 – 51].

Чтобы снизить влияние растительности на результат расчета объема используются данные Global Forest Canopy Height. Это модель, описывающая высоту лесного покрова. Ограничением для её применения является разница в пространственном разрешении с моделями DEM и временная разница в данных.

Протяженность областей с дражными отвалами определяет то, что высота на разных участках может сильно меняться. Для снижения влияния амплитуды высоты на результат расчета, исследуемая область разделяется на сегменты с помощью алгоритма создания полигонов Вороного. Для разделения полигона на части используется центральная линия полигона. Ложные точки ветвления, появляющиеся в результате особенностей полигона, удаляются итеративно. На

очищенном от разрывов и ложных ветвей скелете, через расставленные на заранее заданном расстоянии точки проводятся прямые. Далее для каждого из полигонов с использованием инструментов зональной статистики и цифровых моделей рельефа определяются базовые линии, на основе которых рассчитывается объем.

Для оценки влияния угольного пылевого загрязнения на растительность за основу взят метод временного анализа снимков по вегетационным индексам. Для этого разработан алгоритм, состоящий из двух этапов. Входными данными этого алгоритма являются спутниковые снимки Sentinel-2 исследуемой территории в период таяния снега и в летний период времени за несколько лет.

Первый этап включает в себя определение зоны пылевого загрязнения вблизи области добычи и выполняется для каждого рассматриваемого года. Первым шагом данного этапа является определение области пылевого загрязнения на снимке в период таяния снега. Следующим шагом проводится расчет вегетационного индекса (в работе использовался SARVI) на спутниковом снимке в летний период. По полученным значениям проводится бинаризация по порогу меньше 0,2 для выделения области карьера. После этого, полученная область карьера удаляется из маски области пылевого загрязнения, тем самым получая зону пылевого загрязнения без карьера.

В рамках второго этапа проводится наложение зон пылевого загрязнения за каждый год. Таким образом получают сектора, в которых пылевое загрязнение замечено определенное количество раз. В соответствии с этим определяются названия этих секторов. Данное наложение проводится для каждого года с учетом предыдущих. Затем, в этих секторах проводится расчет вегетационных индексов и их средних значений. По полученным результатам проводится анализ и оценка влияния пылевого загрязнения на растительность вблизи мест добычи угля.

Для идентификации типов леса проводится расчет соответствующих карт с использованием модели нейронной сети DVT. Нейронная сеть DVT представляет собой полносвязанную нейронную сеть с линейными слоями, нормализацией пакетов, активацией ReLU и выходом softmax. Для её обучения использовались спутниковые снимки Sentinel-2 в период максимальной вегетации растений и период конца листопада до выпадения снега за 2019 год. Из снимков отбирались каналы Blue, Green, Red, NIR, SWIR1 и SWIR2. Также в обучении использовались вспомогательные данные, такие как карты Global Forest Canopy Height 2019 и вегетационный индекс SARVI.

SARVI – это спектральный индекс растительности, разработанный для измерения плотности и состояния зелёной растительности по спутниковым снимкам, который обладает устойчивостью к атмосферным и почвенным искажениям. Его расчет проводился по формулам:

$$SARVI = \frac{NIR - RED + \alpha(BLUE - RED)}{NIR + RED - \alpha(BLUE - RED) + L} \cdot (1 + L),$$

где L и α – коэффициенты, компенсирующие влияние отражательной способности почвы и атмосферных искажений. Их обычные значения: $L = 0.5$, $\alpha = 1$.

Оценка состояния растительности в зависимости от её типа проводилась с помощью экспоненциальной регрессии расчета LAI по вегетационному индексу SARVI по формуле:

$$FH_{VI} = FH(VI) = \frac{LAI}{LAI_{\max}} = \frac{a_{VI} \cdot \exp(VI \cdot b_{VI}) + c_{VI}}{LAI_{\max}},$$

где a , b и c – расчётные коэффициенты, зависящие от типа леса (таблица 2); VI – вегетационный индекс SARVI; $LAI_{\max}$ – это верхняя граница диапазона нормальных значений LAI (таблица 3).

Таблица 2. Коэффициенты формулы экспоненциальной регрессии расчета LAI

Типы леса	Коэффициент а	Коэффициент b	Коэффициент с
Хвойный	1.6998	0.1743	– 1.2638
Лиственный	1157.7789	0.0003	– 1157.4778
Смешанный	1720.8242	0.0002	– 1720.496

Таблица 3. Диапазоны нормальных значений LAI

Индекс	Нормальные значения	Тип леса		
		Хвойный	Лиственный	Смешанный
LAI	мин.	0.5086	0.434	0.4713
	макс.	0.6051	0.5175	0.5613

Для исследования содержания тяжелых металлов в почвогрунтах проводился отбор проб для последующего рентгенофлуоресцентного анализа и атомно-абсорбционной спектроскопии. Проводился анализ проб на содержание следующих металлов: Hg, Pb, Cd, As, Cr и Fe. К полученным данным применялись методы

статистического и пространственного анализа. Перед этим распределение набора данных (концентрации микроэлементов в почве) оценивалось методами «Box-plot», Колмогорова-Смирнова ($p < 0,2$) и Шапиро Уилка ($p < 0,05$).

Апробация методов проводилась на следующих территориях: для расчета объема дражных отвалов – р. Семитка вблизи села Бриакан (рисунок 2(1)); для идентификации типов леса и оценки его состояния – Кербинский разрез вблизи села Бриакан (рисунок 2(2)); для оценки влияния пылевого загрязнения на растительность – Правобережный разрез вблизи поселка Чегдомын (рисунок 2(3)); для анализа содержания тяжелых металлов в почве – Солнечный (рисунок 2(4)).

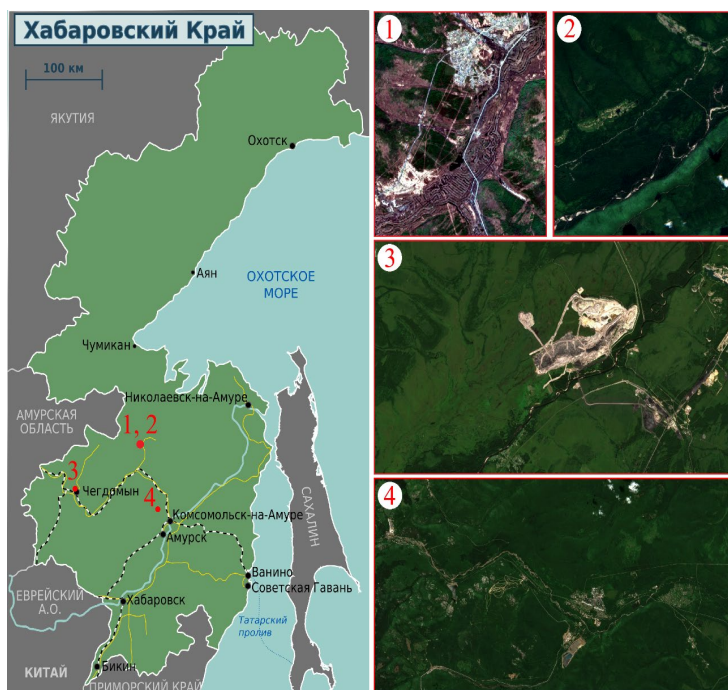


Рис. 2. Территории апробации

4. Результаты. Оценка работы моделей для сегментации территорий с дражными отвалами, угольного пылевого следа, карьеров и хвостохранилищ представлены в таблице 4.

Таблица 4. Оценки точности работы обученных моделей

	Модель	Precision	Recall	F1	IoU
Дражные отвалы	UNet	0.837	0.851	0.844	0.730
	DeepLabV3	0.852	0.889	0.873	0.774
	Segformer	0.784	0.893	0.835	0.717
Угольный пылевой след	UNet	0.943	0.885	0.902	0.837
	DeepLabV3	0.893	0.859	0.865	0.773
	Градиентный бустинг	0.857	0.944	0.898	0.815
	Случайный лес	0.805	0.952	0.873	0.774
Карьеры и хвостохранилища	UNet	0.743	0.758	0.735	0.770
	DeepLabV3	0.644	0.666	0.641	0.721
	Segformer	0.833	0.839	0.822	0.797

DeepLabV3 показала лучшие результаты для дражных отвалов и пылевого загрязнения, а Segformer – для карьеров и хвостохранилищ.

Расчет объема дражных отвалов проводился по снимку за 2019 г. Разработка золотороссыпных месторождений дражным и гидромеханизированным способами в районе им. П. Осипенко Хабаровского края ДФО началась еще в конце прошлого столетия. В связи с многолетним интенсивным освоением минерального сырья здесь накоплены большие объемы отходов золотодобычи, расположенных в верховьях рек Семитка и Бриакан, в границах с. Бриакан. Они загрязняют окружающую среду соединениями ртути и другими токсичными соединениями тяжелых металлов. Отходы обогащения являются потенциальным источником техногенного загрязнения биосферы и относятся ко 2 и 3 классам опасности [51].

На рисунке 3 представлен результат апробации алгоритма разделения полигона территории Бриакан на части для расчета объема.



Рис. 3. Алгоритм разделения полигона на части для расчета объема

В таблице 5 представлены рассчитанные значения объёма апробируемой территории.

Таблица 5. Объемы в границах территорий с отвалами (с. Бриакан), $\text{км}^3 * 10^{-2}$

Модели рельефа	С учетом высоты деревьев			Без учета высоты деревьев		
	Мин.	Сред.	Медиана	Мин.	Сред.	Медиана
ALOS	2.540	0.639	0.831	1.453	0.286	0.310
Copernicus	2.955	0.634	0.777	2.243	0.467	0.524
ASTER	5.528	1.016	1.067	5.002	0.872	0.869
FABDEM	0.837	0.193	0.210	-	-	-
GEDTM	1.505	0.271	0.283	-	-	-

Сравнительный анализ различных цифровых моделей рельефа продемонстрировал существенные расхождения в результатах расчета объемов. На участке вблизи села Бриакан без учета высоты растительности значения варьируются от $1.4 \text{ км}^3 * 10^{-2}$ (ALOS) до $5 \text{ км}^3 * 10^{-2}$ (ASTER). Такие значительные различия обусловлены особенностями самих моделей, сложностью рельефа исследуемых территорий и погрешностью выделенных границ.

Оценка влияния угольного пылевого загрязнения на состояние растительности проводилась в период с 2018 по 2023 гг. На рисунке 4 представлен пример области угольного пылевого загрязнения за 2022 г., маска секторов пылевого загрязнения, полученных за исследуемый период времени.

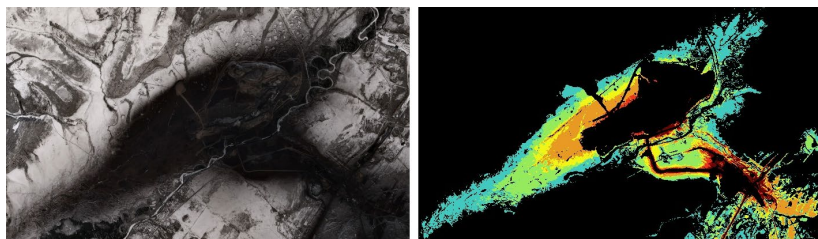


Рис. 4. Область угольного пылевого загрязнения за 2022 г.; полученная маска секторов пылевого загрязнения за период 2018-2023 гг.

На рисунке 5 изображен график средних значений вегетационного индекса SARVI. Из графика видно, что по мере приближения к области карьера заметно ухудшение состояния растительности в рамках одного года. При этом ухудшение оказывается незначительным, так как средние значения индекса меняются в пределах сотых долей. Если же рассматривать изменение

значений за несколько лет, то тенденции к ухудшению не наблюдается. Это можно объяснить тем, что съемка со спутников производится в разных условиях, что может генерировать неконтролируемые, сложно фиксируемые погрешности.

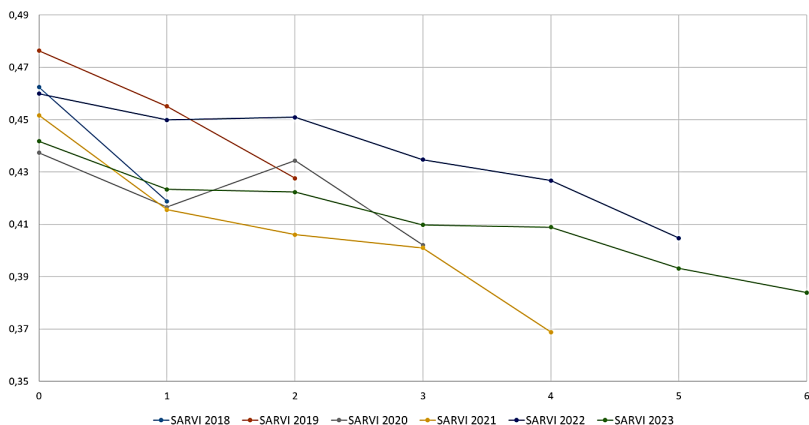


Рис. 5. График рассчитанных средних значений вегетационного индекса SARVI по секторам

Идентификация типов леса и оценка их здоровья проводилась по данным за 2022 г. На рисунке 6 представлена построенная для исследуемой территории карта типов леса, а также карта состояния их здоровья.

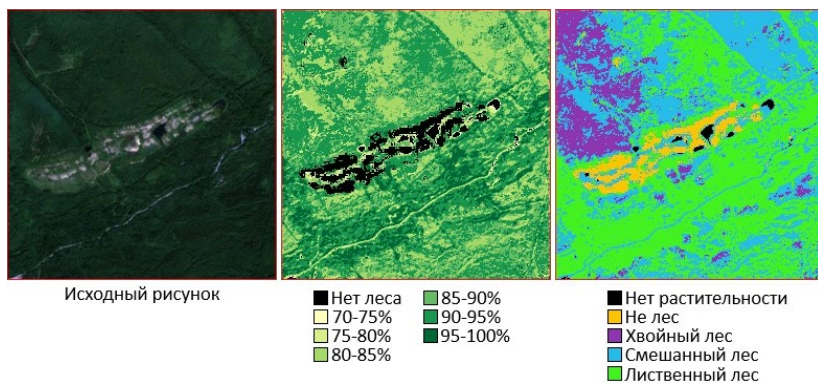


Рис. 6. Построенные карты типов леса и их состояния здоровья

На их основе можно провести анализ изменения на исследуемой территории, а именно анализ изменений площадей: не покрытых растительностью, покрытых кустарниками и травой и покрытых деревьями. Также по картам состояния лесных массивов можно визуальнo оценить влияют ли исследуемые объекты горнопромышленного комплекса на прилегающую лесную растительность, что особенно актуально для хвостохранилищ.

Отбор проб для анализа на содержание тяжелых металлов проводился на территории Солнечного горно-обогачительного комбината (ГОК) – предприятия по добыче и обогащению руд цветных металлов. Расположен данный ГОК в Комсомольском оловорудном районе. Ведется добыча минерального сырья открытым и подземным способом. Климат района муссонно-континентальный, преимущественное направление ветра восточное и западное (вдоль долины). Сочетание климатических особенностей и горно-долинного рельефа создает условия для активного воздушного загрязнения и вторичного, за счет вымывания осадками. Установлено, что в основном процессы миграции происходят в весенне-летний период, во время паводков, сильных дождей и ветров, которые имеют тенденцию к усилению с учетом направления долины р. Силинки [52]. Общая площадь района исследования составила 27394 га. В системе ГОКа функционировали 4 рудника, 2 обогачительные фабрики и 5 хвостохранилищ. Специфика добычи и обогащения оловянной руды на этой территории заключалась в извлечении и переработке огромного объема горной массы, часть которой использовалась, а оставшаяся накапливалась в виде техногенных отходов, размещенных в хвостохранилищах. В районе исследования кроме горнопромышленного сектора, источником поступления химических элементов в окружающую среду являются два населенных пункта общей численностью 13387 человек и автомобильные дороги. Всего было отобрано 10 проб за 2015 и 2020 гг. в местах, отмеченных желтыми точками на рисунке 7.

Описание исходных данных, проб микроэлементов в почве, которые отбирались в 2015, 2020 гг. на десяти площадках (полигонов исследования), расположенных в районе п.г.т. Солнечный приведены в таблице 6.

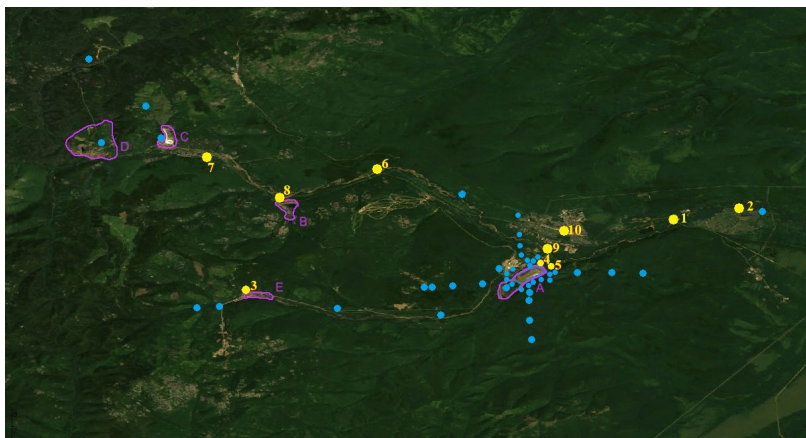


Рис. 7. Карта-схема района исследования Солнечного ГОК

Таблица 6. Описательная статистика содержания микроэлементов на исследуемой территории

2015 год							
Элементы, мг/кг	Mean	Median	CV	SD	SE	Skewness	Kurtosis
Mn	542,57	506,83	36,93	200,38	63,37	0,23	-0,85
Fe	26059,9	23258	41,01	10687,42	3379,66	1,01	0,52
Cu	747,67	575	109,26	816,87	258,32	2,32	6,03
Zn	360,3	321,67	53	190,95	60,38	0,97	0,19
As	1064	490,17	99,71	1060,91	335,49	1,29	0,6
Cd	5,12	4,65	50,51	2,59	0,82	2,45	6,78
Sn	388,57	315,83	70,92	275,56	87,14	0,2	-1,73
Hg	54,59	48	62,13	33,92	10,73	4,38	4,38
Pb	64,2	18,95	192,77	123,76	39,14	3,05	9,46
2020 год							
Mn	661,13	576,67	69,85	461,83	146,04	1,93	4,03
Fe	25765,2	25496,5	35,98	9269,74	2931,35	0,31	-1,35
Cu	585,83	544	316,23	549,36	173,72	0,77	0,32
Zn	749,6	559	91,09	682,82	215,93	0,87	-0,58
As	1220,7	494,5	123,27	1504,79	475,86	1,42	0,56
Cd	2,76	0	135,99	3,75	1,19	0,78	-1,49
Sn	497,77	206	142,31	708,37	224,01	1,7	2,28
Pb	515,47	581,8	72,2	372,17	117,69	-1,38	-1,38

Примечание: Mean – среднее арифметическое; Median – медиана; Min – минимальное значение; Max – максимальное значение; CV (coefficient of variation) – коэффициент вариации, %; SD (standard deviation) – стандартное отклонение; SE (standard error of the mean) – стандартная ошибка среднего значения; Skewness – асимметрия; Kurtosis – эксцесс.

В таблице 7 приведены значения коэффициента вариации содержания микроэлементов в почве в 2015 г. (n=10), 2020 г. (n=10) и в выборке, объединяющей данные за 2015, 2020 гг. (n=20). Коэффициент вариации CV изменяется от 38% до 122% в объединенной выборке, от 37% до 193% в выборке 2025 г. и от 36% до 142%. Данные, приведенные в таблице 7 показывают, что распределение концентраций всех микроэлементов на территории исследования является достаточно неоднородными.

Таблица 7. Анализ неоднородности содержания микроэлементов в почве района вблизи п.г.т. Солнечный

№	Микро-элементы	CV по микроэлементам за 2015, 2020 г. %	CV по микроэлементам за 2015 г. %	CV по микроэлементам за 2020 г. %
1	Mn	58.45	36.93	69.85
2	Fe	37.58	41.01	35.98
3	Cu	102.38	109.26	93.77
4	Zn	95.01	53.00	91.09
5	As	111.15	99.71	123.27
6	Cd	85.34	50.51	135.99
7	Sn	118.72	70.92	142.31
8	Pb	122.69	192.77	72.20

В природе содержание элементов подчиняется нормальному распределению при отсутствии других антропогенных источников поступления загрязняющих веществ [53]. Критериев, которые могут быть использованы для проверки гипотезы о принадлежности выборки нормальному закону, достаточно много. В связи с тем, что выборки концентрации микроэлементов имеют малую длину (n=10), исследование на соответствие нормальному закону проводилось для выборки, значения которой объединяли наблюдения 2015 и 2020 гг. (n=20)). Использовались следующие критерии: совместный критерий проверки на симметричность и нулевой эксцесс, критерий Колмогорова-Смирнова, критерий Шапиро-Уилка, соответствия прямой нормального распределения, анализ коробочных диаграмм «Box-plot» (таблица 8).

Таблица 8. Соответствие данных выборки концентрации микроэлементов за 2015, 2020 гг. нормальному закону распределения

Критерии соответствия нормальному закону распределения					
№	Микроэлементы	Колмогорова-Смирнова	Шапиро-Уилка	Соответствие прямой нормальной распределения	Анализ коробочных диаграмм «Box-plot»
1	As	-	-	-	-
2	Cd	+	+	+	-
3	Cu	+	-	-	-
4	Fe	+	+	+	+
5	Mn	+	-	-	+
6	Pb	-	-	-	-
7	Sn	+	-	-	-
8	Zn	+	-	-	-

Результаты теста Колмогорова-Смирнова, предположительно, показывают ложное соответствие нормальному закону распределения, это может быть связано с малой длиной выборки. Для выборок малой длины рекомендовано ориентироваться на визуализацию соответствия данных прямой нормальной распределения [52].

На основании проведенных исследований, можно сделать вывод о том, что гипотеза о соответствии нормальному закону распределения не отвергается для выборки концентраций Fe и Cd (рисунок 8).

Принято считать, что элементы, имеющие природное происхождение, обычно обладают низкими CV, а элементы, связанные с антропогенными источниками, характеризуются высокими CV и отражают неоднородное распределение концентраций [54]. Анализируя значения коэффициентов вариации в трех выборках, можно предположить, что все они сильно неоднородны, так как CV изменяется от 35 до 192 %, это очень значительная вариация. Вероятно, микроэлементы As, Cu, Mn, Pb, Sn, Zn имеют природное происхождение. Исходя из, этого, можно сказать, что на концентрации Fe и Cd могут влиять как эндогенные, так и экзогенные факторы.

В процессе исследования влияния объектов загрязнения на изменения концентрации микроэлементов, содержащихся в верхнем слое почвы, проводился корреляционный анализ, на основании которого можно определить степень влияния объектов загрязнения на изменение концентрации микроэлементов площадки в зависимости от удаленности площадки исследования от объектов.

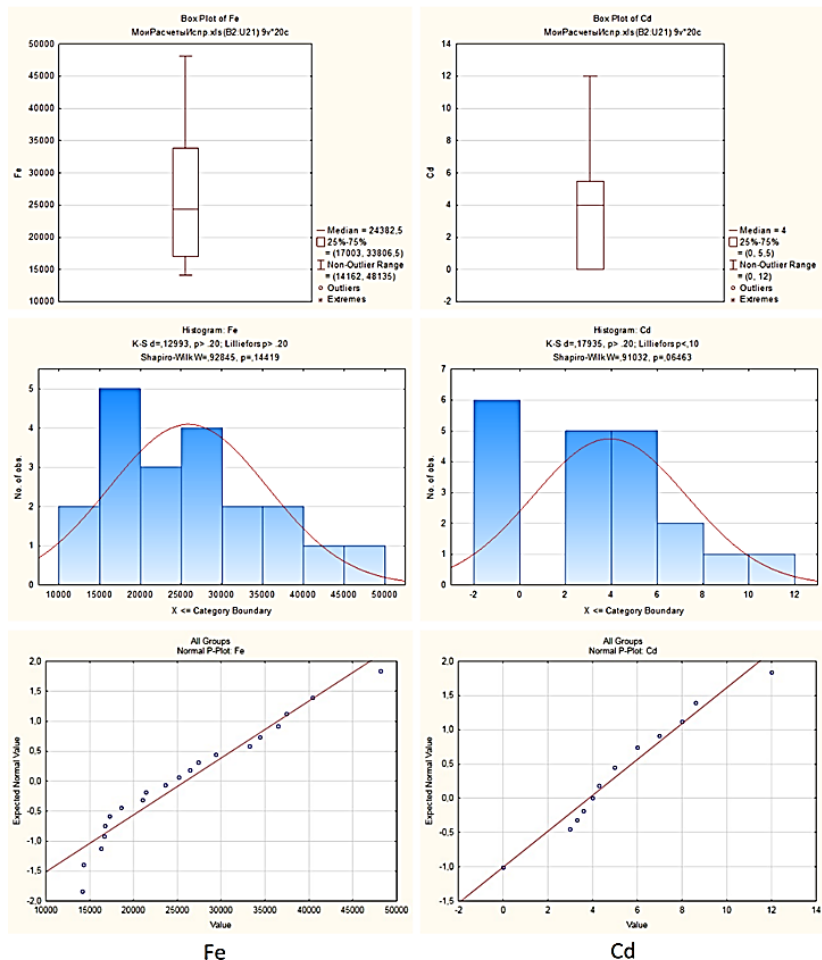


Рис. 8. Диаграмма и гистограмма распределения микроэлементов, диаграмма рассеивания для Fe и Cd

Для составления матрицы корреляции рассчитывалась разность содержания металлов в 2020 и 2015 году по каждой площадке, так как предположительно их концентрация изменялась из-за влияния объектов загрязнения, также рассчитывалась расстояние от каждого загрязняющего объекта до каждой площадки. После этого формировалась матрица корреляции (рисунок 9). Предполагается, что отрицательная корреляция при положительной динамике содержания

самого металла будет свидетельствовать о том, что чем меньше расстояние до объекта, тем больше загрязнение площадки, и наоборот, чем больше расстояние до объекта, тем меньше загрязнение.

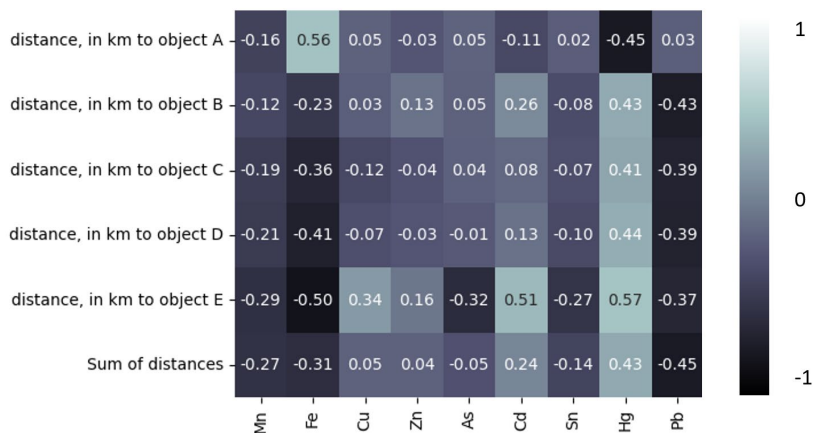


Рис. 9. Матрица корреляций

Наблюдается высокая обратная корреляция содержания свинца (Pb) с расстоянием до всех объектов загрязнения, кроме объекта А, что вместе с положительной динамикой роста концентрации свинца свидетельствует о том, что чем ближе к объектам загрязнения находится площадка, тем выше степень загрязнения свинцом. Отметим, что наибольшее влияние на изменение концентрации свинца оказывает объект загрязнения В. Ртуть из рассмотрения исключена, из-за отсутствия наблюдений за 2020 год. Обратная зависимость, отрицательная корреляция выявлена для железа (Fe), но динамика его изменения отрицательна.

На основании анализа процентного содержания микроэлементов (таблица 9) сделан вывод, что концентрация свинца изменилось приблизительно в семь раз, содержание цинка изменилось в два раза, содержание кадмия изменилось в два раза, содержание олова увеличилось на 30%.

Таблица 9. Изменение содержание микроэлементов в 2020 году относительно 2015 года

Микроэлементы	Mn	Fe	Cu	Zn	As	Cd	Sn	Pb
Изменение, %	21.85	-1.13	-21.65	108.05	14.73	-46.09	28.1	702.91

На рисунке 10 показано изменение концентрации свинца по области интереса. Видно, что наибольшая концентрация свинца наблюдается вблизи объектов С, Е и В, что в совокупности с увлечением концентрации свинца между 2015 и 2020 годами и отрицательной корреляции может свидетельствовать о наличии загрязняющего фактора данных объектов.

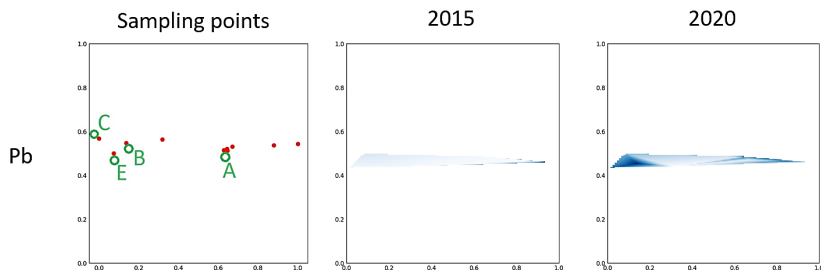


Рис. 10. Визуализация изменения концентрации свинца в 2015, 2020 г. по области интереса

В процессе проведенного исследования выявлено, что основными антропогенными источниками поступления Mn, Fe, Cu, Zn, As, Cd, Sn, Hg, Pb в почвогрунты стали хвостохранилища, в которых хранятся отходы горного производства. Отобранные в зоне воздействия источников пробы почвогрунтов проанализированы с помощью методов статистического анализа. Первичный анализ распределения микроэлементов по асимметрии и эксцессу показал, что для микроэлементов Mn, Cu, Sn гипотеза о нормальности распределения концентрации не отвергается, а распределение концентрации Fe, Zn, As, Cd, Pb отличается от нормального. На основании проведенных исследований, можно сделать вывод о том, что гипотеза о соответствии нормальному закону распределения не отвергается для выборки концентраций Fe и Cd. Анализ коэффициентов вариации свидетельствует о том, что на содержание Ni и Zn могут влиять как эндогенные, так и экзогенные факторы. Наблюдается высокая обратная корреляция содержания свинца с расстоянием до всех объектов загрязнения, кроме объекта А. Анализ процентного содержания микроэлементов показал, что концентрация Pb изменилась в семь раз, Zn и Cd в два раза, Sn увеличилось на 30%. Полученные данные свидетельствуют о том, что происходит миграция исследуемых элементов в объекты окружающей среды. Также установлено наличие процессов гипергенеза с образованием новых соединений, нетипичных для данного района.

5. Заключение. В результате проведенного исследования разработан комплексный подход к организации дистанционного горно-экологического мониторинга объектов минерально-сырьевого комплекса Хабаровского края, основанный на использовании данных дистанционного зондирования земли и анализе результатов полевых исследований.

Разработанные подходы с применением моделей машинного обучения позволили создать инструмент, частично автоматизирующий решения задач мониторинга горнопромышленных территорий.

Модифицированный алгоритм построения диаграммы Вороного позволил эффективно сегментировать протяжённые участки с дражными отвалами, что повысило точность расчётов при использовании различных базовых линий (минимальной, средней и медианной высоты). Выявлено, что на точность оценки объёмов существенно влияют такие факторы, как зарастание отвалов растительностью, эрозионные процессы и аккумуляция наносов, что особенно заметно на старых участках разработки.

При использовании разных моделей рельефа оптимальными являются FABDEM и GEDTM, так как они изначально не включают данные о высоте растительности, зданий и других объектов. Для улучшения результатов необходимо совместное использование данных ДЗЗ и беспилотных летательных аппаратов.

В результате апробации разработанных методов на территориях горнодобывающих предприятий Хабаровского края установлено, что пылевое загрязнение от угольных разрезов оказывает локальное, но незначительное влияние на состояние растительности. Изменения значений вегетационных индексов находятся в пределах сотых долей, при этом сами значения уменьшаются. Для более детального анализа изменений растительности планируется проведение дополнительных полевых исследований.

Анализ содержания тяжелых металлов в почвогрунтах района Солнечного ГОК показал существенное увеличение концентрации свинца, цинка, кадмия и олова за пятилетний период, что может свидетельствовать об активной миграции загрязняющих веществ от объектов складирования и разработки месторождений полезных ископаемых.

Разработанная модель нейронной сети DVT позволяет классифицировать типы леса. На основе результатов классификации и расчете листового индекса LAI оценивается состояние лесов, что обеспечивает возможность выявления зон экологического неблагополучия на прилегающих к горным предприятиям территориях.

Основным недостатком предложенных методов является необходимость заверки полученных результатов в естественных условиях. Авторы представленных материалов планируют решить этот вопрос в процессе предстоящих полевых сезонов.

Перспективы дальнейших исследований связаны с расширением спектра анализируемых параметров окружающей среды, интеграцией данных различных систем дистанционного зондирования, совершенствованием алгоритмов машинного обучения для повышения точности детектирования и классификации объектов, а также разработкой прогностических моделей развития природно-технических систем горнодобывающих регионов на среднесрочную и долгосрочную перспективу.

Работа посвящена памяти ушедшего из жизни Рафаэля Мидхатовича Юсупова – выдающегося ученого в области информатики, информационных технологий и теории управления, основателя и руководителя научных школ по теоретическим основам информатизации общества и по теории чувствительности сложных информационно-управляющих систем, доктора технических наук, профессора, член-корреспондента РАН, заслуженного деятеля науки и техники Российской Федерации, руководителя научного направления СПИИРАН Санкт-Петербургского Федерального исследовательского центра Российской академии наук, директора СПИИРАН (1991-2018 гг.), главного редактора журнала «Информатика и автоматизация».

Литература

1. Wang Z., Zhou W., Jiskani I.M., Ding X., Luo H. Dust pollution in cold region Surface Mines and its prevention and control // *Environmental Pollution*. 2022. vol. 292. DOI: 10.1016/j.envpol.2021.118293.
2. Wang H., Maqbool A., Xiao X., Yang H., Bi W., Bian Z. Seasonal pollution and risk assessment of heavy metals in atmospheric dust from coal mining area // *International Journal of Environmental Science and Technology*. 2022. vol. 19. № 12. pp. 11963–11972.
3. Блиновская Я.Ю., Мазлова Е.А. Выбросы парниковых газов при добыче и переработке угля: состояние проблемы и технологии сокращения // *Ученые записки Российского государственного гидрометеорологического университета*. 2019. № 54. С. 145–154.
4. Мун С.А., Ларин С.А., Глушков А.Н. Влияние роста добычи угля на загрязнение атмосферы и заболеваемость раком легкого в Кемеровской области // *Современные проблемы науки и образования*. 2013. № 1.
5. Афанасов В.И., Вокин В.Н., Заяц В.В., Зеньков И.В., Кирюшина Е.В., Коидрашов П.М., Конов В.Н., Логинова Е.В., Маглинец Ю.А., Нефедов Б.Н., Перфильев Д.А., Раевич К.В., Спирин Т.С., Федоров А.Б., Юронен Ю.П. Угольные разрезы России из космоса. Горные работы и экология нарушенных земель. Красноярск: Сибирский федеральный университет, 2017. 519 с.
6. Liu Y., Heng W., Yue H. Quantifying the coal mining impact on the ecological environment of Gobi open-pit mines // *Science of the total environment*. 2023. vol. 883.

7. Pan Y., Chen M., Wang X., Chen Y. Ecological risk, source apportionment, and influencing factors of heavy metals in soil in a typical lead-zinc mining watershed, Guangxi, China // *Journal of Environmental Chemical Engineering*. 2024. vol. 12. no. 3.
8. Baieta R., Ettler V., Vaněk A., Drahotka P., Křibek B., Nyambe I., Mihaljevič M. Smelter-derived soil contamination in Luanshya, Zambia // *Science of The Total Environment*. 2023. vol. 867.
9. Karn R., Ojha N., Abbas S., Bhugra S. A review on heavy metal contamination at mining sites and remedial techniques // *IOP Conference Series: Earth and Environmental Science*. IOP Publishing, 2021. vol. 796. № 1.
10. Haghizadeh A., Rajabi O., Nezarat A., Hajyani Z., Mina Haghmohammadi M., Hedayatikhah S., Asl S.D., Beni A.A. Comprehensive analysis of heavy metal soil contamination in mining Environments: Impacts, monitoring Techniques, and remediation strategies // *Arabian Journal of Chemistry*. 2024. vol. 17. no. 6.
11. Bakshi S., Banik C., ZhenLi H. The impact of heavy metal contamination on soil health. Managing soil health for sustainable agriculture. Volume 2: Monitoring and management. 2018. pp. 63–95.
12. Шпирт М.Я., Ласкорин Б.Н. Безотходная технология: утилизация отходов добычи и переработки твердых горючих ископаемых. Недра. 1986. 255 с.
13. Леонов С.Б., Федотов К.В., Сенченко А.Е. Промышленная добыча золота из золошлаковых отвалов тепловых электростанций // *Горный журнал*. 1998. Т. 5. С. 67–68.
14. Лупян Е.А., Константинова А.М., Балашов И.В., Кашницкий А.В., Саворский В.П., Панова О.Ю. Разработка системы анализа состояния окружающей среды в зонах расположения крупных промышленных объектов, хвостохранилищ и отвалов // *Современные проблемы дистанционного зондирования Земли из космоса*. 2020. Т. 17. № 7. С. 243–261.
15. Qdais H.A., Begday I.V., Katorgin I.Y., Shkarlet K.Y., Kharin K.V., Bluzhina A.S., Likhovid A.A. Assessment of metals pollution from tailing sites in the North Caucasus Region, Russia // *Mine Water and the Environment*. 2018. vol. 37. no. 4. pp. 815–824.
16. Шаповалов В.В., Козырь Д.А. Ресурсосберегающая технология утилизации породных отвалов горнодобывающих производств // *Известия Томского политехнического университета. Инжиниринг георесурсов*. 2023. Т. 334. № 4. С. 175–184.
17. Чуркин О.Е., Гилярова А.А. Освоение отходов горного производства как инвестиционное направление развития горнорудной промышленности Кольского полуострова // *Экономика, предпринимательство и право*. 2020. Т. 10. № 3. С. 905–916.
18. Alsayed A., Nabawy M.R.A. Stockpile volume estimation in open and confined environments: a review // *Drones*. 2023. vol. 7. no. 8.
19. Ren H., Zhao Y., Xiao W., Hu Z. A review of UAV monitoring in mining areas: Current status and future perspectives // *International journal of coal science & technology*. 2019. vol. 6. pp. 320–333.
20. Таловская А.В., Языков Е.Г., Шахова Т.С., Филимоненко Е.А. Оценка аэротехногенного загрязнения в окрестностях угольных и нефтяных котельных по состоянию снегового покрова (на примере Томской области) // *Известия Томского политехнического университета. Инжиниринг георесурсов*. 2016. Т. 327. № 10. С. 116–130.
21. Казанцева У.Д., Яковенко О.С., Лешукова М.К., Легощин К.В., Лешуков Т.В., Ларионов А.В. Морфологические характеристики, размер и масса пыли PM_{0,1} около угольных карьеров // *Успехи современного естествознания*. 2023. № 2. С. 90–95.

22. Kong S., Lu B., Ji Y., Zhao X., Chen L., Li Z., Han B., Bai Z. Levels, risk assessment and sources of PM10 fraction heavy metals in four types dust from a coal-based city // *Microchemical Journal*. 2011. vol. 98. no. 2. pp. 280–290.
23. Опарин В.Н., Потапов В.П., Гиниятуллина О.Л., Андреева Н.В., Счастливцев Е.Л., Быков А.А. Оценка пылевого загрязнения атмосферы угледобывающих районов Кузбасса в зимний период по данным дистанционного зондирования Земли // *Физико-технические проблемы разработки полезных ископаемых*. 2014. № 3. С. 126–137.
24. Стручкова Г.П., Крупнова Т.Г., Тихонова С.А., Капитонова Т.А. Исследование загрязнения снежного покрова угледобывающих районов с использованием спектральных характеристик // *Горный информационно-аналитический бюллетень (научно-технический журнал)*. 2021. № 12-1. С. 195–203.
25. Krupnova T.G., Rakova O.V., Struchkova G.P., Tikhonova S.A., Kapitonova T.A., Gavrilkina S.V., Bulanova A.V., Yakimova O.N. Insights into particle-bound metal(oids) in winter snow cover: Geochemical monitoring of the Korkinsky coal mine area, south Ural region, Russia // *Sustainability*. 2021. vol. 13. no. 9.
26. Chen W., Wang X., Cui J., Cao X., Pu W., Zheng X., Ran H., Ding J. Radiative forcing of black carbon in seasonal snow of wintertime based on remote sensing over Xinjiang, China // *Atmospheric Environment*. 2021. vol. 247.
27. Ahmad M., Alam K., Tariq S., Blaschke T. Contrasting changes in snow cover and its sensitivity to aerosol optical properties in Hindukush-Karakoram-Himalaya region // *Science of the Total Environment*. 2020. vol. 699.
28. Sahu, R., Gupta R.D. Snow cover area analysis and its relation with climate variability in Chandra basin, Western Himalaya, during 2001–2017 using MODIS and ERA5 data // *Environmental Monitoring and Assessment*. 2020. vol. 192(8). DOI: 10.1007/s10661-020-08442-8.
29. Тигеев А.А., Аксенов Н.В., Московченко Д.В., Пожитков Р.Ю. Оценка пылевого загрязнения атмосферы наземными и дистанционными методами (на примере г. Тобольск) // *Географический вестник*. 2021. № 2(57). С. 121–134.
30. Nie X., Hu Z., Ruan M., Zhu Q., Sun H. Remote-sensing evaluation and temporal and spatial change detection of ecological environment quality in coal-mining areas // *Remote Sensing*. 2022. vol. 14. no. 2.
31. Li Z., Zhao Y., Ren H., Sun Y. A Novel Index for Detecting Bare Coal in Open-Pit Mining Areas Based on Landsat Imagery // *Remote Sensing*. 2024. vol. 16. no. 24.
32. Xia N., Hai W., Song G., Tang M. Identification and monitoring of coal dust pollution in Wucaiwan mining area, Xinjiang (China) using Landsat derived enhanced coal dust index // *Plos one*. 2022. vol. 17. no. 4.
33. Yu H., Zahidi I. Environmental hazards posed by mine dust, and monitoring method of mine dust pollution using remote sensing technologies: An overview // *Science of the total environment*. 2023. vol. 864.
34. Wei X., Chang N.B., Bai K., Gao W. Satellite remote sensing of aerosol optical depth: advances, challenges, and perspectives // *Critical Reviews in Environmental Science and Technology*. 2020. vol. 50. no. 16. pp. 1640–1725.
35. Антонинова Н.Ю., Рыбникова Л.С., Славиковская Ю.О., Шубина Л.А. Эколого-экономические аспекты выбора направлений реабилитации территорий размещения промышленных отходов горно-металлургического комплекса // *Горная промышленность*. 2022. № S1. С. 71–77.
36. Мусина Г.А. Сравнительный анализ современных методов деформационного мониторинга // *Интерэкспо ГЕО-Сибирь*. 2023. Т. 4. № 2. С. 82–90.
37. Багаев С.М., Медведева Е.В. Сегментация спутниковых снимков на основе сверточной нейронной сети u-net. Цифровая обработка сигналов и ее применение (DSPA-2021): докл. 23-й междунар. конф. 2021. С. 218–222.

38. Kumar A., Gorai A.K. Development of a deep convolutional neural network model for detection and delineation of coal mining regions // *Earth Science Informatics*. 2023. vol. 16. no. 2. pp. 1151–1171.
39. Gallwey J., Robiati C., Coggan J., Vogt D., Eyre M. A Sentinel-2 based multispectral convolutional neural network for detecting artisanal small-scale mining in Ghana: Applying deep learning to shallow mining // *Remote Sensing of Environment*. 2020. vol. 248.
40. Ronneberger O., Fischer P., Brox T. U-net: Convolutional networks for biomedical image segmentation // *Proceedings of the 18th international conference Medical image computing and computer-assisted intervention (MICCAI)*, part III. 2015. pp. 234–241.
41. Chen L.C., Papandreou G., Florian S., Hartwig A. Rethinking atrous convolution for semantic image segmentation // *arXiv preprint arXiv:1706.05587*. 2017.
42. Xie E., Wang W., Yu Z., Anandkumar A., Alvarez J.M., Luo P. SegFormer: Simple and efficient design for semantic segmentation with transformers // *Advances in neural information processing systems*. 2021. vol. 34. pp. 12077–12090.
43. Cao H., Wang Y., Chen J., Jiang D., Zhang X., Tian Q., Wang M. Swin-unet: Unet-like pure transformer for medical image segmentation // *European conference on computer vision*. Cham: Springer Nature Switzerland. 2022. pp. 205–218.
44. Bentjac C., Csörgő A., Martínez-Muñoz G. A comparative analysis of gradient boosting algorithms // *Artificial Intelligence Review*. 2021. vol. 54(3). pp. 1937–1967.
45. Biau G., Scornet E. A random forest guided tour // *Test*. 2016. vol. 25. no. 2. pp. 197–227.
46. Khasanov K. Evaluation of ASTER DEM and SRTM DEM data for determining the area and volume of the water reservoir // *IOP Conference Series: Materials Science and Engineering*. IOP Publishing, 2020. vol. 883. no. 1.
47. Tadono T., Ishida H., Oda F., Naito S., Minakawa K., Iwamoto H. Precise global DEM generation by ALOS PRISM // *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*. 2014. vol. 2. pp. 71–76.
48. Li H., Zhao J., Yan B., Yue L., Wang L. Global DEMs vary from one to another: an evaluation of newly released Copernicus, NASA and AW3D30 DEM on selected terrains of China using ICESat-2 altimetry data // *International Journal of Digital Earth*. 2022. vol. 15. no. 1. pp. 1149–1168.
49. Marsh C.B., Harber P., Pomeroy J.W. Validation of FABDEM, a global bare-earth elevation model, against UAV-lidar derived elevation in a complex forested mountain catchment // *Environmental Research Communications*. 2023. vol. 5. no. 3.
50. Ho Y.F., Grohmann C.H., Lindsay J., Reuter I.H., Parente L., Witjes M., Hengl T. GEDTM30: Global Ensemble Digital Terrain model at 30 m and derived multiscale terrain variables. *PeerJ*. 2025. vol. 13. DOI: 10.7717/peerj.19673.
51. Филатова М.Ю. Оценка экологической ситуации в границах влияния золотодобычи (для обоснования создания лесных плантаций) // *Горный информационно-аналитический бюллетень (научно-технический журнал)*. 2023. № 8. С. 27–44.
52. Лемешко Б.Ю. Критерии проверки отклонения распределения от нормального закона. Руководство по применению. 2014. 192 с.
53. Zhao Y., Wang Z., Sun W., Huang B., Shi X., Ji J. Spatial interrelations and multi-scale sources of soil heavy metal variability in a typical urban–rural transition area in Yangtze River Delta region of China // *Geoderma*. 2010. vol. 156. no. 3–4. pp. 216–227.
54. Yuan G.L., Sun T.H., Han P., Li J., Lang X.X. Source identification and ecological risk assessment of heavy metals in topsoil using environmental geochemical mapping: typical urban renewal area in Beijing, China // *Journal of geochemical exploration*. 2014. vol. 136. pp. 40–47.

Смагин Сергей Иванович — д-р физ.-мат. наук, профессор, академик РАН, главный научный сотрудник, ВЦ ДВО РАН. Область научных интересов: математическое моделирование, информационные и вычислительные технологии. Число научных публикаций — 200. smagin@ccfebras.ru; улица Ким Ю Чена, 65, 680000, Хабаровск, Россия; р.т.: +7(421)222-7267.

Озарян Юлия Александровна — канд. техн. наук, ведущий научный сотрудник, ИГД ДВО РАН. Область научных интересов: геоэкология, геоинформатика, изучение воздействия горнопромышленного комплекса на компоненты окружающей среды и разработка мероприятий по его снижению. Число научных публикаций — 100. ozaryanigd@gmail.com; улица Тургенева, 51, 680000, Хабаровск, Россия; р.т.: +7(421)232-7927.

Усиков Виталий Игнатьевич — канд. экон. наук, ведущий научный сотрудник, ИГД ДВО РАН. Область научных интересов: геоэкология, геоинформатика, изучение воздействия горнопромышленного комплекса на компоненты окружающей среды и разработка мероприятий по его снижению. Число научных публикаций — 98. v-i-usikov@yandex.ru; улица Тургенева, 51, 680000, Хабаровск, Россия; р.т.: +7(421)232-7927.

Смагин Алексей Сергеевич — канд. техн. наук, старший научный сотрудник, ИГД ДВО РАН. Область научных интересов: компьютерное зрение, машинное обучение, интеллектуальный анализ данных. Число научных публикаций — 30. smaginkhv@gmail.com; улица Тургенева, 51, 680000, Хабаровск, Россия; р.т.: +7(421)222-7267.

Цыгулёв Кирилл Сергеевич — младший научный сотрудник, ВЦ ДВО РАН. Область научных интересов: компьютерное зрение, машинное обучение, интеллектуальный анализ данных, дистанционное зондирование. Число научных публикаций — 24. kirill.tsygulev@mail.ru; улица Ким Ю Чена, 65, 680000, Хабаровск, Россия; р.т.: +7(421)222-7469.

Секриеру Роман Артурович — инженер, ВЦ ДВО РАН. Область научных интересов: компьютерное зрение, машинное обучение, интеллектуальный анализ данных, дистанционное зондирование. Число научных публикаций — 17. roma.sekrieru@gmail.com; улица Ким Ю Чена, 65, 680000, Хабаровск, Россия; р.т.: +7(421)222-7469.

Орлов Сергей Александрович — младший научный сотрудник, ВЦ ДВО РАН. Область научных интересов: компьютерное зрение, машинное обучение, интеллектуальный анализ данных, дистанционное зондирование. Число научных публикаций — 25. programmist00@gmail.com; улица Ким Ю Чена, 65, 680000, Хабаровск, Россия; р.т.: +7(421)222-7469.

Кожевникова Татьяна Владимировна — научный сотрудник, ВЦ ДВО РАН. Область научных интересов: компьютерное зрение, машинное обучение, интеллектуальный анализ данных, дистанционное зондирование. Число научных публикаций — 137. ktvsl@mail.ru; улица Ким Ю Чена, 65, 680000, Хабаровск, Россия; р.т.: +7(421)222-7469.

Поддержка исследований. Исследование выполнено за счет гранта РНФ № 24-11-20029, <https://rscf.ru/project/24-11-20029/> и гранта Правительства Хабаровского края (Соглашение № 107C/2024 от 31.07.2024 г.).

S. SMAGIN, Y. OZARYAN, V. USIKOV, A. SMAGIN, K. TSYGULEV,
R. SEKRIERU, S. ORLOV, T. KOZHEVNIKOVA
**AUTOMATION OF REMOTE ENVIRONMENTAL MONITORING
OF MINERAL RESOURCE COMPLEX FACILITIES IN
KHBAROVSK KRAI**

Smagin S., Ozaryan Y., Usikov V., Smagin A., Tsygulev K., Sekrieru R., Orlov S., Kozhevnikova T. **Automation of Remote Environmental Monitoring of Mineral Resource Complex Facilities in Khabarovsk Krai.**

Abstract. The paper addresses the problem of comprehensive mining and environmental monitoring of mineral resource complex facilities using Earth remote sensing data. The structure of an information system is presented, including modules for data management, processing and analysis of satellite imagery. Earth remote sensing data were used as source information. Processing and analysis of this information was performed to assess the impact of geo-resource extraction and processing on the state of the natural environment at mining enterprise locations and adjacent territories, and to study biota restoration in disturbed areas. Convolutional and transformer neural networks are applied to automate the identification of mining complex objects. Segmentation models for dredge tailings, open-pit mines, tailing dams and coal dust pollution were trained on prepared datasets including annotated satellite data of studied objects in seven regions of the Russian Federation. Methods are proposed for: assessing the impact of coal dust pollution on vegetation based on vegetation indices; mapping forest types using a fully-connected neural network; calculating volumes of dredge tailings based on digital elevation models, polygon centerline and Voronoi diagram construction algorithm; and determining heavy metal content in soil based on analysis of samples collected over a multi-year period using mathematical statistics methods. A correlation analysis was performed on data regarding element content in soil samples and water in the zone of influence of polymetallic ore mining. The proposed methods were tested in the territories of mining enterprises in Khabarovsk Krai. Point analysis of accumulated results from long-term observations and current data on the present state of natural environment objects allows forecasting the development of the studied natural-technical systems in the medium term.

Keywords: information system, monitoring, mining, remote sensing, mineral deposit, environmental safety, computer vision, machine learning.

References

1. Wang Z., Zhou W., Jiskani I.M., Ding X., Luo H. Dust pollution in cold region Surface Mines and its prevention and control. *Environmental Pollution*. 2022. vol. 292. DOI: 10.1016/j.envpol.2021.118293.
2. Wang H., Maqbool A., Xiao X., Yang H., Bi W., Bian Z. Seasonal pollution and risk assessment of heavy metals in atmospheric dust from coal mining area. *International Journal of Environmental Science and Technology*. 2022. vol. 19. № 12. pp. 11963–11972.
3. Blinovskaja Ja.Ju., Mazlova E.A. [Greenhouse gases emissions at coal production and processing: problem status and decrease technologies]. *Uchenye zapiski Rossijskogo gosudarstvennogo gidrometeorologicheskogo universiteta – Scientific notes of the Russian State Hydrometeorological University*. 2019. no. 54. pp. 145–154. (In Russ.).
4. Mun S.A., Larin S.A., Glushkov A.N. The influence of mining on atmosphere contamination and lung cancer in the Kemerovo region. *Sovremennye problemy nauki i obrazovaniya – Modern Problems of Science and Education*. 2013. no. 1. (In Russ.).

5. Afanasov V.I., Vokin V.N., Zayats V.V., Zenkov I.V., Kiryushina E.V., Kondrashov P.M., Konov V.N., Loginova E.V., Maglits Yu.A., Nefedov B.N., Perfilov D.A., Raevich K.V., Spirin T.S., Fedorov A.B., Yuronen Yu.P. [Coal Open Pits of Russia from Space. Mining Operations and Ecology of Disturbed Lands]. Krasnoyarsk: Sibirskiy federalnyy universitet – Siberian Federal University, 2017. 519 p. (In Russ.).
6. Liu Y., Heng W., Yue H. Quantifying the coal mining impact on the ecological environment of Gobi open-pit mines. *Science of the total environment*. 2023. vol. 883.
7. Pan Y., Chen M., Wang X., Chen Y. Ecological risk, source apportionment, and influencing factors of heavy metals in soil in a typical lead-zinc mining watershed, Guangxi, China. *Journal of Environmental Chemical Engineering*. 2024. vol. 12. no. 3.
8. Baieta R., Ettler V., Vaněk A., Drahota P., Křibek B., Nyambe I., Mihaljevič M. Smelter-derived soil contamination in Luanshya, Zambia. *Science of The Total Environment*. 2023. vol. 867.
9. Karn R., Ojha N., Abbas S., Bhugra S. A review on heavy metal contamination at mining sites and remedial techniques. *IOP Conference Series: Earth and Environmental Science*. IOP Publishing, 2021. vol. 796. № 1.
10. Haghighizadeh A., Rajabi O., Nezarat A., Hajyani Z., Mina Haghmohammadi M., Hedayatikhah S., Asl S.D., Beni A.A. Comprehensive analysis of heavy metal soil contamination in mining Environments: Impacts, monitoring Techniques, and remediation strategies. *Arabian Journal of Chemistry*. 2024. vol. 17. no. 6.
11. Bakshi S., Banik C., ZhenLi H. The impact of heavy metal contamination on soil health. *Managing soil health for sustainable agriculture*. Volume 2: Monitoring and management. 2018. pp. 63–95.
12. Shpirt M.I.A., Laskorin B.N. [Waste-Free Technology: Utilization of Wastes from Mining and Processing of Solid Fossil Fuels]. Nedra – Bowels. 1986. 255 p. (In Russ.).
13. Leonov S.B., Fedotov K.V., Senchenko A.E. [Industrial Gold Extraction from Ash-Slag Dumps of Thermal Power Plants]. *Gorniy zhurnal – Mining Journal*. 1998. vol. 5. pp. 67–68. (In Russ.).
14. Lupyan E.A., Konstantinova A.M., Balashov I.V., Kashnitskiy A.V., Savorskiy V.P., Panova O.Yu. [Development of a system for analyzing the state of environment in areas of large industrial facilities, tailings and dumps]. *Sovremennyye problemy distantsionnogo zondirovaniya Zemli iz kosmosa – Modern Problems of Remote Sensing of the Earth from Space*. 2020. vol. 17. no. 7. pp. 243–261. (In Russ.).
15. Qdais H.A., Begday I.V., Katorgin I.Y., Shkarlet K.Y., Kharin K.V., Bluzhina A.S., Likhovid A.A. Assessment of metals pollution from tailing sites in the North Caucasus Region, Russia. *Mine Water and the Environment*. 2018. vol. 37. no. 4. pp. 815–824.
16. Shapovalov V.V., Kozyr D.A. [Resource-Saving Technology for Utilization of Waste Dumps of Mining Industries]. *Izvestiya Tomskogo politekhnicheskogo universiteta. Inzhiniring georesurov – Bulletin of the Tomsk Polytechnic University. Geo Assets Engineering*. 2023. vol. 334. no. 4. pp. 175–184. (In Russ.).
17. Churkin O.E., Gilyarova A.A. [Mining wastes managements as an investment direction for the development of the Kila Peninsula mining industry]. *Ekonomika, predprinimatelstvo i pravo – Journal of economics, entrepreneurship and law*. 2020. vol. 10. no. 3. pp. 905–916. (In Russ.).
18. Alsayed A., Nabawy M.R.A. Stockpile volume estimation in open and confined environments: a review. *Drones*. 2023. vol. 7. no. 8.
19. Ren H., Zhao Y., Xiao W., Hu Z. A review of UAV monitoring in mining areas: Current status and future perspectives. *International journal of coal science & technology*. 2019. vol. 6. pp. 320–333.

20. Talovskaya A.V., Yazikov E.G., Shakova T.S., Filimonenko E.A. [Assessment of Aerotechnogenic Pollution: case study in the vicinity of coal-fired and oil-fired local boiler in Tomsk Region)]. *Izvestiya Tomskogo politekhnicheskogo universiteta. Inzhiniring georesursov – Bulletin of the Tomsk Polytechnic University. Geo Assets Engineering*. 2016. vol. 327. no. 10. pp. 116–130. (In Russ.).
21. Kazantseva U.D., Yakovlenko O.S., Leshukova M.K., Legoshchin K.V., Leshukov T.V., Lariionov A.V. [Morphological characteristics, size and mass of PM0.1 dust near the open coal mines]. *Uspekhi sovremennogo estestvoznaniya – Advances in Current Natural Science*. 2023. no. 2. pp. 90–95. (In Russ.).
22. Kong S., Lu B., Ji Y., Zhao X., Chen L., Li Z., Han B., Bai Z. Levels, risk assessment and sources of PM10 fraction heavy metals in four types dust from a coal-based city. *Microchemical Journal*. 2011. vol. 98. no. 2. pp. 280–290.
23. Oparin V.N., Potapov V.P., Ginyatulina O.L., Andreeva N.V., Shchastlivtsev E.L., Bykov A.A. [Assessment of Dust Pollution of the Atmosphere in Coal Mining Areas of Kuzbass in Winter Period Based on Remote Sensing Data]. *Fiziko-tekhnicheskie problemy razrabotki poleznykh iskopaemykh – Physico-Technical Problems of Mineral Development*. 2014. no 3. pp. 126–137. (In Russ.).
24. Struchkova G.P., Krupnova T.G., Tikhonova S.A., Kapitonova T.A. [Snow cover pollution analysis in coal mining regions using color response curves]. *Gorniy informatsionno-analiticheskiy byulleten (nauchno-tekhnicheskiy zhurnal) – Mining Informational and Analytical Bulletin (Scientific and Technical Journal)*. 2021. no. 12-1. pp. 195–203. (In Russ.).
25. Krupnova T.G., Rakova O.V., Struchkova G.P., Tikhonova S.A., Kapitonova T.A., Gavrilkina S.V., Bulanova A.V., Yakimova O.N. Insights into particle-bound metal(loids) in winter snow cover: Geochemical monitoring of the Korkinsky coal mine area, south Ural region, Russia. *Sustainability*. 2021. vol. 13. no. 9.
26. Chen W., Wang X., Cui J., Cao X., Pu W., Zheng X., Ran H., Ding J. Radiative forcing of black carbon in seasonal snow of wintertime based on remote sensing over Xinjiang, China. *Atmospheric Environment*. 2021. vol. 247.
27. Ahmad M., Alam K., Tariq S., Blaschke T. Contrasting changes in snow cover and its sensitivity to aerosol optical properties in Hindukush-Karakoram-Himalaya region. *Science of the Total Environment*. 2020. vol. 699.
28. Sahu, R., Gupta R.D. Snow cover area analysis and its relation with climate variability in Chandra basin, Western Himalaya, during 2001–2017 using MODIS and ERA5 data. *Environmental Monitoring and Assessment*. 2020. vol. 192(8). DOI: 10.1007/s10661-020-08442-8.
29. Tigeev A.A., Aksenov N.V., Moskovchenko D.V., Pozhitkov R.Yu. [Assessment of Atmospheric Dust Contamination with Ground-based and Remote Sensing Methods (on the example of the Town of Tobolsk)]. *Geograficheskiy vestnik – Geographical Bulletin*. 2021. no. 2(57). pp. 121–134. (In Russ.).
30. Nie X., Hu Z., Ruan M., Zhu Q., Sun H. Remote-sensing evaluation and temporal and spatial change detection of ecological environment quality in coal-mining areas. *Remote Sensing*. 2022. vol. 14. no. 2.
31. Li Z., Zhao Y., Ren H., Sun Y. A Novel Index for Detecting Bare Coal in Open-Pit Mining Areas Based on Landsat Imagery. *Remote Sensing*. 2024. vol. 16. no. 24.
32. Xia N., Hai W., Song G., Tang M. Identification and monitoring of coal dust pollution in Wucaiwan mining area, Xinjiang (China) using Landsat derived enhanced coal dust index. *Plos one*. 2022. vol. 17. no. 4.
33. Yu H., Zahidi I. Environmental hazards posed by mine dust, and monitoring method of mine dust pollution using remote sensing technologies: An overview. *Science of the total environment*. 2023. vol. 864.

34. Wei X., Chang N.B., Bai K., Gao W. Satellite remote sensing of aerosol optical depth: advances, challenges, and perspectives. *Critical Reviews in Environmental Science and Technology*. 2020. vol. 50. no. 16. pp. 1640–1725.
35. Antoninova N.Yu., Rybnikova L.S., Slavikovskaya Yu.O., Shubina L.A. [Environmental and Economic Aspects of Selecting Reclamation Directions for Industrial Mining and Metallurgical Waste Disposal Sites]. *Gornaya promyshlennost – Mining Industry*. 2022. no. S1. pp 71–77. (In Russ.).
36. Musina G.A. [Comparative analysis of modern methods of deformation monitoring]. *Interjekspo GEO-Sibir' – Interexpo GEO-Siberia*. 2023. vol. 4. no. 2. pp. 82–90. (In Russ.).
37. Bagaev S.M., Medvedeva E.V. [Segmentation of satellite images based on the u-net convolutional neural network]. *Cifrovaja obrabotka signalov i ee primenenie (DSPA-2021): dokl. 23-j mezhdunar. konf. – Digital signal processing and its applications (DSPA-2021): Proceedings of the 23rd International Conference*. 2021. pp. 218–222. (In Russ.).
38. Kumar A., Gorai A.K. Development of a deep convolutional neural network model for detection and delineation of coal mining regions. *Earth Science Informatics*. 2023. vol. 16. no. 2. pp. 1151–1171.
39. Gallwey J., Robiati C., Coggan J., Vogt D., Eyre M. A Sentinel-2 based multispectral convolutional neural network for detecting artisanal small-scale mining in Ghana: Applying deep learning to shallow mining. *Remote Sensing of Environment*. 2020. vol. 248.
40. Ronneberger O., Fischer P., Brox T. U-net: Convolutional networks for biomedical image segmentation. *Proceedings of the 18th international conference Medical image computing and computer-assisted intervention (MICCAI)*, part III. 2015. pp. 234–241.
41. Chen L.C., Papandreou G., Florian S., Hartwig A. Rethinking atrous convolution for semantic image segmentation. *arXiv preprint arXiv:1706.05587*. 2017.
42. Xie E., Wang W., Yu Z., Anandkumar A., Alvarez J.M., Luo P. SegFormer: Simple and efficient design for semantic segmentation with transformers. *Advances in neural information processing systems*. 2021. vol. 34. pp. 12077–12090.
43. Cao H., Wang Y., Chen J., Jiang D., Zhang X., Tian Q., Wang M. Swin-unet: Unet-like pure transformer for medical image segmentation. *European conference on computer vision*. Cham: Springer Nature Switzerland. 2022. pp. 205–218.
44. Bentéjac C., Csörgő A., Martínez-Muñoz G. A comparative analysis of gradient boosting algorithms. *Artificial Intelligence Review*. 2021. vol. 54(3). pp. 1937–1967.
45. Biau G., Scornet E. A random forest guided tour. *Test*. 2016. vol. 25. no. 2. pp. 197–227.
46. Khasanov K. Evaluation of ASTER DEM and SRTM DEM data for determining the area and volume of the water reservoir. *IOP Conference Series: Materials Science and Engineering*. IOP Publishing, 2020. vol. 883. no. 1.
47. Tadono T., Ishida H., Oda F., Naito S., Minakawa K., Iwamoto H. Precise global DEM generation by ALOS PRISM. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*. 2014. vol. 2. pp. 71–76.
48. Li H., Zhao J., Yan B., Yue L., Wang L. Global DEMs vary from one to another: an evaluation of newly released Copernicus, NASA and AW3D30 DEM on selected terrains of China using ICESat-2 altimetry data. *International Journal of Digital Earth*. 2022. vol. 15. no. 1. pp. 1149–1168.
49. Marsh C.B., Harber P., Pomeroy J.W. Validation of FABDEM, a global bare-earth elevation model, against UAV-lidar derived elevation in a complex forested mountain catchment. *Environmental Research Communications*. 2023. vol. 5. no. 3.

50. Ho Y.F., Grohmann C.H., Lindsay J., Reuter I.H., Parente L., Witjes M., Hengl T. GEDTM30: Global Ensemble Digital Terrain model at 30 m and derived multiscale terrain variables. *PeerJ*. 2025. vol. 13. DOI: 10.7717/peerj.19673.
51. Filatova M.Yu. [Evaluation of ecological conditions in the influence zone of gold mining (manmade forest justification)]. *Gorniy informatsionno-analiticheskiy byulleten (nauchno-tekhnicheskiy zhurnal) – Mining Informational and Analytical Bulletin (Scientific and Technical Journal)*. 2023. no. 8. pp. 27–44. (In Russ.).
52. Lemesheko B.Yu. Criteria for Checking Deviation of Distribution from the Normal Law. Application Guide. 2014. 192 p. (In Russ.).
53. Zhao Y., Wang Z., Sun W., Huang B., Shi X., Ji J. Spatial interrelations and multi-scale sources of soil heavy metal variability in a typical urban–rural transition area in Yangtze River Delta region of China. *Geoderma*. 2010. vol. 156. no. 3–4. pp. 216–227.
54. Yuan G.L., Sun T.H., Han P., Li J., Lang X.X. Source identification and ecological risk assessment of heavy metals in topsoil using environmental geochemical mapping: typical urban renewal area in Beijing, China. *Journal of geochemical exploration*. 2014. vol. 136. pp. 40–47.

Smagin Sergey — Ph.D., Dr.Sci., Professor, Academician of the RAS, Chief researcher, Computing Center FEB RAS. Research interests: mathematical modeling, information and computing technologies. The number of publications — 200. smagin@ccfebras.ru; 65, Kim Yu Chena St., 680000, Khabarovsk, Russia; office phone: +7(421)222-7267.

Ozaryan Yulia — Ph.D., Leading researcher, Institute of Mining FEB RAS. Research interests: geocology, geoinformatics, study of mining industry impacts on environmental components and development of measures to reduce them. The number of publications — 100. ozaryanigd@gmail.com; 51, Turgenev St., 680000, Khabarovsk, Russia; office phone: +7(421)232-7927.

Usikov Vitaly — Ph.D., Leading researcher, Institute of Mining FEB RAS. Research interests: geocology, geoinformatics, study of mining industry impacts on environmental components and development of measures to reduce them. The number of publications — 98. v-i-usikov@yandex.ru; 51, Turgenev St., 680000, Khabarovsk, Russia; office phone: +7(421)232-7927.

Smagin Aleksey — Ph.D., Senior researcher, Institute of Mining FEB RAS. Research interests: computer vision, machine learning, data mining. The number of publications — 30. smaginkhv@gmail.com; 51, Turgenev St., 680000, Khabarovsk, Russia; office phone: +7(421)222-7267.

Tsygulev Kirill — Junior researcher, Computing Center FEB RAS. Research interests: computer vision, machine learning, data mining, remote sensing. The number of publications — 24. kirill.tsygulev@mail.ru; 65, Kim Yu Chena St., 680000, Khabarovsk, Russia; office phone: +7(421)222-7469.

Sekrieru Roman — Engineer, Computing Center FEB RAS. Research interests: computer vision, machine learning, data mining, remote sensing. The number of publications — 17. roma.sekrieru@gmail.com; 65, Kim Yu Chena St., 680000, Khabarovsk, Russia; office phone: +7(421)222-7469.

Orlov Sergey — Junior researcher, Computing Center FEB RAS. Research interests: computer vision, machine learning, data mining, remote sensing. The number of publications — 25.

programmist00@gmail.com; 65, Kim Yu Chena St., 680000, Khabarovsk, Russia; office phone: +7(421)222-7469.

Kozhevnikova Tatyana — Researcher, Computing Center FEB RAS. Research interests: computer vision, machine learning, data mining, remote sensing. The number of publications — 137. ktvsl@mail.ru; 65, Kim Yu Chena St., 680000, Khabarovsk, Russia; office phone: +7(421)222-7469.

Acknowledgements. This research is supported by the Russian Science Foundation (grant 24 11 20029 – <https://rscf.ru/project/24-11-20029/>) and the Government of Khabarovsk Krai (agreement 107C/2024 dated 31.07.2024).

А.М. ФЕДОРОВ, И.О. ДАТЬЕВ, И.Г. ВИШНЯКОВ
**МЕТОД ИНТЕГРАЦИИ БОЛЬШИХ ЯЗЫКОВЫХ МОДЕЛЕЙ
В АЛГОРИТМЫ ФОКУСИРОВАННОГО МОНИТОРИНГА
ОТКРЫТЫХ ДАННЫХ СОЦИАЛЬНЫХ МЕДИА**

Федоров А.М., Датьев И.О., Вишняков И.Г. Метод интеграции больших языковых моделей в алгоритмы фокусированного мониторинга открытых данных социальных медиа.

Аннотация. Актуальность исследования определяется важностью и сложностью выполнения оперативных обобщений постоянно возрастающего массива пользовательских сообщений в социальных сетях. Предлагается уменьшить сложность решаемой задачи за счет использования роботизированных алгоритмов и их автоматизированной интеллектуальной фокусировки на целевые платформы, доступности данных и их объема. Рассматривается способность больших языковых моделей (LLMs) генерировать высококачественные, связные и контекстно-зависимые аннотации (рефераты), подходящие для динамической природы неструктурированных, «зашумленных» данных социальных сетей. Представлены особенности технологии RAG LLM при реферировании публикаций социальных сетей. Основным недостатком языковых моделей является нестабильность и сложность прослеживания результатов для подтверждения фактической точности. Авторами предложен гибридный метод для обобщения сообщений в социальных сетях за определенный период времени. Метод заключается в комплексном и вариативном сочетании классических способов извлечений данных из их хранилищ, а также реферативных и генеративных возможностей больших языковых моделей. Большие языковые модели использованы для векторизации анализируемых данных. Применение алгоритмов кластеризации к полученным векторным представлениям позволило повысить стабильность и качество результатов. В рамках технологии RAG возможности больших языковых моделей расширены средствами интеллектуального поиска в используемой для хранения исходных данных базе MongoDB. В работе представлены три конвейера, каждый из которых является вариантом реализации метода и обладает преимуществами и недостатками в различных условиях применения. Приведены используемые для оценки конвейеров метрики и произведен сравнительный анализ. В целом, метод позволяет уменьшить конфабуляции большой языковой модели и получать обобщения публикаций за разные временные периоды в режиме реального времени. Предложенный метод применяется на практике в разработанной авторами системе мониторинга открытых данных социальных медиа.

Ключевые слова: социальные медиа, публикации, обобщение, большие языковые модели, генерация дополненная поиском, интеллектуальные агенты, гибридный метод.

1. Введение. В настоящее время активно развиваются средства интеллектуальной обработки данных. Одним из самых мощных и универсальных инструментов в этой области являются технологии искусственного интеллекта, реализующие их нейронные сети, в частности большие языковые модели (LLM – large language models). Флагманами разработки в этой сфере регулярно демонстрируются новые технологии, средства и получаемые с их помощью результаты.

Такое мощное развитие индустрии технологий искусственного интеллекта требует вложения очень больших ресурсов: материальных, энергетических, производственных, научных, кадровых и др. В свою очередь разработчики среднего звена используют доступные флагманские результаты, развивают их, в меру имеющихся сил совершенствуют и применяют на практике. Несмотря на все достижения, потенциал развития в области технологий искусственного интеллекта остается очень высоким. Многие эксперты сейчас говорят о том, что аббревиатуру ИИ скорее можно расшифровать как «имитация интеллекта», и что настоящий искусственный интеллект только еще впереди.

Тем не менее, доступный ИИ-инструментарий в виде больших языковых моделей ускорил решение задач обработки и генерирования текста. Для управления работой таких моделей уже не нужно быть программистом. Языком управления стал естественный язык, с помощью которого составляется запрос (промт). Однако даже в этом случае следует понимать, что от способа формулирования, формирования, оформления, наполнения запросов зависит качество получаемого результата.

В настоящей работе авторы решают задачу обобщения публикаций социальных сетей на основе результатов мониторинга созданной авторами программной системы получения и анализа данных социальных медиа [1] с использованием программных веб-роботов (ботов) для сбора сообщений с платформ VK, Telegram, Whatsap. Обобщение или суммаризация публикаций – получение (синтез, генерация) тематических абстрактов (аннотаций, рефератов, дайджестов, тем), другими словами, получение сокращенного изложения содержания исходного текста. Практической целью данной процедуры является решение различных задач автоматизированной обработки текста, в том числе, в данной работе, сокращение объема текста для дальнейшей обработки оператором системы мониторинга.

Исходными данными являются тексты пользовательских сообщений, получаемые в процессе мониторинга открытых данных социальных медиа. Объем таких текстов в общем случае ничем не ограничен. В работе реализован методологический подход и рассматриваются все возможные варианты объема исследуемых сообщений: 1) помещаются в контекст, 2) не помещаются в контекст. Счетный, но бесконечный характер числа пользовательских сообщений позволяет выбрать такое их количество, что оно не будет помещаться в контекстное окно используемых языковых моделей, даже с учетом постоянного роста их возможностей.

На основе заданного временного интервала для текстовой выборки публикаций социальных сетей необходимо сгенерировать обобщения (резюме, рефераты, аннотации, дайджесты, темы) и минимизировать возможные ошибки, обеспечить стабильность результата и его прослеживаемость, т.е. подтверждение на основе исходных данных.

Формально данная задача поставлена так: для получения оперативных и качественных тематических обобщений над элементами массива текстов *ДТ* объемом *ОДТ* найти удовлетворяющий заданному критерию *K* вариант реализации конвейера обработки на основе интеллектуального компонента в виде большой языковой модели *ЯМ* с объемом контекста *ОЯМ*. Комплексный критерий *K* состоит из двух: 1) «*воспроизводимость*» и 2) «*прослеживаемость*». Обе составляющие *K* используются для оценки негативных эффектов – галлюцинаций/конфабуляций, возникающих в результате использования генеративной функции *ЯМ*. Первый критерий позволяет оценить стабильность получения результата при неизменных начальных условиях, а второй – возможность восстановить исходные данные на основе полученного результата. Дополнительным условием является решение поставленной задачи с учетом разных соотношений *ОЯМ* и *ОДТ*.

Таким образом, основная обладающая научной новизной задача, поставленная в данном исследовании, заключается в выявлении, изучении, системном анализе и практической реализации способов получения генеративных обобщений при снижении конфабуляций языковых моделей, используемых в конвейерах решений.

2. Социальные сети и большие языковые модели. Большой объем пользовательского контента на платформах социальных сетей способствует повышению актуальности разработки методов резюмирования (аннотирования, реферирования, обобщения, суммаризации) сообщений за определенные временные периоды. Искусственный интеллект, в частности достижения в обработке естественного языка (NLP), привели к появлению сложных методов автоматического резюмирования, адаптированных к особенностям специфических данных социальных сетей. Автоматическое резюмирование текста (ATS – Automatic Text Summarization) значительно изменилось за время своего существования. Алгоритмы резюмирования могут быть извлекающими (выбор ключевых предложений), абстрактными (генерация новых предложений), а также гибридными. Ранние методы основывались на статистических моделях и системах на основе правил, в основном извлекая ключевые

предложения. Улучшению результатов аннотирования также способствовали методы тематического моделирования, в том числе метод LDA (Latent Dirichlet Allocation)[2], более подробный обзор которых и применение к текстам социальных сетей представлены в работе [3]. В настоящее время область решения сместилась в сторону машинного обучения, при этом предварительно обученные большие языковые модели (LLM), такие как GPT, BERT, DeepSeek, GigaChat, YandexGPT демонстрируют значительные улучшения в создании связных и контекстно-зависимых резюме на основе «зашумленных» публикаций социальных сетей. В работе [4] представлен сравнительный анализ методов суммаризации текста с использованием больших языковых моделей. Авторы работы [5] отмечают, что интеграция технологий ИИ и генеративного ИИ предлагает многообещающие способы решения задач исследовательского поиска и резюмирования.

Сообщения в социальных сетях представляют собой совокупность текста и различных метаданных, например, эмодзи, эмодзики и мультимедиа. Нарушение последовательности слов и неформальный язык, использующиеся в социальных сетях, являются серьезным препятствием для использования только статистических методов, в том числе алгоритмов выделения ключевых слов, таких как Rake [6], YAKE [7], TextRank [8]. Реферирование этого контента требует моделей, которые могут эффективно обрабатывать сообщения, обладающие этими особенностями. В исследовании [9] проведен обзор гибридных методов реферирования и предложена система, объединяющая семантические (основанные на значении) и статистические (основанные на частоте) методы для обобщения постов в Instagram, что позволило повысить точность суммаризации.

В работе [10] освещается использование представлений на основе векторов, методов снижения размерности и технологий генеративного ИИ (например, LLM с подсказками Chain of Thought (CoT) – цепочки рассуждений [11, 12]) для выявления тем в больших открытых данных социальных сетей.

LLM превосходит другие методы суммаризации по связности результата и адаптивности к различным предметным областям, включая социальные сети. Способность выполнять контекстное и малочастотное обучение позволяет LLM генерировать краткие и контекстно релевантные обобщения даже из динамического и неформального контента, найденного на этих платформах [4].

Для извлечения и обобщения доминирующих тем из данных социальных сетей за определенный временной отрезок также

используются агентные генеративные системы ИИ, что позволяет масштабировать и адаптировать тематический анализ, что представлено в работе [10].

Все больше внимания уделяется тому, чтобы сделать сгенерированные технологиями ИИ резюме более прозрачными. Концепция "отслеживаемого текста" представлена в [13], и направлена на связывание контента резюме с его первоначальными источниками для повышения доверия пользователей и интерпретируемости выходных данных ИИ.

3. Дополненная поиском генерация (RAG). Технология Retrieval-Augmented Generation (RAG) – это гибридный подход в обработке естественного языка, который сочетает в себе способности к генерации больших языковых моделей с внешними системами извлечения информации [14]. Этот подход помогает LLM преодолевать такие ограничения, как галлюцинации, устаревшие знания и отсутствие специфики предметной области, позволяя моделям получать доступ к актуальным и релевантным внешним данным и использовать их во время генерации. Системы RAG обычно состоят из двух основных компонентов: т.н. ретривер (компонент поиска) и генератор. Ретривер находит и извлекает релевантные документы или фрагменты данных из внешней базы знаний на основе входного запроса. Генератор формирует свой вывод, основываясь как на исходном запросе, так и на извлеченном контексте, чтобы сгенерировать окончательный ответ. RAG быстро набирает популярность в различных областях, включая ответы на вопросы из различных предметных областей и суммаризацию текстов. В компоненте поиска могут использоваться и сочетаться различные методы, например, разреженный поиск (традиционный поиск по ключевым словам, алгоритм BM25), плотный поиск (семантический поиск с использованием векторизации на основе модели Sentence Transformers).

Авторы работы [15] отмечают, что для эффективного извлечения релевантной информации из больших неструктурированных данных социальных сетей требуются продвинутые стратегии поиска, и предлагают использовать "sentence window retrieval" и "auto-merging", чтобы найти наиболее точные и контекстуально релевантные фрагменты.

В работе [16] авторы оценивали различные стратегии поиска для технологии RAG и пришли к выводу, что гибридный поисковик (объединяющий разреженные и плотные методы, использующий расширение запроса и обратное ранговое слияние) показал наилучшие

результаты. В качестве общих причин галлюцинаций указываются: поиск, не приведший к результату и недостаточная (неполноценная) генерация. При гибридном подходе улучшилась релевантность поиска документов, понизились показатели галлюцинаций и повысилась точность ответов при тестировании на таких бенчмарках, как HaluBench.

Многоагентные гибридные фреймворки помимо классического поиска, могут включать несколько специализированных «агентов» или модулей. Каждый агент может фокусироваться на другом аспекте рассуждения или интеграции знаний. Примером является фреймворк RAG-KG-IL [17], объединяющий RAG для ответов на основе доказательств, граф знаний (KG) для предоставления структурированного фактического контекста, а также итеративное обучение (IL) для постоянного улучшения результата и исправления ошибок. Это многоагентное сотрудничество помогает улучшить рассуждения, уменьшить галлюцинации и динамически выбирать и объединять источники знаний.

Технология RAG является ключевой при разработке больших языковых моделей (LLM). Надежность систем RAG находится в стадии изучения. Авторы [18] предлагают фреймворк, который оценивает системы RAG по шести ключевым измерениям: фактичность, надежность, справедливость, прозрачность, подотчетность и конфиденциальность. Следует отметить, что решение не является исчерпывающим, авторы лишь предпринимают попытку заложить структурированную основу для будущих исследований и предоставить практические идеи для повышения надежности систем RAG в реальных приложениях. Человеческая экспертная оценка до сих пор остается наиболее важной.

4. Особенности RAG для суммаризации постов в социальных сетях. Социальные сети содержат огромное количество разнообразной и постоянно меняющейся информации. RAG позволяет LLM извлекать наиболее релевантные данные, обеспечивая более точные и актуальные сводки. Например, в исследовании, касающемся ответов на медицинские вопросы с использованием данных социальной сети Reddit, двухслойная RAG-система показала сравнимую производительность с GPT-4 в релевантности и связности в условиях ограниченных ресурсов [19].

LLM могут "галлюцинировать", то есть генерировать ложную информацию. Благодаря RAG, модель получает фактические данные из базы знаний, что уменьшает вероятность таких ошибок и делает сводки более надежными [20].

RAG может учитывать специфический контекст социальных медиа, включая сленг, сокращения и быстро меняющиеся тренды. В работе [21] авторы продемонстрировали возможности технологии RAG для прогнозирования реакций сообщества на посты в социальных сетях, учитывая идеологическое и эмоциональное разнообразие.

Одной из важнейших тем в исследованиях ИИ является вопрос о том, какой из подходов – технология RAG, тонкая настройка (fine-tuning) [22, 23] моделей или гибридный подход – лучше? Каждый метод имеет свои преимущества и недостатки, поэтому необходимо понимать их применимость.

Тонкая настройка подразумевает обновление весов предварительно обученного LLM на специфическом для предметной области наборе данных и требует повторного обучения каждый раз при поступлении новой информации, что в свою очередь, для больших моделей занимает много времени и требует вычислительных затрат. RAG обеспечивает актуальные ответы за счет извлечения знаний в реальном времени, но обладает большей задержкой из-за предварительного этапа поиска. Гибридный подход сочетает тонкую настройку для адаптации к предметной области с RAG для обновлений в реальном времени. Таким образом, выбор между тонкой настройкой, RAG или гибридным подходом зависит от компромисса между адаптивностью в реальном времени, вычислительными затратами и специфичностью предметной области [24].

5. Многошаговое промптирование или цепочки рассуждений. При работе с языковыми моделями очень важным элементом является т.н. промптирование, т.е. использование правильных запросов-подсказок. В работе [25] авторы исследуют различные варианты подсказок и демонстрируют соответствующие улучшения выдачи языковой модели.

Наиболее перспективным для решения задач резюмирования представляется метод CoT (Chain-of-Thought) – цепочки рассуждений – это метод, который побуждает LLM генерировать промежуточные шаги рассуждения перед тем, как дать окончательный ответ. Это улучшает прозрачность и качество рассуждений модели. Работы, содержащие прямое использование CoT с RAG для суммаризации публикаций социальных сетей все ещё находятся на ранних этапах развития, однако представляются перспективными по следующим причинам. CoT может сделать процесс суммаризации более понятным, показывая, как модель сформировала резюме. Социальные сети могут содержать сложные дискуссии и взаимосвязи между постами. CoT, в сочетании с RAG, может помочь LLM выявить

эти сложные взаимодействия и создать более содержательные резюме. Например, в работе [26] авторы используют графовый поиск для многошагового рассуждения, что концептуально схоже с CoT. Путем извлечения и интеграции информации из взаимосвязанных документов (например, цепочек в социальных сетях), предлагаемый фреймворк GRAG (Graph Retrieval-Augmented Generation) позволяет LLM выполнять пошаговые рассуждения с использованием графов, улучшая аннотирование и качество выводов.

Обобщение сообщений в социальных сетях за определенные периоды времени с помощью технологий ИИ – это быстро развивающаяся область, значительно усиленная гибридными методами и большими языковыми моделями RAG LLM. Эти технологии позволяют создавать эффективные и контекстно-релевантные резюме, которые полезны для таких приложений, как мониторинг событий и динамический (временной) тематический анализ. Текущие исследования продолжают решать проблемы, связанные со спецификой данных социальных сетей, а также прозрачностью и прослеживаемостью сгенерированных резюме.

При применении методов извлекающего резюмирования может отсутствовать связность. Абстрактное резюмирование более связно, но требует больших вычислительных ресурсов. Тематическое извлечение на основе эмбедингов хорошо работает для выявления тем, но требует больших наборов предварительно подготовленных данных. Временное тематическое моделирование эффективно для обнаружения событий и временного анализа, но может упускать нюансы контекста. Технологии генеративного ИИ с подсказками в виде цепочки рассуждений (CoT) создает более высококачественные резюме, но зависит от возможностей LLM.

Большие языковые модели (LLM) являются мощным средством для решения задачи резюмирования, но склонны к «галлюцинациям» или конфабуляциям – генерации результатов, которые фактически неверны или не подтверждаются доказательствами. Гибридные методы, которые объединяют LLM с другими стратегиями, такими как расширенные методы поиска (в частности технология RAG) или многоагентные системы, наиболее перспективны для повышения надежности LLM и уменьшения их подверженности галлюцинациям.

RAG LLM представляют собой значительный шаг вперед в суммаризации постов социальных сетей, предлагая улучшенную фактическую точность, контекстуальную привязку и адаптируемость к динамичному контенту. Методологические инновации в поиске и инженерии подсказок, включая развивающиеся исследования

промптирования цепочек рассуждения, расширяют возможности этих систем. Необходимы масштабируемые, адаптируемые фреймворки и надежные оценочные метрики для результатов резюмирования публикаций в социальных сетях. На сегодняшний день, человеческая экспертная оценка остается наиболее важной.

6. Метод интеграции больших языковых моделей в алгоритмы фокусированного мониторинга. Используемый в рамках системы мониторинга социальных медиа механизм фокусировки предполагает динамическую корректировку заданий на мониторинг с учетом поступающих данных. Процесс такой корректировки представляет собой состоящих из нескольких шагов конвейер, в рамках которого полученные на предыдущем шаге результаты передаются на следующий и т.д. Работоспособность конвейера зависит от качества получаемых на каждом шаге данных. Важным является повышение этого качества.

Идея метода заключается в создании надежных и стабильных конвейеров, обеспечивающих качественный результат работы.

В рамках решения задачи обобщения множества сообщений в понятие качества также включается гарантированная возможность проверки полученного результата выборкой исходных данных.

Получение результирующих обобщений может рассматриваться и в качестве самостоятельной цели, и в качестве промежуточного этапа процедуры фокусировки. В любом случае от результата в первую очередь требуется стабильная воспроизводимость и прослеживаемость. Первое свойство можно проверить серией экспериментов. Второе свойство требует работы с дополнительными конструкциями, позволяющими на основе полученных результатов найти исходные данные.

Последовательность действий (блок-схема алгоритма) метода для получения искомого результата представлена на рисунке 1 в виде стандартной UML-диаграммы деятельности. Однако изображение в таком виде менее выразительно, так как не позволяет продемонстрировать ряд важных специфических аспектов, к которым в том числе относятся и объектная модель реализации алгоритма, и временная последовательность действий, и параллельность взаимодействия серверных и клиентских акторов. По этой причине здесь и далее в статье для иллюстрации реализованных алгоритмов авторами используются диаграммы последовательности, как более подходящий инструмент из многообразной палитры унифицированного языка моделирования UML.

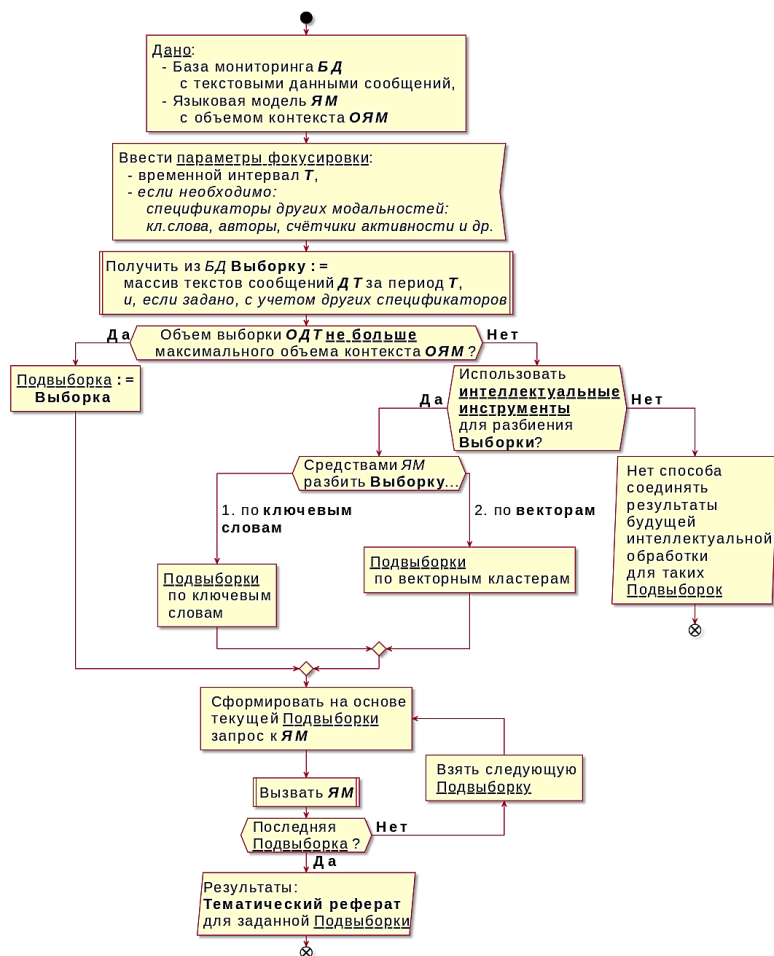


Рис. 1. Блок-схема метода получения обобщений для сообщений

Далее рассматриваются три конвейера, в которых несколькими способами решается поставленная задача.

6.1. Прямой запрос к большой языковой модели. Тривиальным вариантом получения генеративных обобщений (списка тем) является непосредственный запрос к большой языковой модели. Предварительно по заданным критериям в базе данных мониторинга производится поиск необходимых данных. Структурированный список полученных текстов сообщений включается в контекст запроса. Пример формулировки

запроса: «Выдели пять наиболее крупных тем в представленном ниже списке пользовательских сообщений: {список сообщений}».

В данном случае на основе технологии RAG большая языковая модель, используя свое умение работать с текстами на естественном языке и основываясь только на содержании запроса, в качестве результата выдаст список формулировок тем.

На рисунке 2 в виде UML диаграммы последовательности представлен порядок действий данного варианта реализации алгоритма получения тем для заданного списка сообщений.

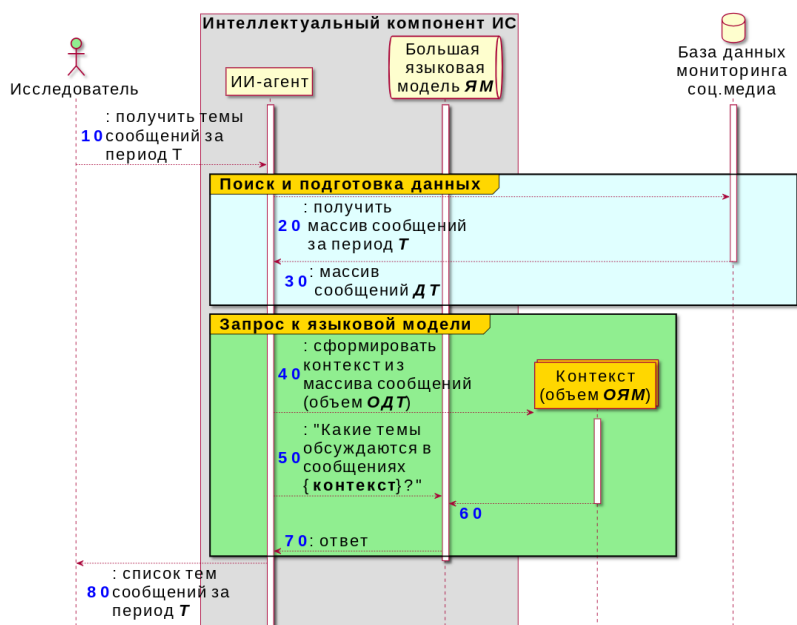


Рис. 1. Извлечение тем из списка сообщений путем прямого запроса к языковой модели

Особенностью данного подхода является то, что при повторении последовательности действий при одних и тех же исходных данных формулировки тем в целом семантически будут соответствовать ранее полученным вариантам, но синтаксически будут различаться. Эта нестабильность связана с вероятностными свойствами генеративной функции языковой модели. Снизить вероятность таких изменений можно с помощью конфигурационного параметра «температура модели», задав ему самое минимальное значение.

Преимуществом данного подхода является классическая для прикладных языковых моделей одношаговая функциональная реализация и зависящая лишь от объема контекста скорость выполнения. Для его реализации достаточно сформировать один запрос к большой языковой модели, разместив в нем список необходимых для анализа сообщений.

6.2. Предварительное определение ключевых слов. Для повышения стабильности результатов получения генеративных обобщений (тем) из данных социальных медиа предложен вариант, основанный на предварительном получении ключевых слов и последующем их использовании в фокусировке алгоритмов именования тем.

В отличие от предложенного в предыдущем пункте варианта здесь на первом этапе с помощью большой языковой модели на основании множества сообщений формируются списки наиболее релевантных ключевых слов по каждой потенциальной теме.

На следующем этапе списки полученных ключевых слов используются для поиска содержащих их сообщений. В рамках данного исследования для извлечения таких сообщений использовался интеллектуальный поиск, который является встроенным средством в систему управления базами данных MongoDB. В данном случае очень полезным оказывается свойство интеллектуального поиска по нахождению заданных ключевых слов с учетом их словоформ.

На финальном этапе полученные на основе отбора по ключевым словам множества сообщений отдельными запросами обрабатываются с помощью большой языковой модели. Результатом такой обработки являются формулировки тем для каждой группы сообщений.

Таким образом, представленный выше конвейер обработки данных позволяет получить искомый набор тем и дополнительно список ключевых слов, с помощью которых можно найти сообщения, относящиеся к каждой теме.

Усложнение представленного варианта обработки данных введением в него промежуточной стадии компенсируется более стабильными результатами. Дополнительным свойством получаемых результатов является прослеживаемость (прозрачность), которая в большинстве случаев требуется в качестве обязательной характеристики используемого на практике решения. Также данный подход позволяет работать с исходными данными, объемы которых превышают максимальный размер запроса к языковой модели. В этом случае ключевые слова для тем должны быть получены отдельными запросами для каждого из подмножеств исходных данных. В итоге

полученные частичные данные объединяются и уже в таком виде обрабатываются дальше.

На рисунке 3 в виде UML диаграммы последовательности представлен порядок действий варианта реализации алгоритма получения обобщений через предварительную обработку соответствующих ключевых слов.

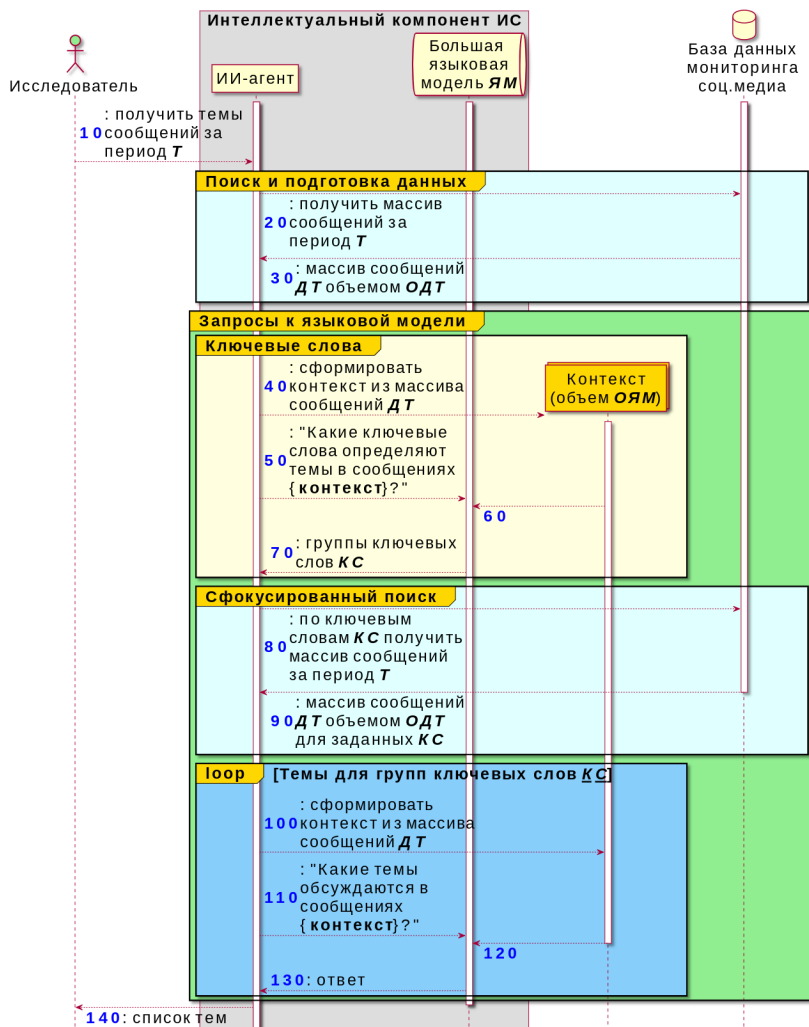


Рис. 2. Извлечение тем из списка сообщений на основе обработки ключевых слов

На рисунке 4 в виде UML диаграммы последовательности представлен вариативный этап работы конвейера, учитывающий ситуацию, когда объем исходных данных превышает максимальный размер контекста запроса к языковой модели.

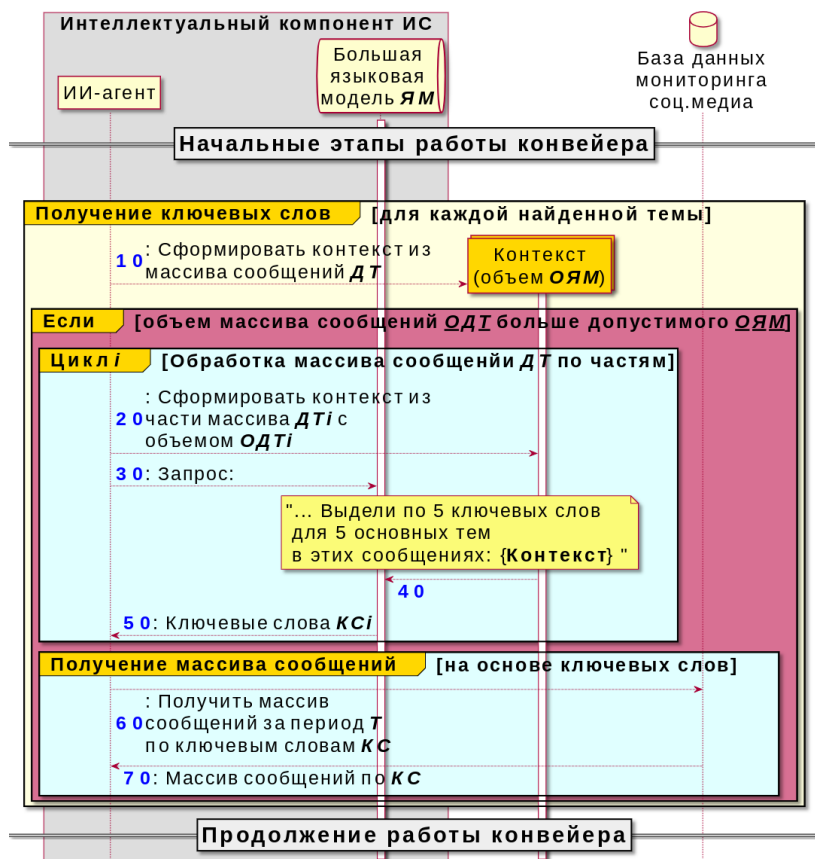


Рис. 3. Этап предварительного отбора ключевых сообщений

6.3. Векторизация и кластеризация. В данном варианте большая языковая модель главным образом используется в качестве инструмента для размещения текстов социальных медиа в своем векторном семантическом пространстве.

Целевые сообщения векторизуются и записываются в векторную базу данных. В данной работе используется векторная база данных Milvus.

Далее полученное множество векторов с помощью алгоритма HDBSCAN [27] разбивается на кластеры. Данный алгоритм кластеризации предлагается в качестве рабочего примера разработчиками базы Milvus. Однако для исследовательских целей в будущих работах планируется использовать различные алгоритмы кластеризации, т.к. в рамках решаемой задачи результат разбиения на кластеры играет большую роль.

Заключительным действием является генерация с помощью большой языковой модели обобщений (формулировок тем) для множества сообщений, входящих в каждый полученный кластер.

На рисунке 5 в виде UML диаграммы последовательности представлен порядок действий варианта реализации алгоритма получения тем через предварительную кластеризацию векторов исходных сообщений.

Особенностью данного варианта является то, что группировка сообщений по темам производится явно с помощью алгоритмов кластеризации. Это повышает точность и прослеживаемость результатов. Также преимуществом является то, что ограничение на количество анализируемых сообщений задается характеристиками векторной базы данных, а не максимальным количеством токенов, из которых формируется промпт для языковой модели. Фактически векторизации и последующей кластеризации можно подвергать всё множество сообщений из базы мониторинга, в то время как для формирования промпта количество таких сообщений будет значительно ограничено.

Основные компоненты предложенного метода, в виде фрагментов программного кода, зарегистрированы авторами в Роспатент [28] и доступны по ссылке https://git.iimm.ru/iimm-pub/2025-iimm-rid2-fed_dat_vish-soc_mon_kw_themes_gen.

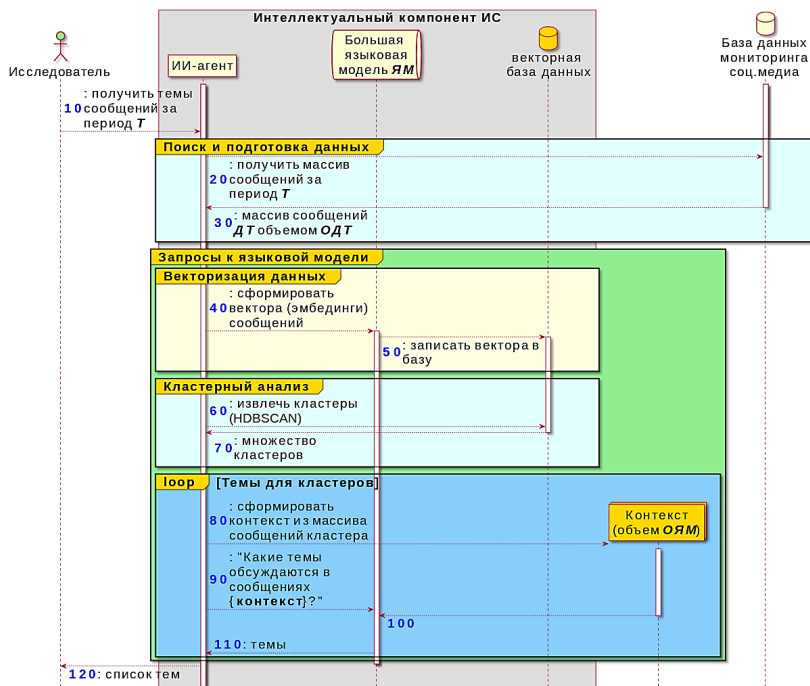


Рис. 5. Извлечение тем из списка сообщений с помощью кластеризации векторов

7. Комплекс метрик для оценки конвейеров метода

7.1. Стабильная воспроизводимость результата. Точность на

а) синтаксическом б) семантическом уровне. Уровень стабильности определяется количеством повторений результата при повторной генерации при одинаковых начальных условиях. Более строгий синтаксический способ предполагает точное синтаксическое равенство получаемых результатов. Семантический вариант предполагает смысловое равенство результатов.

7.2. Прослеживаемость результата.

а) синтаксическим б) интеллектуальным (например, MongoDB) поиском. Проверка

Факт прослеживаемости результата определяется возможностью на его основе найти в исходном наборе данных элементы, благодаря которым результат получен.

В данном случае хорошим примером является получение в качестве результата множества ключевых слов, поиск которых в

исходных данных приведет к множеству текстов, являющихся основой полученного результата. Такого рода операция нужна для проверки достоверности полученных результатов.

7.3. Объем обрабатываемых (исходных) данных.

Ограничения на объем а) контекста языковой модели или б) хранилища данных для «обогащения» (дополнения) запроса.

Классическое использование технологии RAG накладывает ограничение на объем данных, на основе которого формируется результат. Ограничение зависит от измеряемого в токенах максимального размера текста запроса, в который включаются и управляющие инструкции и непосредственно сами данные. В режиме конвейеризации можно делать несколько таких запросов и тем самым каждый раз иметь возможность использовать запросы максимального объема. Однако, если использовать большую языковую модель в качестве векторизатора, то можно работать одновременно со всем объемом данных. Векторные хранилища имеют достаточно большой объем и позволяют разместить все необходимые для работы данные. Некоторые задачи успешно решаются с помощью непосредственной обработки векторов. Например, одним из этапов конвейера обработки данных может быть кластеризация, которую как раз и можно решать, напрямую обращаясь к векторному хранилищу.

8. Прогнозирование результатов экспериментов.

Потенциальные результаты работы представленных конвейеров соответствуют особенностям их построения. Ожидаемые результаты запланированных экспериментов представлены в таблице 1, а ниже дано соответствующее обоснование сделанных прогнозов.

Таблица 1. Варианты реализации метода и соответствующие показатели

Вариант реализации метода	Классический RAG LLM	Конвейерный RAG LLM	Кластерный анализ векторов LLM
Метрика			
«Воспроизводимость» (Стабильность)	-	-+	++
«Прослеживаемость»	-	++	++
«Объем данных»	-	-+	++

Одношаговый вариант конвейера с комплексным запросом к большой языковой модели позволяет сразу получить и список тем, и список соответствующих ключевых слов. Генеративный характер работы модели предопределяет синтаксическую вариативность результата и весьма вероятные эффекты проявления «конфабуляции».

Смысловая составляющая таких результатов сохраняется. Однако поиск подтверждений в исходных данных на основе ключевых слов не дает гарантированного результата.

Конвейер с предварительной генерацией ключевых слов и последующим поиском соответствующих им исходных данных имеет большую стабильность. Это связано с тем, что запрос к языковой модели о получении генеративных обобщений формируется на основании отобранных по ключевым словам сообщений. Сами же ключевые слова проходят дополнительную проверку на присутствие в сообщениях, что гарантирует возможность использования их для нахождения исходных сообщений.

Третий вариант также основывается на предварительном этапе, в рамках которого в векторном пространстве выделяются кластеры сообщений. Каждый запрос к языковой модели ограничен только относящимися к определенному кластеру сообщениями, что значительно повышает стабильность получаемого результата. В этом случае начальная кластеризация гарантирует возможность однозначного сопоставления результата обобщения и соответствующих исходных данных.

Таким образом, более качественными вариантами реализации конвейеров генеративных обобщений можно считать те, в которых упрощается структура запроса при обращении к языковой модели, а на отдельном этапе обеспечивается сохранение признаков, позволяющих найти необходимые для контроля прослеживаемости исходные данные.

9. Результаты эксперимента. На основе представленных трех конвейеров проведена серия экспериментов, результаты которых оценены с помощью трех метрик, приведенных выше.

В экспериментах использованы два набора данных, сформированных на основе реальных данных из системы мониторинга домовых чатов:

- Набор данных №1 (Дс1): 280 сообщений, 18509 символов.
 - Набор данных №2 (Дс2): 911 сообщений, 59211 символов.
- Для запросов и векторизации использована языковая модель:
- bambucha/saiga-llama3:latest, температура модели 0.1.

Суть эксперимента состояла в том, чтобы на каждом наборе данных несколько раз проверить работу каждого конвейера и оценить полученные результаты. Каждый конвейер на каждом наборе данных отработал десять раз. Обобщенные результаты приведены в таблице 2.

Оценка по метрике «Объем данных» определяется характеристиками специально подобранных наборов данных. Первый

набор сформирован на основе данных суточного мониторинга домашних чатов. Объем сообщений позволяет загрузить их в контекст одного запроса к языковой модели. Таким образом становится возможна работа одноэтапного конвейера, в рамках которой все необходимые сведения о данных получаются в результате единого запроса. Второй набор данных сформирован на основе более продолжительного мониторингового периода. В этом случае суммарный объем текстов сообщений превышает ограничения на размер контекста запроса к языковой модели. В результате для обработки таких данных либо требуется предварительное их уплотнение, либо применение соответствующих таким условиям технологий. Для этих целей как раз подходят второй и третий вид описываемых в данной работе конвейеров. В таблице 2 отражены результаты экспериментов со всеми вариантами конвейеров в отношении первого (Дс1) и второго (Дс2) наборов данных.

Оценки по метрике «Воспроизводимость (стабильность)» сделаны с помощью экспертов. В результате десяти запусков конвейера для одного и того же набора данных получены десять наборов по пять обобщающих формулировок тем в каждом. Сформирована сравнительная таблица, в которую полученные наборы тем записаны и по строкам, и по столбцам, а на их пересечении (кроме диагональных элементов) эксперты выставили оценку соответствия. Для данной работы эксперты оценивали смысловую эквивалентность тематических наборов. Пара тематических наборов полностью (на 100%) соответствует друг другу, когда для каждой темы из одного набора найдена эквивалентная тема из сравниваемого набора. Если для темы не нашлось эквивалентной, то оценка соответствия пропорционально уменьшается (на $1/5$, т.е. на 20%). В итоге, среднее значение по сравнительным таблицам всех экспертов дает искомую оценку по метрике «Воспроизводимость».

На первом наборе данных лучшие результаты продемонстрировал второй вариант. Это связано с тем, что обработка каждой темы в отдельном запросе дает более стабильные формулировки. Противоположный результат продемонстрировал первый вариант, так как использование одного запроса сразу для всех тем повышает уровень вариативности их формулировок. Третий вариант продемонстрировал самый плохой результат. Это связано с тем, что используемый набор данных не имеет достаточного объема, для того чтобы их качественно сформировать и обработать на кластерном уровне. В то же время, на более объемном наборе данных третий вариант продемонстрировал наилучший результат.

Оценки по метрике «Прослеживаемость» получены автоматически. Алгоритм расчета формально соответствует классической метрике из машинного обучения Precision (точность), которая рассчитывается как отношение количества правильно классифицированных положительных объектов (True Positives) к общему числу положительных предсказаний модели (True Positives + False Positives). В описываемых экспериментах положительными исходами стали полученные для темы ключевые слова, поиск по которым в наборе данных дал ненулевые результаты. Как и предполагалось до эксперимента, некоторая часть сгенерированных языковой моделью ключевых слов, возможно, по смыслу отражает суть темы, но по факту не позволила найти в наборе данных ни одного содержащего их сообщения. Оценка по данной метрике произведена как с помощью оригинальных ключевых слов, так и с помощью их нормализованных (стемминг) вариантов. В таблице 2 данная метрика имеет два соответствующих представления. Результаты экспериментов для всех рассматриваемых конвейеров показали в среднем примерно одинаковые результаты. Однако для наиболее реального на практике набора данных (Дс2) свое преимущество продемонстрировал третий вариант. Это подтверждает тот факт, что использование генеративных возможностей языковых моделей вносит дополнительную вариативность в получаемые результаты. С другой стороны, использование языковых моделей для формирования векторных представлений требует дополнительной специфической программной обработки результатов. Золотой серединой является конвейерное использование языковых моделей с подходящими вариантами предварительного уплотнения, фильтрации и фрагментирования исходных данных, включаемых в контекст запроса.

Таблица 2. Оценки результатов эксперимента

Вариант реализации метода	Классический RAG LLM		Конвейерный RAG LLM		Кластерный анализ векторов LLM	
	Дс1	Дс2	Дс1	Дс2	Дс1	Дс2
Метрика	Дс1	Дс2	Дс1	Дс2	Дс1	Дс2
«Воспроизводимость» (стабильность)	60%	-	80%	62%	66%	100%
«Прослеживаемость»	68%	-	68%	63%	65%	89%
«Прослеживаемость» (стемминг)	81%	-	81%	70%	70%	94%
Объем данных	+	-	+	+	+	+

9. Заключение. Предложен и опробован на практике гибридный метод интеграции больших языковых моделей в алгоритмы фокусированного мониторинга открытых данных социальных медиа. В качестве примера практического применения метода реализованы три разнотипных конвейера для получения генеративных обобщений на основе множества сообщений социальных медиа. Каждый конвейер включает в себя этап обращения к большой языковой модели. С помощью сравнительного анализа показано, что особенности реализации конвейеров определяют качество получаемого с помощью них результата. В данном исследовании качество определено воспроизводимостью результатов, их прослеживаемостью, а также допустимым объемом исходных данных.

Литература

1. Датьев И.О., Федоров А.М., Ревякин А.А. Фокусированный сбор и обработка открытых данных социальных медиа // *Онтология проектирования*. 2024. Т. 14. № 4(54). С. 569–581. DOI: 10.18287/2223-9537-2024-14-4-569-581.
2. Blei D.M., Ng A.Y., Jordan M.I. Latent Dirichlet Allocation // *Journal of Machine Learning Research*. 2003. vol. 3. pp. 993–1022.
3. Федоров А.М., Датьев И.О., Вишняков И.Г. Проектирование информационной системы комплексного тематического анализа больших данных социальных медиа // *Онтология проектирования*. 2024. Т. 14. № 1(51). С. 55–70. DOI: 10.18287/2223-9537-2024-14-1-55-70.
4. Zhang Y., Jin H., Meng D., Wang J., Tan J. A Comprehensive Survey on Process-Oriented Automatic Text Summarization with Exploration of LLM-Based Methods. *arXiv preprint arXiv:2403.02901*. 2024.
5. Glickman M., Zhang Y. AI and Generative AI for Research Discovery and Summarization // *Harvard Data Science Review*. 2024. vol. 6(2). DOI: 10.1162/99608f92.7f9220ff.
6. Rose S., Engel D., Cramer N., Cowley W. Automatic Keyword Extraction from Individual Documents // *Text mining: applications and theory*. 2010. pp. 1–20. DOI: 10.1002/9780470689646.ch1.
7. Campos R., Mangaravite V., Pasquali A., Jorge A., Nunes C., Jatowt A. YAKE! Keyword extraction from single documents using multiple local features // *Information Sciences*. 2020. vol. 509. pp. 257–289. DOI: 10.1016/j.ins.2019.09.013.
8. Mihalcea R., Tarau P. TextRank: Bringing Order into Text // *Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing*. Barcelona, Spain: Association for Computational Linguistics, 2004. pp. 404–411.
9. Tabanmehr Z., Akhtarkavan E. Automatic summarization of Instagram social network posts by combining semantic and statistical approaches // *6th International Conference on Pattern Recognition and Image Analysis (IPRIA)*. 2023. pp. 1–6. DOI: 10.1109/IPRIA59240.2023.10147186.
10. Ghali M.K., Farrag A., Lam S., Won, D. Beyondwords is All You Need: Agentic Generative Ai Based Social Media Themes Extractor. URL: <https://ssrn.com/abstract=5155101> (дата обращения: 26.10.2025)
11. Wei J., Wang X., Schuurmans D., Bosma M., Ichter B., Xia F., Chi E., Le Q., Zhou D. Chain-of-thought prompting elicits reasoning in large language models // *Proceedings*

- of the 36th International Conference on Neural Information Processing Systems (NIPS'22). NY, USA: Curran Associates Inc., Red Hook, 2022. pp. 24824–24837.
12. Kojima T., Gu S.S., Reid M., Matsuo Y., Iwasawa Y. Large language models are zero-shot reasoners // In Proceedings of the 36th International Conference on Neural Information Processing Systems (NIPS '22). NY, USA: Curran Associates Inc., Red Hook, 2022. pp. 22199–22213.
13. Kambhampettu H., Flores J., Head A. Traceable Text: Deepening Reading of AI-Generated Summaries with Phrase-Level Provenance Links. arXiv preprint arXiv:2409.13099. 2024. DOI: 10.48550/arXiv.2409.13099.
14. Zhao S., Yang Y., Wang Z., He Z., Qiu L.K., Qiu L. Retrieval Augmented Generation (RAG) and Beyond: A Comprehensive Survey on How to Make your LLMs use External Data More Wisely. arXiv preprint arXiv:2409.14924. 2024. DOI: 10.48550/arXiv.2409.14924.
15. Hidayaturrahman, Prawira I. Retrieval-Augmented Generation for Social Media Content Creation with Sentence Window and Auto-Merging Retrieval. International Conference on Informatics, Multimedia, Cyber and Information System (ICIMCIS). 2024. pp. 608–613. DOI: 10.1109/ICIMCIS63449.2024.10956838.
16. Zhang W., Zhang J. Hallucination Mitigation for Retrieval-Augmented Large Language Models: A Review // Mathematics. 2025. vol. 13(5). DOI: 10.3390/math13050856.
17. Yu H.Q., McQuade F. RAG-KG-IL: A Multi-Agent Hybrid Framework for Reducing Hallucinations and Enhancing LLM Reasoning through RAG and Incremental Knowledge Graph Learning Integration // arXiv preprint arXiv:2503.13514. 2025.
18. Zhou Y., Liu Y., Li X., Jin J., Qian H., Liu Z., Li C., Dou Z., Ho T.-Y., Yu P. Trustworthiness in Retrieval-Augmented Generation Systems: A Survey. arXiv preprint arXiv:2409.10102. 2024. DOI: 10.48550/arXiv.2409.10102.
19. Das S., Ge Y., Guo Y., Rajwal S., Hairston J., Powell J., Walker D., Peddireddy S., Lakamana S., Bozkurt S., Reyna M., Sameni R., Xiao Y., Kim S., Chandler R., Hernandez N., Mowery D., Wightman R., Love J., Spadaro A., Perrone J., Sarker A. Two-Layer Retrieval-Augmented Generation Framework for Low-Resource Medical Question Answering Using Reddit Data: Proof-of-Concept Study // Journal of Medical Internet Research. 2025. vol. 27. DOI: 10.2196/66220.
20. Gupta S., Ranjan R., Singh S.N. A Comprehensive Survey of Retrieval-Augmented Generation (RAG): Evolution, Current Landscape and Future Directions. arXiv preprint arXiv:2410.12837. 2024. DOI: 10.48550/arXiv.2410.12837.
21. Sun D., Lyu Y., Li J., Chen Y., Wang T., Kimura T., Abdelzaher T. SCRAG: Social Computing-Based Retrieval Augmented Generation for Community Response Forecasting in Social Media Environments. arXiv preprint arXiv:2504.16947v1. 2025. DOI: 10.48550/arXiv.2504.16947.
22. Wu X.-K., Chen M., Li W., Wang R., Lu L., Liu J., Hwang K., Hao Y., Pan Y., Meng Q., et al. LLMFine-Tuning: Concepts, Opportunities, and Challenges // Big Data Cogn. Comput. 2025. vol. 9. DOI: 10.3390/bdcc9040087.
23. Song Y., Lv C., Zhu K., et al. LoRA fine-tuning of Llama3 large model for intelligent fishery field // Discov Computing. 2025. vol. 28. DOI: 10.1007/s10791-025-09663-6.
24. Ramachandran A. Advancing Retrieval-Augmented Generation (RAG): Innovations, Challenges, and the Future of AI Reasoning. 2025.
25. Bsharat S.M., Myrzakhan A., Shen Z. Principled instructions are all you need for questioning llama-1/2, gpt-3.5/4 // arXiv preprint arXiv:2312.16171. 2023.
26. Hu Y., Lei Z., Zhang Z., Pan B., Ling C., Zhao L. GRAG: Graph Retrieval-Augmented Generation // Findings of the Association for Computational Linguistics: NAACL 2025. 2025. pp. 4145–4157. DOI: 10.18653/v1/2025.findings-naacl.232.

27. Campello R.J.G.B., Moulavi D., Sander J. Density-Based Clustering Based on Hierarchical Density Estimates // *Advances in Knowledge Discovery and Data Mining (PAKDD 2013). Lecture Notes in Computer Science*. 2013. vol. 7819. pp. 160–172. DOI: 10.1007/978-3-642-37456-2_14.
28. Федоров А.М., Датьев И.О., Вишняков И.Г. Модуль системы мониторинга социальных медиа: извлечение ключевых слов и генерация названий тем // Роспатент: Свидетельство о государственной регистрации программы для ЭВМ № 2025668928 от 21 июля 2025 г.

Федоров Андрей Михайлович — канд. техн. наук, ведущий научный сотрудник, заместитель директора по научной работе, ИИММ КНЦ РАН; доцент кафедры, кафедра информатики и вычислительной техники, филиал МАУ в г. Апатиты. Область научных интересов: разработка моделей и технологий информационной поддержки для регионального управления, мониторинг и анализ данных социальных медиа. Число научных публикаций — 60. a.fedorov@ksc.ru; улица Ферсмана, 24а, 184209, Апатиты, Мурманская область, Россия; р.т.: +7(81555)740-50.

Датьев Игорь Олегович — канд. техн. наук, старший научный сотрудник, ученый секретарь, ИИММ КНЦ РАН. Область научных интересов: разработка моделей и технологий для обработки больших данных, мониторинг и анализ данных социальных медиа. Число научных публикаций — 87. i.datyev@ksc.ru; улица Ферсмана, 24а, 184209, Апатиты, Мурманская область, Россия; р.т.: +7(81555)740-50.

Вишняков Иван Геннадьевич — аспирант, системный администратор, ИИММ КНЦ РАН. Область научных интересов: разработка информационных систем анализа больших открытых данных социальных медиа. Число научных публикаций — 6. i.vishnyakov@ksc.ru; улица Ферсмана, 24а, 184209, Апатиты, Мурманская область, Россия; р.т.: +7(81555)740-50.

Поддержка исследований. Исследование выполнено в рамках государственного задания ИИММ КНЦ РАН Министерства науки и высшего образования РФ, тема НИР: «Методы и технологии создания интеллектуальных информационных систем для поддержки развития сложных динамических систем с региональной спецификой в условиях неопределённости и риска» (шифр темы FMEZ-2025-0053).

A. FEDOROV, I. DATYEV, I. VISHNYAKOV
**THE METHOD FOR INTEGRATING LARGE LANGUAGE
MODELS INTO ALGORITHMS FOR FOCUSED MONITORING
OF OPEN SOCIAL MEDIA DATA**

Fedorov A., Datyev I., Vishnyakov I. The Method for Integrating Large Language Models into Algorithms for Focused Monitoring of Open Social Media Data.

Abstract. The relevance of the study is determined by the importance and complexity of performing rapid summarization of a vast array of user-generated content on social networks. It is proposed to reduce the complexity of the problem by using robotic algorithms and their automated intelligent focusing on specific platforms, data availability, and data volumes. The paper examines the ability of large language models (LLMs) to generate high-quality, coherent, and context-sensitive annotations (summaries) that are suitable for the dynamic nature of unstructured, noisy social network data. The features of the RAG LLM technology for summarizing social network publications are presented. The main disadvantage of language models is instability and the difficulty of tracking the results to confirm factual accuracy. The authors propose a hybrid method for summarizing social media posts over a given period of time. The method involves a complex and variable combination of classical methods for extracting data from their repositories, as well as the abstractive and generative capabilities of large language models. Large language models are used to vectorize the analyzed data. The application of clustering algorithms to the obtained vector representations made it possible to increase the stability and quality of the results. Within the RAG technology, the capabilities of large language models are expanded by means of intelligent search in the MongoDB database used to store the original data. The paper presents three pipelines, each of which is a variant of the method implementation, and has advantages and disadvantages in various application conditions. The metrics used to evaluate the pipelines are given, and a comparative analysis is performed. Overall, the method allows us to reduce the confabulations of a large language model and obtain annotations of publications for different time periods in real-time. The proposed method is used in practice in the open social media data monitoring system developed by the authors.

Keywords: social media, posts, text summarization, LLMs, RAG, AI agents, hybrid method.

References

1. Datyev I.O., Fedorov A.M., Reviakin A.A. [Focused collection and processing of open social media data]. *Ontologija proektirovaniya – Ontology of designing*. 2024. vol. 14. no. 4(54). pp. 569–581. DOI: 10.18287/2223-9537-2024-14-4-569-581. (In Russ.).
2. Blei D.M., Ng A.Y., Jordan M.I. Latent Dirichlet Allocation. *Journal of Machine Learning Research*. 2003. vol. 3. pp. 993–1022.
3. Fedorov A.M., Datyev I.O., Vishnyakov I.G. [Designing an information system for integrated topic analysis of social media big data]. *Ontologija proektirovaniya – Ontology of designing*. 2024. vol. 14. no. 1(51). pp. 55–70. DOI: 10.18287/2223-9537-2024-14-1-55-70. (In Russ.).
4. Zhang Y., Jin H., Meng D., Wang J., Tan J. A Comprehensive Survey on Process-Oriented Automatic Text Summarization with Exploration of LLM-Based Methods. *arXiv preprint arXiv:2403.02901*. 2024.

5. Glickman M., Zhang Y. AI and Generative AI for Research Discovery and Summarization. *Harvard Data Science Review*. 2024. vol. 6(2). DOI: 10.1162/99608f92.7f9220ff.
6. Rose S., Engel D., Cramer N., Cowley W. Automatic Keyword Extraction from Individual Documents. *Text mining: applications and theory*. 2010. pp. 1–20. DOI: 10.1002/9780470689646.ch1.
7. Campos R., Mangaravite V., Pasquali A., Jorge A., Nunes C., Jatowt A. YAKE! Keyword extraction from single documents using multiple local features. *Information Sciences*. 2020. vol. 509. pp. 257–289. DOI: 10.1016/j.ins.2019.09.013.
8. Mihalcea R., Tarau P. TextRank: Bringing Order into Text. *Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing*. Barcelona, Spain: Association for Computational Linguistics, 2004. pp. 404–411.
9. Tabanmehr Z., Akhtarkavan E. Automatic summarization of Instagram social network posts by combining semantic and statistical approaches. *6th International Conference on Pattern Recognition and Image Analysis (IPRIA)*. 2023. pp. 1–6. DOI: 10.1109/IPRIA59240.2023.10147186.
10. Ghali M.K., Farrag A., Lam S., Won, D. Beyondwords is All You Need: Agentic Generative Ai Based Social Media Themes Extractor. Available at: <https://ssrn.com/abstract=5155101> (accessed 26.10.2025)
11. Wei J., Wang X., Schuurmans D., Bosma M., Ichter B., Xia F., Chi E., Le Q., Zhou D. Chain-of-thought prompting elicits reasoning in large language models. *Proceedings of the 36th International Conference on Neural Information Processing Systems (NIPS'22)*. NY, USA: Curran Associates Inc., Red Hook, 2022. pp. 24824–24837.
12. Kojima T., Gu S.S., Reid M., Matsuo Y., Iwasawa Y. Large language models are zero-shot reasoners. In *Proceedings of the 36th International Conference on Neural Information Processing Systems (NIPS '22)*. NY, USA: Curran Associates Inc., Red Hook, 2022. pp. 22199–22213.
13. Kambhamettu H., Flores J., Head A. Traceable Text: Deepening Reading of AI-Generated Summaries with Phrase-Level Provenance Links. *arXiv preprint arXiv:2409.13099*. 2024. DOI: 10.48550/arXiv.2409.13099.
14. Zhao S., Yang Y., Wang Z., He Z., Qiu L.K., Qiu L. Retrieval Augmented Generation (RAG) and Beyond: A Comprehensive Survey on How to Make your LLMs use External Data More Wisely. *arXiv preprint arXiv:2409.14924*. 2024. DOI: 10.48550/arXiv.2409.14924.
15. Hidayaturrahman, Prawira I. Retrieval-Augmented Generation for Social Media Content Creation with Sentence Window and Auto-Merging Retrieval. *International Conference on Informatics, Multimedia, Cyber and Information System (ICIMCIS)*. 2024. pp. 608–613. DOI: 10.1109/ICIMCIS63449.2024.10956838.
16. Zhang W., Zhang J. Hallucination Mitigation for Retrieval-Augmented Large Language Models: A Review. *Mathematics*. 2025. vol. 13(5). DOI: 10.3390/math13050856.
17. Yu H.Q., McQuade F. RAG-KG-IL: A Multi-Agent Hybrid Framework for Reducing Hallucinations and Enhancing LLM Reasoning through RAG and Incremental Knowledge Graph Learning Integration. *arXiv preprint arXiv:2503.13514*. 2025.
18. Zhou Y., Liu Y., Li X., Jin J., Qian H., Liu Z., Li C., Dou Z., Ho T.-Y., Yu P. Trustworthiness in Retrieval-Augmented Generation Systems: A Survey. *arXiv preprint arXiv:2409.10102*. 2024. DOI: 10.48550/arXiv.2409.10102.
19. Das S., Ge Y., Guo Y., Rajwal S., Hairston J., Powell J., Walker D., Peddireddy S., Lakamana S., Bozkurt S., Reyna M., Sameni R., Xiao Y., Kim S., Chandler R., Hernandez N., Mowery D., Wightman R., Love J., Spadaro A., Perrone J., Sarker A. Two-Layer Retrieval-Augmented Generation Framework for Low-Resource Medical

- Question Answering Using Reddit Data: Proof-of-Concept Study. *Journal of Medical Internet Research*. 2025. vol. 27. DOI: 10.2196/66220.
20. Gupta S., Ranjan R., Singh S.N. A Comprehensive Survey of Retrieval-Augmented Generation (RAG): Evolution, Current Landscape and Future Directions. *arXiv preprint arXiv:2410.12837*. 2024. DOI: 10.48550/arXiv.2410.12837.
21. Sun D., Lyu Y., Li J., Chen Y., Wang T., Kimura T., Abdelzaher T. SCRAP: Social Computing-Based Retrieval Augmented Generation for Community Response Forecasting in Social Media Environments. *arXiv preprint arXiv:2504.16947v1*. 2025. DOI: 10.48550/arXiv.2504.16947.
22. Wu X.-K., Chen M., Li W., Wang R., Lu L., Liu J., Hwang K., Hao Y., Pan Y., Meng Q., et al. LLMFine-Tuning: Concepts, Opportunities, and Challenges. *Big Data Cogn. Comput.* 2025. vol. 9. DOI: 10.3390/bdcc9040087.
23. Song Y., Lv C., Zhu K., et al. LoRA fine-tuning of Llama3 large model for intelligent fishery field. *Discov Computing*. 2025. vol. 28. DOI: 10.1007/s10791-025-09663-6.
24. Ramachandran A. Advancing Retrieval-Augmented Generation (RAG): Innovations, Challenges, and the Future of AI Reasoning. 2025.
25. Bsharat S.M., Myrzakhan A., Shen Z. Principled instructions are all you need for questioning llama-1/2, gpt-3.5/4. *arXiv preprint arXiv:2312.16171*. 2023.
26. Hu Y., Lei Z., Zhang Z., Pan B., Ling C., Zhao L. GRAG: Graph Retrieval-Augmented Generation. *Findings of the Association for Computational Linguistics: NAACL 2025*. 2025. pp. 4145–4157. DOI: 10.18653/v1/2025.findings-naacl.232.
27. Campello R.J.G.B., Moulavi D., Sander J. Density-Based Clustering Based on Hierarchical Density Estimates. *Advances in Knowledge Discovery and Data Mining (PAKDD 2013). Lecture Notes in Computer Science*. 2013. vol. 7819. pp. 160–172. DOI: 10.1007/978-3-642-37456-2_14.
28. Fedorov A.M., Datyev I.O., Vishnyakov I.G. [Social media monitoring system module: extracting keywords and generating topic titles]. *Rospatent: Certificate of state registration of a computer program No. 2025668928 dated July 21, 2025. (In Russ.)*.

Fedorov Andrey — Ph.D., Leading researcher, deputy director for research, IIMM KSC RAS; Associate professor of the department, Department of informatics and computer engineering, Apatity branch of MAU. Research interests: development of models and technologies for information support for regional management, monitoring and analysis of social media data. The number of publications — 60. a.fedorov@ksc.ru; 24a, Fersmana St., 184209, Apatity, Murmansk region, Russia; office phone: +7(81555)740-50.

Datyev Igor — Ph.D., Senior researcher, scientific secretary, IIMM KSC RAS. Research interests: development of models and technologies for big data processing, monitoring and analysis of social media data. The number of publications — 87. i.datyev@ksc.ru; 24a, Fersmana St., 184209, Apatity, Murmansk region, Russia; office phone: +7(81555)740-50.

Vishnyakov Ivan — Postgraduate student, system administrator, IIMM KSC RAS. Research interests: development of information systems for analyzing big open data in social media. The number of publications — 6. i.vishnyakov@ksc.ru; 24a, Fersmana St., 184209, Apatity, Murmansk region, Russia; office phone: +7(81555)740-50.

Acknowledgements. The work is supported by the Ministry of Science and Higher Education of the Russian Federation. Topic title: Methods and technologies for creating intelligent information systems to support the development of complex dynamic systems with regional specificity in conditions of uncertainty and risk (No. FMEZ-2025-0053).

В.Ю. Осипов

УСТОЙЧИВОСТЬ ИНТЕЛЛЕКТУАЛЬНОЙ ОБРАБОТКИ СИГНАЛОВ ПОТОКОВЫМИ РЕКУРРЕНТНЫМИ НЕЙРОННЫМИ СЕТЯМИ С НЕПРЕРЫВНЫМ ОБУЧЕНИЕМ

Осипов В.Ю. Устойчивость интеллектуальной обработки сигналов потоковыми рекуррентными нейронными сетями с непрерывным обучением.

Аннотация. Интеллектуальная нейросетевая обработка различных сигналов с непрерывным обучением представляет большой как научный, так и практический интерес. Для такой обработки наряду с другими решениями применяют потоковые импульсные рекуррентные нейронные сети (РНС) с расширенными функциональными возможностями. Однако для этих РНС во многом не исследованы вопросы обеспечения их устойчивого функционирования. Преследуется цель повышения устойчивости потоковых импульсных РНС за счет разработки новых методов интеллектуальной обработки сигналов с непрерывным обучением. В интересах этого уточняются возможности этих РНС и анализируются подходы к обеспечению их устойчивости при обучении и генерации сигналов. Осуществлена формализация потоковой РНС в виде относительно конечного операционного автомата. Предложен новый метод устойчивой интеллектуальной обработки сигналов усовершенствованной РНС с непрерывным обучением. Рассмотрены схемы решения различных интеллектуальных задач анализа и синтеза предлагаемым методом. Приведены результаты моделирования, подтверждающие работоспособность предложенного метода и возможности повышения устойчивости РНС при непрерывном обучении и генерировании сигналов. Показана достижимость устойчивого баланса между запоминаемой и постепенно забываемой информацией в РНС с оперативной адаптацией к изменяющимся внешним условиям.

Ключевые слова: нейронная сеть, интеллектуальность, устойчивость, непрерывное обучение, генерация сигналов.

1. Введение. Успешность решения интеллектуальных задач потоковыми импульсными рекуррентными нейронными сетями (РНС) во многом зависит от устойчивости их функционирования. В качестве таких задач для РНС могут выступать распознавание и прогнозирование различных событий, генерация текстов и видеoinформации, нейруправление трудно формализуемыми процессами, фильтрация сигналов и другие [1]. Творческие задачи должны реализовываться РНС в общем случае в условиях изменения законов поведения воспринимаемых потоков сигналов.

Для оперативной адаптации потоковых РНС [2, 3] и других нейронных сетей [4 – 6] к изменяющимся условиям предусматривается непрерывное обучение, со своей спецификой их функционирования. При таком обучении с ограниченным объемом памяти РНС должны не только запоминать входные сигналы и результаты обработки, но и забывать устаревшую информацию. Должен соблюдаться некоторый баланс между запоминанием и забыванием информации. Обучение

решению одной творческой задачи не должно приводить к катастрофическому забыванию информации о другой задаче. Потокосые РНС не должны перегружаться входными сигналами. Следует исключать перевозбуждение этих РНС. Число одновременно возбужденных нейронов и текущий суммарный вес (проводимостей) синапсов РНС должны находиться в допустимых пределах.

Для соблюдения этих требований необходимо учитывать специфику потокосых импульсных РНС с непрерывным обучением, а также иметь соответствующие методы обеспечения устойчивости таких сетей.

Несмотря на наличие в составе потокосых импульсных РНС блока управления, позволяющего в некоторой мере повысить их устойчивость, проблема, связанная с нею, остается. Заметим, что эта проблема характерна не только для потокосых РНС, но и других нейронных сетей с непрерывным обучением.

Известны решения, касающиеся общих подходов к обеспечению устойчивости РНС [7–9]. Предложен ряд критериев, моделей и методов для реализации асимптотической и равномерной устойчивости по Ляпунову [10–12]. С формальных позиций эта устойчивость зависит от характеристик входных потоков сигналов, функций переходов РНС и ее элементов из одних состояний в другие, а также характеристик состояний сети (содержания ее памяти).

Если с требованиями к устойчивости РНС относительно прозрачная ситуация, то с конкретными методами их соблюдения много открытых вопросов. Они касаются методов устойчивого нейросетевого запоминания полезной информации и забывания устаревших данных, вызова из памяти результатов обработки, исключения перевозбуждения РНС и других аспектов. Необходима разработка новых методов устойчивой интеллектуальной обработки сигналов потокосыми РНС с непрерывным обучением.

В интересах решения этой задачи анализируются особенности потокосых РНС с непрерывным обучением и известные методы обеспечения устойчивости. Рассматривается обобщенная модель потокосой РНС в виде относительно конечного автомата. Предлагается новый метод устойчивой интеллектуальной обработки сигналов усовершенствованной потокосой РНС с непрерывным обучением и генерацией связанных сигналов. Раскрываются особенности решения интеллектуальных задач предлагаемым методом. Приводятся результаты моделирования. Формулируются практические рекомендации по обеспечению устойчивости потокосых РНС.

2. Особенности потоковых РНС и известных методов обеспечения устойчивости. Еще до появления нейронных сетей – трансформеров [13], составляющих основу современных больших языковых моделей ChatGPT, Deep Seek и подобных, в [2] показаны возможности генерации потоковыми РНС последовательностей слов и других сигналов с продвижением их по сети без искажений. При этом обращено внимание на необходимость обеспечения устойчивого функционирования этих РНС, как при непрерывном обучении, так и генерации сигналов с запоминанием результатов. Для этого предусматривался поиск методов, обеспечивающих баланс между возбуждающими и тормозящими воздействиями в потоковых РНС.

Особенностями таких РНС выступают следующие. На входы потоковых РНС подаются последовательные совокупности единичных импульсов (СЕИ), несущие всю информацию об исходных сигналах. Продвигаясь по РНС СЕИ связываются друг с другом, а результаты связывания запоминаются на элементах сети. В РНС формируется постоянно обновляемая модель воспринимаемых потоков сигналов, которая используется для решения творческих задач.

В базовом варианте это двухслойные РНС (рисунок 1) с блоком управления ассоциативным взаимодействием обрабатываемых сигналов в зависимости от текущих состояний слоев.

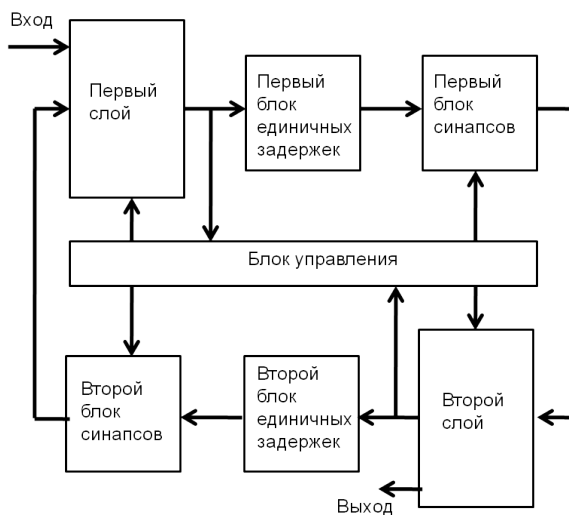


Рис. 1. Обобщенная структура потоковой импульсной рекуррентной нейронной сети

В этой РНС каждый нейрон одного слоя в общем случае связан синапсами со всеми нейронами другого слоя. Нейроны одного и того же слоя связей не имеют. Время задержки единичных импульсов в образуемых двухслойных контурах РНС меньше времени невосприимчивости нейронов после возбуждения.

За счет реализуемых пространственных сдвигов СЕИ, передаваемых от слоя к слою, РНС могут наделяться различными (линейными, спиральными, петлевыми и другими) логическими структурами. При этом слои РНС разделяются на одинаковые логические поля, а обрабатываемые СЕИ продвигаются вдоль слоев. Однозначное соответствие между входным и выходным языками РНС обеспечивается за счет приоритетности сильных связей между нейронами.

За счет блока управления может осуществляться вызов данных из памяти РНС в различных направлениях, а также регулироваться загрузка сети СЕИ. Такими РНС могут одновременно обрабатываться сигналы различных модальностей.

Для решения творческих задач с непрерывным обучением могут применяться многоуровневые потоковые импульсные РНС. В таких РНС непрерывное обучение реализуется прежде всего на нижних уровнях, а решение творческих задач осуществляется на высоких уровнях. Также для этого могут использоваться пары идентичных потоковых РНС [14]. В этом случае первая РНС функционирует в режиме непрерывного обучения, а на второй сети решаются творческие задачи на основе информации, периодически считываемой с первой РНС.

Без непрерывного обучения потоковых РНС невозможно оперативно адаптироваться к изменяющимся внешним условиям, что, несомненно, приводит к снижению качества выдаваемых результатов. Однако непрерывное обучение РНС с ограниченными возможностями памяти должно предусматривать не только запоминание новых сигналов, но и забывание ненужной информации. При этом в РНС не должен проявляться эффект катастрофического забывания ранее полученной информации при решении новых задач. Для разрешения этого противоречия при обучении последовательными данными применяют подходы на основе регуляризации, динамических архитектур и дополнительной памяти [8]. Подходы регуляризации предусматривают ограничения на обновление определенных параметров РНС. К ним относятся эластичная консолидация веса [15], синаптический интеллект [16], использование слоев расширения [17], перекрывающихся и ортогональных представлений [18]. Для

преодоления катастрофического забывания применяют контекстно-зависимые случайные подсети с упорядочиванием изменения весов [8]. Подходы динамической архитектуры предусматривают добавление нейронных ресурсов [19]. Методы, связанные с дополнительной памятью, основаны на использовании вторичных сетей для генерации шаблонов для повторения ранее изученной информации, а также внешней памяти, увеличивающейся при поступлении новой информации [8]. Также ориентируются на применимость гиперсетей, позволяющих запоминать информацию с ее сжатием. Известен также ряд работ, посвященных исследованию асимптотической устойчивости различных по структуре РНС [11, 20 – 22]. В нашем случае интерес представляют, прежде всего, методы обеспечения равномерной устойчивости РНС с непрерывным обучением.

Обычно для динамических систем устойчивость понимают в терминах критерия типа «ограниченный вход – ограниченный выход» (bounded input – bounded output stability criterion – BIBO) [7]. Согласно этому критерию выход системы не должен неограниченно возрастать в результате подачи ограниченного входного сигнала [23]. Заметим, что BIBO устойчивость – это конкретный вид робастности, говорящий об устойчивости системы при ограниченных входных сигналах. Критерий хорошо подходит для линейных динамических систем. Считают [7], что критерий BIBO не целесообразно применять к нейронным сетям, так как подобные нелинейные динамические системы являются BIBO устойчивыми, поскольку в конструкцию нейронов встроена нелинейность с насыщением. Однако этот критерий иногда может быть полезен для анализа устойчивости РНС с непрерывным обучением. Когда речь идет о нелинейных динамических системах, к которым относятся нейронные сети, то подразумевают устойчивость по Ляпунову [7]. При этом заметим, что BIBO устойчивость можно рассматривать как специальный случай устойчивости по Ляпунову.

В соответствии с теорией устойчивости Ляпунова [10] функционирование РНС можно считать равномерно устойчивым, когда для любого положительного числа ε можно указать положительное число δ , такое, что из неравенства $\|N(0) - N_0\| < \delta$ следует неравенство $\|N(t) - N_0\| < \varepsilon$ при $t > t_0$, где $\|\dots\|$ – Евклидова норма; N_0 – равновесное состояние нейросетевой системы; $N(0)$, $N(t)$ – состояния РНС на моменты времени $t_0 = 0$ и $t > t_0$. Заметим, что для каждого вида состояний РНС могут формироваться свои неравенства. Так в качестве $N(t)$ может вступать число нейронов $M(t)$, возбужденных в РНС на текущий момент времени. По отношению к синапсам РНС $N(t)$ представимо значением их суммарных весов $W_{\Sigma}(t)$. Таким образом,

при равномерно устойчивом функционировании РНС при выполнении неравенств $\|M(0)-M_0\|<\delta_1$, $\|W_\Sigma(0)-W_0\|<\delta_2$ должны выполняться условия: $\|M(t)-M_0\|<\varepsilon_1$ и $\|W_\Sigma(t)-W_0\|<\varepsilon_2$ при $t > t_0$.

Обеспечение выполнения этих условий применительно к потоковым РНС в некоторой мере выполнимо в рамках известных методов. Однако они не учитывают специфику этих РНС. В том числе известные подходы регуляризации должны учитывать особенности функционирования потоковых РНС.

3. Потоковая РНС как относительно конечный операционный автомат. В интересах поиска путей расширения возможностей устойчивой интеллектуальной обработки сигналов импульсными потоковыми РНС, рассмотрим их обобщенную модель.

При формализации функционирования потоковой РНС в дискретном времени и дискретном пространстве состояний функцию переходов ее из одного внутреннего состояния в другое можно записать в виде

$$\mathbf{h}_{t+1} = F_t^h(\mathbf{h}_t, \mathbf{SSP}_t^a, \mathbf{SSP}_{t-1}^a, \mathbf{SSP}_{t-2}^a, \dots, \mathbf{SSP}_{t-k}^a), \quad (1)$$

а функцию выходов как

$$\mathbf{SSP}_{t+1}^b = F_t^S(\mathbf{h}_t, \mathbf{SSP}_t^a, \mathbf{SSP}_{t-1}^a, \mathbf{SSP}_{t-2}^a, \dots, \mathbf{SSP}_{t-k}^a), \quad (2)$$

где $\mathbf{h}_{t+1}$ – внутреннее состояние РНС на момент времени $t+1$, которое включает состояния $\mathbf{x}_{t+1}$ нейронов и состояния $\mathbf{w}_{t+1}$ синапсов; $\mathbf{SSP}_t^a, \mathbf{SSP}_{t-1}^a, \mathbf{SSP}_{t-2}^a, \dots, \mathbf{SSP}_{t-k}^a$ – состояния входа РНС в виде СЕИ (обозначенных как SSP) на моменты времени $t, t-1, t-2, \dots, t-k$; $\mathbf{SSP}_{t+1}^b$ – состояние выхода РНС.

Состояния $\mathbf{h}_t, \mathbf{SSP}_t^a, \mathbf{SSP}_t^b$ и функции F_t^h, F_t^S , характеризующие РНС на t -й момент времени, должны удовлетворять определенным условиям (3) – (7):

$$\mathbf{h}_t \in \mathbf{DH}(\mathbf{h}_{t-1}); \quad (3)$$

$$\mathbf{SSP}_t^a \in \mathbf{DA}(\mathbf{h}_{t-1}); \quad (4)$$

$$\mathbf{SSP}_t^b \in \mathbf{DB}(\mathbf{h}_{t-1}); \quad (5)$$

$$F_t^h \in \mathbf{FH}(\mathbf{h}_{t-1}); \quad (6)$$

$$F_t^S \in \mathbf{FS}(\mathbf{h}_{t-1}). \quad (7)$$

Согласно (3) внутреннее состояние РНС на t -й момент времени должно относиться к множеству $\mathbf{DH}(\mathbf{h}_{t-1})$ допустимых внутренних состояний для этого момента, определенных относительно $t-1$. Условие (4) показывает, что состояния входа РНС на t -е моменты времени, ограничиваются множеством $\mathbf{DA}(\mathbf{h}_{t-1})$. Условие (5) отражает ограниченность возможных состояний выходов РНС. Они должны принадлежать множеству $\mathbf{DB}(\mathbf{h}_{t-1})$ допустимых результатов относительно момента времени $t-1$. По условию (6) функция переходов F_t^h РНС на t -й момент должна относиться к множеству $\mathbf{FH}(\mathbf{h}_{t-1})$ допустимых функций. Множество $\mathbf{FH}(\mathbf{h}_{t-1})$ функций отражает возможные переходы РНС из одного внутреннего состояния в другое на t -й момент времени. Согласно условию (7) функция F_t^S выходов РНС на t -й момент времени должна принадлежать множеству $\mathbf{FS}(\mathbf{h}_{t-1})$ допустимых функций относительно $t-1$ момента.

Система выражений (1) – (7) представляет собой математическую модель РНС. В данном случае РНС формализуется в виде относительно конечного операционного перестраиваемого автомата. Все параметры такого автомата в общем случае конечны только на интервале одного шага работы. Они переопределяются относительно предыдущего шага.

С учетом этого РНС на каждый t -й момент времени можно формализовать автоматом $ARNN_t$ с восьмью параметрами,

$$ARNN_t = \{ \mathbf{h}_t, F_t^h, F_t^S, \mathbf{DA}(\mathbf{h}_{t-1}), \mathbf{DH}(\mathbf{h}_{t-1}), \mathbf{DB}(\mathbf{h}_{t-1}), \mathbf{FH}(\mathbf{h}_{t-1}), \mathbf{FS}(\mathbf{h}_{t-1}) \}. \quad (8)$$

Переход РНС в виде автомата $ARNN_t$ к $ARNN_{t+1}$ можно записать в виде

$$F_t^h : ARNN_t, \mathbf{SSP}_t^a \rightarrow ARNN_{t+1}. \quad (9)$$

Этот автомат, формализующий потоковую импульсную РНС, является перестраиваемой системой.

Устойчивость функционирования РНС согласно (1) – (9) во многом определяется ограничениями на внутренние ее состояния и состояния входов. При этом условие (3) представимо в виде ограничений на состояния нейронов и синапсов сети. Также условие (4) может предусматривать отдельные временные и пространственные ограничения на СЕИ, поступающие на различные входы РНС. Зависимость DH , DA , DB , FH , FS от h_{t-1} в (3) – (7) применительно к импульсной потоковой РНС можно объяснить следующими положениями. Каждый нейроны этой РНС после возбуждения переходит в состояние рефрактерности (временной невосприимчивости). Нахождение ряда нейронов в состояниях рефрактерности на момент $t-1$ влечет изменения допустимых множеств DH , DA , DB , FH , FS для момента t .

Опираясь на эту модель, поясним в следующем разделе предлагаемый метод устойчивой интеллектуальной обработки сигналов усовершенствованной потоковой РНС.

4. Метод. Раскроем предлагаемый метод на примере потоковой импульсной РНС с линейными схемами продвижения сигналов (последовательностей СЕИ) вдоль слоев (рисунок 2). На уровне нейросетевых каналов эта РНС представима в виде рисунка 3.

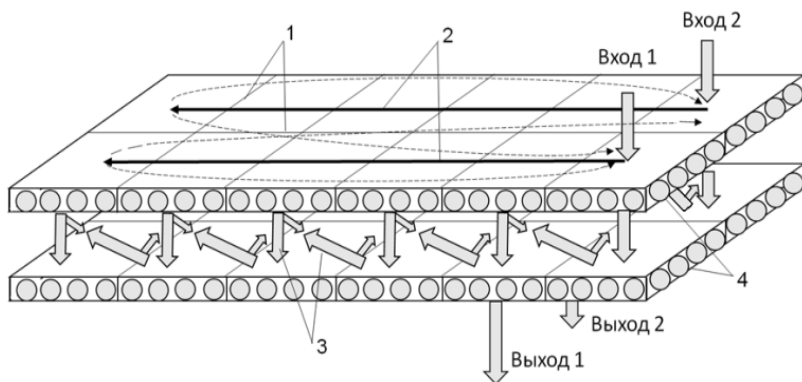


Рис. 2. Схема потоковой РНС с линейной структурой: 1 – линии разбивки слоев на логические поля за счет реализуемых пространственных сдвигов СЕИ, передаваемых от слоя к слою; 2, 3 – направления продвижения СЕИ вдоль и между слоями; 4 – нейроны. Штрихпунктирными стрелками показаны направления рекуррентного ассоциативного вызова СЕИ из памяти в режиме генерации сигналов

В такой РНС, чем больше длина нейросетевых каналов, тем лучше внимание и корректнее вызов сигналов из памяти по содержанию обрабатываемой информации.

В структурном плане эта РНС отличается от известных решений дополнительной несимметричностью ассоциативного связывания СЕИ в сети и вызова их из памяти.

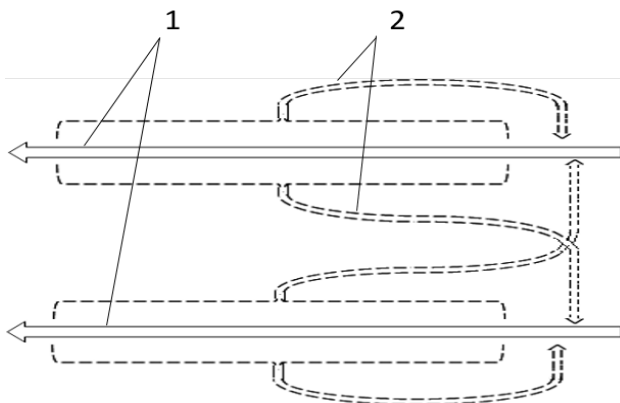


Рис. 3. Структура потоковой РНС на уровне нейросетевых каналов: 1 – каналы; 2 – рекуррентные ассоциативные взаимодействия обрабатываемых сигналов

Помимо реализации в нейросетевых каналах сети генерации сигналов (за счет рекуррентного вызова их из памяти), предусматривается перекрестный рекуррентный вызов сигналами одной модальности сигналами другой модальности. При этом выходы РНС смещены относительно входов на одно логическое поле. Отличия в части правил функционирования касаются особенностей предварительного преобразования входных сигналов и ввода их в сеть, механизмов связывания, запоминания и забывания в ней информации, аспектов регулирования порогами возбуждения нейронов в сети, условий устойчивого непрерывного обучения и генерирования сигналов.

С учетом этого и условий Ляпунова [10] предлагается новый метод устойчивой интеллектуальной обработки сигналов в потоковой РНС. В обобщенном виде этот метод можно свести к следующим положениям:

1. Подача сигналов на входы РНС, предварительно преобразованных в последовательности совокупностей единичных импульсов (СЕИ) по уточненным правилам. При этом СЕИ должны удовлетворять требованиям (4).

2. Передача СЕИ от слоя к слою, с продвижением их вдоль слоев по принятым маршрутам без искажений. Этот пункт метода и последующие должны реализовываться в рамках условий (1) – (3), (5) – (7) со своей спецификой.

3. Связывание в РНС поступающих совокупностей единичных импульсов при продвижении их по сети только с последующими СЕИ.

4. Запоминание результатов этого связывания на элементах потоковой РНС.

5. Изменение порогов возбуждения каждого нейрона в любом из логических полей слоев РНС в зависимости от числа уже возбужденных в этом поле нейронов.

6. При снижении загрузки РНС входными СЕИ меньше допустимого значения генерирование сетью последовательностей СЕИ посредством ассоциативного рекуррентного вызова их из памяти сети.

7. При возрастании загрузки РНС входными СЕИ снижение и прекращение генерирования сетью связанных сигналов. Потенциально управление загрузкой РНС входными сигналами может осуществляться в зависимости от содержания обрабатываемой в сети информации.

Поясним эти положения метода. В качестве исходных сигналов для обработки в РНС могут выступать различные воздействия (акустические, электрические, радио, оптические, химические и другие сигналы). Перед подачей в РНС они должны раскладываться, в общем случае, на пространственно-частотные или пространственно-химические составляющие. Каждая составляющая должна преобразовываться в последовательность единичных импульсов с частотой и фазой следования, как функциями от амплитуды и фазы составляющей. В частности, исходный сигнал $s(t)$ можно разложить в гармонический ряд Фурье (10):

$$s(t) = \sum_{n=0}^{\infty} s_n(t) = \sum_{n=0}^{\infty} A_n \cos(n\omega t + \varphi_n), \quad (10)$$

где A_n – амплитуда гармоник; $n\omega$ – частота гармоник, $\omega = 2\pi/T$; φ_n – фаза гармоник; T - период.

Каждую гармоническую составляющую этого ряда можно преобразовать в последовательность единичных импульсов с частотой следования F_n , как функцией от амплитуды A_n , и с задержкой τ_n , однозначно связанной с фазой φ_n :

$$F_n = 1/(2W_0R/A_n^2 + \tau_p + \tau_r), \quad (11)$$

$$\tau_n = 2W_0R/A_n^2 + \varphi_n/nw. \quad (12)$$

В (11), (12) приняты обозначения: W_0 – пороговая энергия, при которой происходит возбуждение первичных модельных нейронов в датчиках; R – входное сопротивление; τ_p – длительность формируемых единичных импульсов; τ_r – время рефрактерности, невосприимчивости нейронов после возбуждения. При этом заметим, что номер n каждой составляющей исходного сигнала должен совпадать с номером нейрона во входном поле РНС, а число таких последовательностей должно быть ограничено допустимыми пределами.

Для преобразования результатов обработки последовательностей единичных импульсов в соответствующие им исходные сигналы можно реализовать обратную процедуру. В интересах этого требуется определить частоту F_n^* следования единичных импульсов и их относительные задержки $\tau_{n\_z}^*$. Для определения F_n^* достаточно знать число импульсов, попадающих в предварительно заданный временной интервал [24]. Для установления значений $\tau_{n\_z}^*$ можно измерять относительные задержки единичных импульсов в их последовательностях. Параметры $W_0, R, \tau_p, \tau_r, nw$ считаются известными. С учетом этого амплитуды A_n^* и фазы φ_n^* восстановленных составляющих определяются как (13), (14):

$$A_n^* = (2W_0R/(1/F^* - \tau_p - \tau_r))^{\frac{1}{2}}, \quad (13)$$

$$\varphi_n^* = (\tau_{n\_z}^* - \frac{2W_0R}{(A^*)^2}) \cdot nw. \quad (14)$$

Таким образом, все параметры исходного сигнала $s^*(t)$ определены, и он может быть успешно восстановлен с применением современных методов цифрового синтеза. Возможны и другие преобразования [24].

В дискретном времени формируемые на входе РНС последовательности единичных импульсов можно рассматривать как последовательные совокупности этих импульсов. Если в качестве исходного сигнала выступает один обычный кадр изображения (при обработке цифрового видеопотока), он также должен преобразовываться в последовательность совокупностей единичных импульсов с длительностью, не превышающей период повторения этих кадров. В частном случае при векторном представлении входных сигналов каждая последовательная совокупность может содержать всего один импульс. Например, так можно кодировать последовательность слов в предложениях, если пространство логических полей в слоях нейронной сети позволяет отражать все слова из используемого словаря.

При этом каждая $i+1$ СЕИ не должна содержать импульсов из i -й СЕИ. Кроме этого ввод СЕИ в РНС должен осуществляться с периодом, меньше времени рефрактерности нейронов (невосприимчивости их после возбуждения). При непрерывном обучении это позволяет обеспечить баланс в РНС между возбуждающими и тормозящими воздействиями нейронов.

Передача СЕИ от слоя к слою с продвижением их вдоль слоев по принятому маршруту без искажений реализуется за счет приоритетности сильных связей в РНС. При передаче СЕИ по сети нейроны передающего слоя возбуждают нейроны принимающего слоя, который затем становится передающим слоем. Состояние $x_i(t_{r+1})$ каждого нейрона принимающего слоя РНС на момент времени t_{r+1} можно выразить как,

$$x_i(t_{r+1}) = \begin{cases} +1, & \text{когда } I_{\Sigma i}(t_r) \geq ID_i(t_r), \quad t_{i\text{ож}} \geq t_{i\text{реф}} \\ -1, & \text{если } t_{i\text{ож}} < t_{i\text{реф}} \\ 0, & \text{в других случаях} \end{cases}, \quad (15)$$

$$I_{\Sigma i}(t_r) = \sum_{j=1}^{J_e} w_{ij}(t_r) \cdot x_j(t_r). \quad (16)$$

В выражениях (15), (16) приняты обозначения: $I_{\Sigma i}(t_r)$ – суммарный ток на входе i -го нейрона на момент t_r ; $x_j(t_r)$ – состояние j -го нейрона передающего слоя РНС на момент времени t_r ; $w_{ij}(t_r)$ – вес (проводимость) синапса, связывающего j -й нейрон с i -м нейроном принимающего слоя; J_e – число возбужденных нейронов передающего слоя, по которым осуществляется суммирование весов синапсов; $ID_i(t_r)$ – порог возбуждения i -го нейрона на момент времени t_r ; $t_{i_{ож}}$ – время ожидания i -го нейрона после последнего возбуждения; $t_{i_{реф}}$ – время невосприимчивости i -го нейрона после последнего возбуждения. Согласно (15) при нахождении нейрона в состоянии возбуждения $x_i(t_{r+1}) = 1$, а в состоянии невосприимчивости $x_i(t_{r+1}) = -1$. В состоянии ожидания $x_i(t_{r+1}) = 0$.

Как и в [3] веса $w_{ij}(t_r)$ синапсов определяются через произведение их весовых коэффициентов $k_{ij}(t_r)$ и функций ослабления $\mu_{ij}(t_r)$ расходящихся и сходящихся единичных образов (17) – (18),

$$w_{ij}(t_r) = k_{ij}(t_r) \cdot \mu_{ij}(t_r), \quad (17)$$

$$k_{ij}(t_r) = (1 - \exp(-\gamma \cdot g_{ij}(t_r))) / (1 + \exp(-\gamma \cdot g_{ij}(t_r))) = \text{th}(\gamma \cdot g_{ij}(t_r) / 2), \quad (18)$$

$$g_{ij}(t_r) = g_{ij}(t_{r-1}) \pm \Delta g_{ij}(t_{r-1}). \quad (19)$$

В (18) γ – постоянный коэффициент; $g_{ij}(t_r)$ – на момент t_r суммарное число единичных импульсов, прошедших через соответствующий синапс, с учетом их знака. Значение $\Delta g_{ij}(t_{r-1})$ зависит от состояний, взаимодействующих i -го и j -го нейронов. Если возбужденный j -й нейрон вызвал возбуждение i -го нейрона, то значение $\Delta g_{ij}(t_{r-1})$ положительное. Когда сгенерированный j -м

нейроном сигнал воздействует на i -й нейрон в состоянии невосприимчивости, то $\Delta g_{ij}(t_{r-1})$ отрицательно. В остальных случаях $\Delta g_{ij}(t_{r-1}) = 0$. Функции ослабления $\mu_{ij}(t_r)$, за счет изменения которых реализуются пространственные сдвиги СЕИ, можно определить согласно [3].

Для исключения перевозбуждения РНС порог возбуждения каждого i -го нейрона в состоянии ожидания согласно (15) предлагается определять, исходя из следующих правил. В каждом логическом поле сначала возбуждается нейрон с наибольшим значением $I_{\Sigma i}(t_r)$, преодолевающим порог возбуждения этого нейрона в состоянии ожидания, с одновременным увеличением порогов возбуждения для других нейронов в этом поле. Затем может возбудиться другой нейрон с наибольшим значением суммарного тока на входе, если это значение превышает повышенный порог и так далее. Пороги возбуждения у оставшихся нейронов в состоянии ожидания возвращаются к исходному состоянию к моменту t_{r+1} .

Согласно предлагаемому методу увеличение числа возбуждающих воздействий на нейроны приводит к увеличению числа тормозящих воздействий, а возрастание числа последних влечет снижение числа первых. При таких условиях всегда можно добиться, чтобы суммарное число возбужденных нейронов в каждом слое РНС не превышало заданного значения $B_{зад}$.

$$\sum_{i(j)=1}^{I(J)} x_{i(j)}(t_r) < B_{зад}, \quad (20)$$

Также когда соблюдаются начальные условия на момент времени t_0 с учетом (17)–(19), (20) и правил ввода СЕИ в РНС применительно к суммарному весу синапсов справедливо условие (21):

$$\left\| \sum_{s=1}^2 \sum_{i=1}^I \sum_{j=1}^J w^s_{ij}(t_r) - W_0 \right\| < \varepsilon_2, \text{ при } t_r > t_0, \quad (21)$$

Согласно этому условию суммарный вес синапсов двухслойной РНС как при обучении, так и при генерации сигналов будет находиться в пределах некоторой «трубки».

В соответствии с методом, когда загрузка РНС входными сигналами больше допустимого значения, генерирование сетью последовательностей СЕИ посредством ассоциативного рекуррентного вызова их из памяти сети невозможно. Входные сигналы, продвигаясь по сети по сильным связям, подавляют возможные ассоциативные вызовы СЕИ из памяти. При этих условиях РНС функционирует только в режиме непрерывного обучения. При снижении загрузки РНС входными сигналами ниже допустимого значения осуществимо генерирования сетью связанных сигналов посредством вызова их из памяти.

5. Особенности решения интеллектуальных задач предлагаемым методом. В отличие от [3, 25] в предлагаемом методе исключается ассоциативный вызов СЕИ из памяти РНС в направлении продвижения сигналов по сети. Однако сохраняется рекуррентный вызов. Это приводит к тому, что вся интеллектуальная обработка сигналов базируется на решении задач прямого анализа. К таким задачам относится последовательный вызов из памяти сигналов на основе ранее воспринятых, прогнозирование и распознавание событий, перевод речи, оценивание эффективности реализуемых процессов и другие. Формально эти задачи на уровне нейросетевой обработки совокупностей единичных импульсов, обозначенных, как и в (1), (2), в виде **SSP**, сводятся к схеме (22):

$$\{\mathbf{SSP}_{t-k}, \dots, \mathbf{SSP}_{t-2}, \mathbf{SSP}_{t-1}, \mathbf{SSP}_t\} \rightarrow \mathbf{SSP}_{t+1} \text{ при } t = \overline{1, T}. \quad (22)$$

Согласно этой схеме ищется $\mathbf{SSP}_{t+1}$ на момент времени $t+1$, исходя из ранее введенных в РНС совокупностей $\{\mathbf{SSP}_{t-k}, \dots, \mathbf{SSP}_{t-2}, \mathbf{SSP}_{t-1}, \mathbf{SSP}_t\}$ и содержания ее памяти.

Решение обратных задач возможно по этой же схеме, но при обучении РНС потоками сигналов в обратном порядке (23):

$$\{\mathbf{SSP}_t, \mathbf{SSP}_{t-1}, \mathbf{SSP}_{t-2}, \dots, \mathbf{SSP}_{t-k}\} \rightarrow \mathbf{SSP}_{t-(k+1)} \text{ при } t = \overline{1, T}. \quad (23)$$

Также обратные задачи могут сводиться к поиску более ранних сигналов, относительно которых можно прийти к известному результату $\mathbf{SSP}_{зад}$. В этом случае схема обработки предусматривает многократное решение задач обычного прямого анализа (24):

$$\left\{ \begin{aligned} &\{\mathbf{SSP}_{t-k}^1, \dots, \mathbf{SSP}_{t-2}^1, \mathbf{SSP}_{t-1}^1, \mathbf{SSP}_t^1\} \rightarrow \mathbf{SSP}_{t+1}^1 \\ &\dots \\ &\{\mathbf{SSP}_{t-k}^f, \dots, \mathbf{SSP}_{t-2}^f, \mathbf{SSP}_{t-1}^f, \mathbf{SSP}_t^f\} \rightarrow \mathbf{SSP}_{t+1}^f, \\ &\dots \\ &\{\mathbf{SSP}_{t-k}^D, \dots, \mathbf{SSP}_{t-2}^D, \mathbf{SSP}_{t-1}^D, \mathbf{SSP}_t^D\} \rightarrow \mathbf{SSP}_{t+1}^D \end{aligned} \right. \quad (24)$$

при t от 1 до T , и определение одной или несколько совокупностей
единичных импульсов, например, из последовательности

$$\{\mathbf{SSP}_{t-k}^f), \dots, \mathbf{SSP}_{t-2}^f, \mathbf{SSP}_{t-1}^f, \mathbf{SSP}_t^f\}.$$

при условии (25):

$$\mathbf{SSP}_{t+1}^f \cong \mathbf{SSP}_{3q\partial}. \quad (25)$$

К обратным задачам относятся, прежде всего, задачи обоснования требований к процессам и системам, определение первопричин наступления событий и другие.

При решении задач синтеза согласно предлагаемому методу считаются известными одна $\mathbf{SSP}_{0\_зад}$ или несколько начальных СЕИ, а также одна $\mathbf{SSP}_{e\_зад}$ или несколько конечных СЕИ. Требуется найти последовательность СЕИ, позволяющую перейти от $\mathbf{SSP}_{0\_зад}$ к $\mathbf{SSP}_{e\_зад}$ при удовлетворении ряду условий. В качестве таких условий могут выступать ограничения на число шагов, на время перехода и другие. Схему решения задач синтеза можно представить также в виде многократной реализации задач анализа (24), но при условиях, например, когда

$$\mathbf{SSP}_{t-k}^f = \mathbf{SSP}_0 \text{ и } \mathbf{SSP}_{t+1}^f = \mathbf{SSP}_e$$

$$W_z(\{\mathbf{SSP}_{t-k}^f, \dots, \mathbf{SSP}_{t-2}^f, \mathbf{SSP}_{t-1}^f, \mathbf{SSP}_t^f, \mathbf{SSP}_{t+1}^f\}) \in W_{z\_don}, z = \overline{1, Z}. \quad (26)$$

В (26) $W_z(.)$ – z-е показатели, которые должны относиться к областям допустимых значений $W_{z\_доп}$. Результатом решения этой задачи выступает последовательность СЕИ:

$$\{SSP_{t-k}^f = SSP_{0\_зад}, \dots, SSP_{t-2}^f, SSP_{t-1}^f, SSP_t^f, SSP_{t+1}^f = SSP_{e\_зад}\}.$$

Учитывая эти особенности, можно утверждать, что все виды задач могут решаться на основе прямой задачи анализа, лежащей в основе принципов работы усовершенствованной потоковой РНС. При этом все умение взаимодействовать с интеллектуальной системой на основе этой РНС, чтобы получить интересующий результат, может сводиться к правильному воздействию на нее. Некоторые примеры решения задач согласно подобным схемам приведены в [25].

Заметим, что в предлагаемом методе не предусматривается управление физическим временем работы РНС. Для предсказания будущих событий и обращение к прошлому времени обучение РНС должно осуществляться со связыванием основных сигналов со свойствами времени их проявления. В дальнейшем это позволяет обращаться к свойствам времени и вызывать из памяти РНС связанные с ними сигналы. Для управления модельным временем можно изменять масштаб вызываемых из памяти РНС сигналов.

В частном случае, усовершенствованная РНС применима и в системах управления физическим временем. В этом случае необходимо использовать две идентичные усовершенствованные РНС. Первая РНС должна функционировать только в режиме непрерывного обучения, воспринимать и запоминать последовательности СЕИ. На второй РНС на основе периодического считывания информации о состоянии первой РНС могут решаться прикладные задачи. Так при решении задачи прогнозирования согласно схеме (22) выполняются следующие действия. С первой РНС считываются во вторую РНС текущие состояния нейронов, несущие информацию о введенной в первую сеть последовательности СЕИ (левая часть формулы (22)). Также во вторую РНС считываются данные о текущих значениях весов синапсов первой сети. Затем вторая РНС начинает функционировать в ускоренном времени и вызывать из памяти результаты прогнозирования (правая часть формулы (22)) без прерывания работы первой РНС.

Однако заметим, что решение интеллектуальных задач потоковой рекуррентной нейронной сетью, реализуемой программным путем, требует существенных вычислительных ресурсов. Так

пространство только самих нейронов в РНС $3^{2 \cdot J}$ вариантов, где J - число нейронов в каждом слое, а пространство весов синапсов - $K^{2 \cdot J^2}$. При этом число возможных состояний K одного взятого синапса зависит от значения Δg_{ij} и может исчисляться десятками. При числе нейронов в РНС $J=1000$, для кодирования возможных состояний сети требуется около 6,5 Мб, а при $J=10^6$ уже около 6500 Мб. К этому добавим, что непрерывное обучение РНС, в том числе при генерации новых сигналов, предусматривает перевод сети из одних состояний в другие, что кратно увеличивает сложность программного моделирования. Перспективным путем выступает аппаратная реализация РНС на аналоговых нейроморфных процессорах с мемристивной памятью. При этом появляется возможность использования в РНС нейронов с внутренней памятью и дополнительной обработкой сигналов на их входах [25]. Это позволяет существенно расширить память РНС и повысить избирательность обращения к ней по содержанию обрабатываемых сигналов.

Несомненно, в частном случае отказавшись от непрерывного обучения, можно существенно снизить вычислительную сложность моделирования нейросетевой обработки сигналов. Однако это не позволяет обеспечить высокие возможности РНС по обработке сигналов и событий с изменяющимися законами поведения. Во всех случаях необходимо обеспечивать устойчивое функционирование РНС.

6. Результаты моделирования. В интересах подтверждения работоспособности предложенного метода проводилось моделирование. На входы потоковой РНС подавались последовательности СЕИ, состоящие из 2, 3, 4 и 7 и 15 единичных импульсов, несущих информацию о закодированных сигналах. При этом каждая последующая $i+1$ СЕИ не содержала импульсов из i -й СЕИ. Ввод СЕИ в РНС осуществлялся с периодом, меньше времени невосприимчивости нейронов после возбуждения. Длительность таких последовательностей составляла 1500 СЕИ. В качестве базовой потоковой импульсной РНС выступала сеть с числом нейронов в каждом слое 630 единиц. Слои сети за счет реализуемых пространственных сдвигов СЕИ, передаваемых от слоя к слою, разбивались на логические поля размером 6×7 нейронов. Обрабатываемые СЕИ посредством этих сдвигов продвигались вдоль слоев по линейной схеме, как на рисунках 1, 2. Связывание в РНС поступающих совокупностей единичных импульсов при продвижении их по сети осуществлялось только с последующими СЕИ. Нейроны

одного и того же слоя связей друг с другом не имели и функционировали согласно (15) – (19).

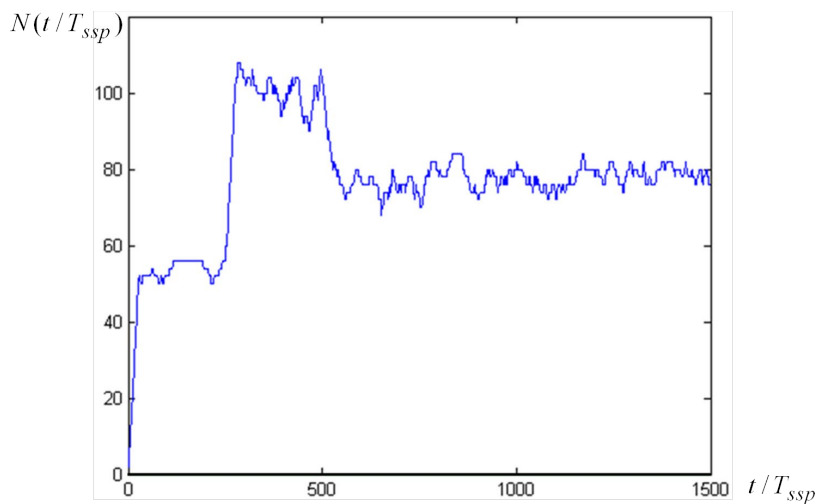
Моделирование не ограничивалось одной РНС с числом нейронов в слоях 630 единиц и размерами полей 6×7 . Базовая модель позволяла моделировать РНС с разным количеством логических полей (размерами 6×7) в слоях. В частности, исследовались также РНС с числом нейронов в слоях 840 и 1050 единиц. Получаемые результаты принципиально не отличались, только с увеличением числа нейронов в слоях существенно возрастало время моделирования. Обработывались не только фиксированные, но и случайно генерируемые последовательности СЕИ.

Анализировались текущая загрузка РНС единичными импульсами, текущий суммарный вес синапсов сети, а также возможности ассоциативного рекуррентного вызова сигналов из ее памяти. На рисунке 4(а, б) показано, как изменяется загрузка базовой РНС (с числом нейронов в каждом слое 630 единиц) единичными импульсами в режиме непрерывного обучения при обработке последовательности длиной 1500 СЕИ с временем рефрактерности нейронов сети, равным одному шагу ее работы.

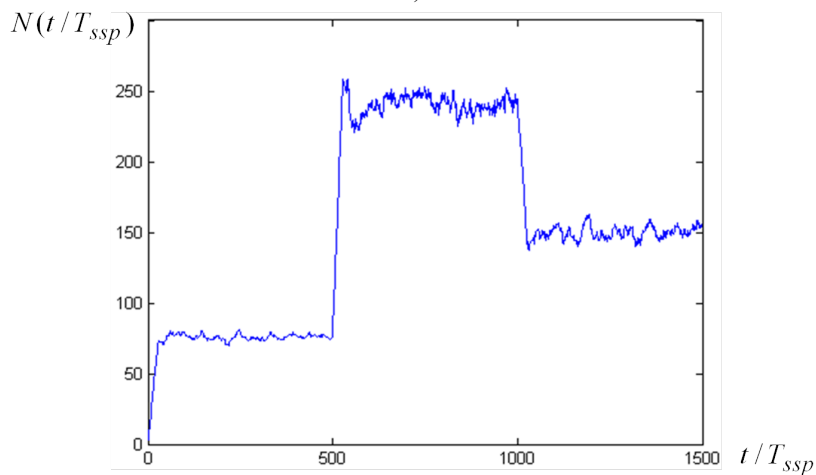
При восприятии последовательности СЕИ, состоящих на интервалах 0 – 250, 251 – 500 и 501 – 1500, соответственно, из двух, четырех и трех импульсов, график этой загрузки отражен на рисунке 4(а). Для последовательности СЕИ, с числом импульсов 3, 15, 7 импульсов, соответственно на интервалах 0 – 250, 251 – 500 и 501 – 1500 график загрузки другой (рисунок 4(б)).

Согласно рисунку 4(б) первый и второй всплески, обусловлены оперативным заполнением РНС соответствующими СЕИ до моментов начала покидания ими сети. В установившихся режимах в первом случае число импульсов в РНС колеблется около отметки 70 единиц, а во втором случае – около 240 единиц. После подачи в РНС СЕИ с числом импульсов, равным 7, вместо 15, колебания происходят около отметки 150 единиц.

Из анализа рисунка 4 и других аналогичных экспериментов сделан вывод, что на интервале длиной 1500 последовательных СЕИ загрузка РНС изменяется в пределах некоторой «трубки», что удовлетворяет критериям Ляпунова для равномерной устойчивости.



а)

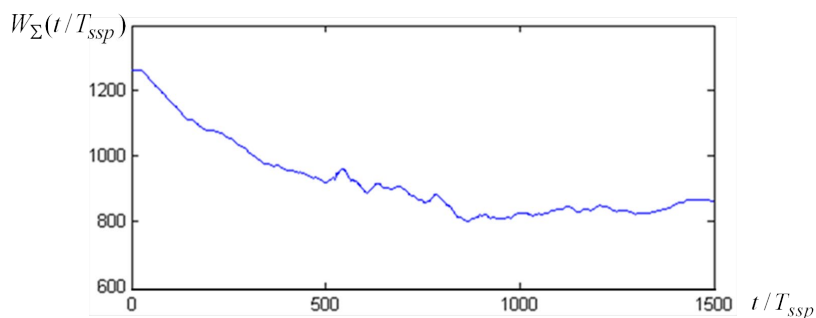


б)

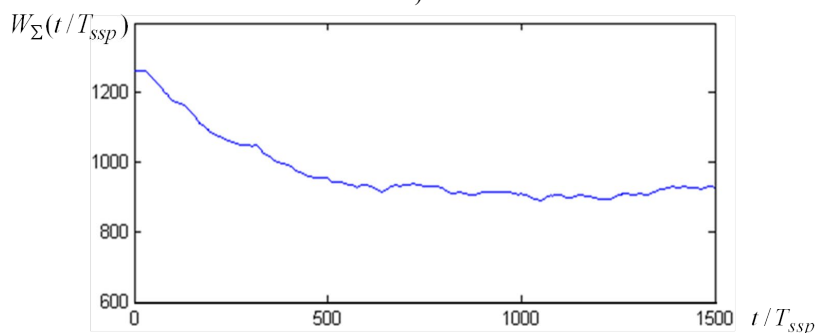
Рис. 4. Примеры загрузки $N(t/T_{ssp})$ РНС единичными импульсами в режиме непрерывного обучения при обработке последовательности СЕИ с суммарной длиной 1500 совокупностей: T_{ssp} – период повторения СЕИ, равный одному шагу работы сети (десятые доли секунды); t – текущее время, кратное T_{ssp}

Установлено, что суммарный вес $W_{\Sigma}(t/T_{ssp})$ синапсов (рисунок 5(а, б)), являясь положительным по значению, с увеличением длины последовательности обрабатываемых СЕИ стремится к некоторой точке равновесия. Такое стремление удовлетворяет требованиям устойчивости, в части отрицательного приращения по времени функции Ляпунова для дискретного времени.

Рисунок 5(а) соответствует изменению суммарного веса синапсов для загрузки РНС согласно рисунку 4(б). При загрузке РНС последовательностями СЕИ, состоящими только из трех импульсов, имеем рисунок 5(б). Такие зависимости можно объяснить тем, что число введенных в сеть единичных импульсов, содержащихся в нечетных СЕИ, стремится к числу этих импульсов в четных СЕИ. В следствие этого число положительных взаимодействий сигналов в РНС стремится к числу отрицательных их ассоциаций.



а)



б)

Рис. 5. Изменение суммарного веса $W_{\Sigma}(t/T_{ssp})$ синапсов при обработке последовательности длиной 1500 СЕИ

Заметим, что начальная точка суммарного веса синапсов равна 1260 – числу сильных связей в сети с весами, равными единице. Для рассмотренного примера на рисунке 5(б) число положительных ассоциаций между обрабатываемыми импульсами при длине последовательности 1500 СЕИ составило 4212295 единиц, а отрицательных – 4362269 единиц. Отклонения их друг от друга лежат в пределах 3,5%.

Если допустить, что $W_{\Sigma}(t/T_{ssp})$ и t/T_{ssp} являются непрерывными величинами, то графики на рисунке 5 можно аппроксимировать функцией (27),

$$W_{\Sigma}^*(x) = W_0 - \Delta W(1 - \exp(-\lambda \cdot x)), \quad (27)$$

где W_0 – значение суммарных весов синапсов на начальный момент; ΔW – предельный спад значений этих весов; λ – интенсивность спада суммарных значений весов. Эту функцию можно отнести к функции Ляпунова. Производная от нее по x , как функции времени, имеет отрицательное значение и при $x \rightarrow \infty$ стремится к нулю (28),

$$\frac{dW_{\Sigma}^*(x)}{dx} = -\Delta W \cdot \lambda \cdot \exp(-\lambda \cdot x) \leq 0. \quad (28)$$

Значения ΔW и λ в (27), (28) можно определить с применением метода наименьших квадратов [26].

Один из вариантов распределения значений (19) положительных и отрицательных рекуррентных ассоциаций сигналов через синапсы между двумя логическими полями (размером $42=6 \times 7$ нейронов) разных слоев РНС показан на рисунке 6. Вертикальная и горизонтальная оси на рисунке 6 соответствуют номерам нейронов, взаимодействующих двух полей разных слоев РНС. Цветными пикселями обозначены синапсы, через которые реализуются ассоциации между нейронами. Цвет каждого синапса отражает суммарное число единичных импульсов прошедших через него с учетом их знака. Справа на рисунке 6 цветовая шкала значений.

Из анализа рисунка 6 видно, что такое распределение носит относительно равномерный характер в части положительных и отрицательных взаимодействий сигналов. Это свидетельствует о том, что в явном виде катастрофическое забывание ранее запомненных сигналов в анализируемой РНС отсутствует.

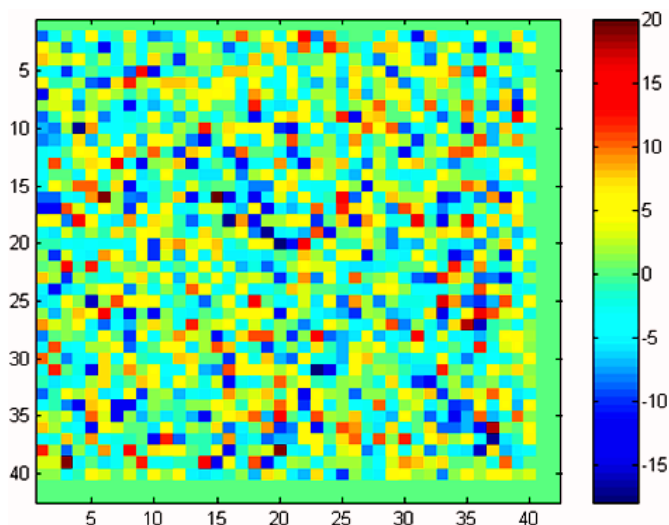


Рис. 6. Распределение значений положительных и отрицательных рекуррентных ассоциаций сигналов между двумя логическими полями нейронов в РНС

Пример состояния первого слоя РНС при непрерывном обучении приведен на рисунке 7. Согласно рисунку 7 вводимые в сеть СЕИ, содержащие по три импульса, продвигаются вдоль слоев справа налево и ассоциируются. При этом исключается ассоциативный вызов из нейросетевой памяти ранее запомненных сигналов.

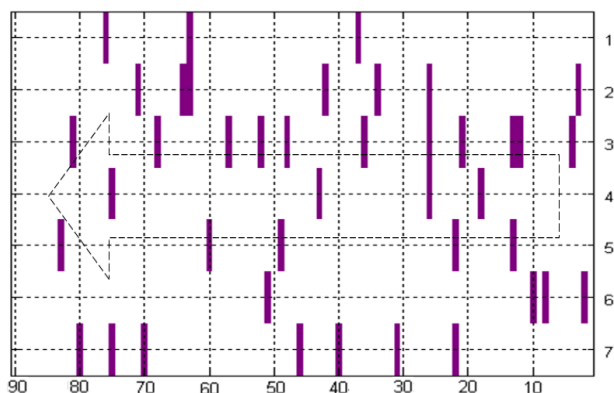
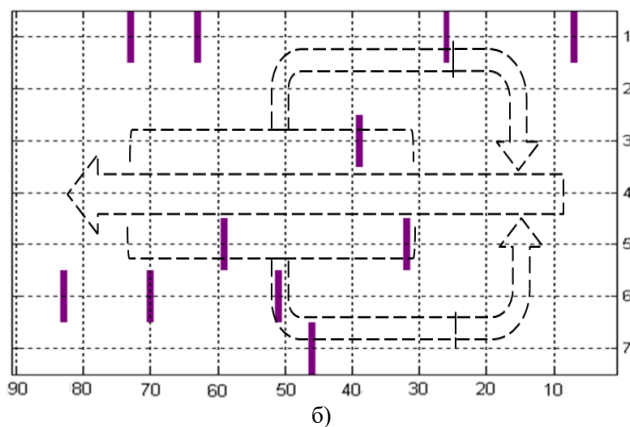
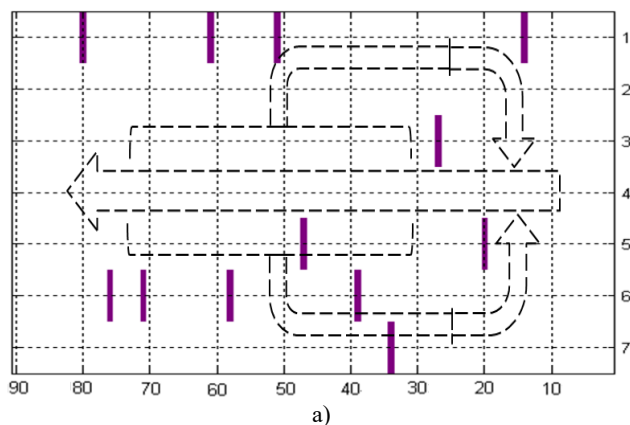


Рис. 7. Пример состояния первого слоя РНС при непрерывном обучении

Вертикальная ось на рисунке 7 соответствует номерам нейронов в столбцах слоя сети, а горизонтальная ось – числу таких столбцов. Жирными линиями отражены возбужденные нейроны на рассматриваемый момент времени.

Подтверждено, что после непрерывного обучения, например, последовательностями СЕИ, состоящими только из одиночных импульсов, при снижении загрузки РНС входными сигналами сеть способна устойчиво генерировать СЕИ, ассоциативно вызывая их из памяти. Примеры такой генерации приведены на рисунке 8.



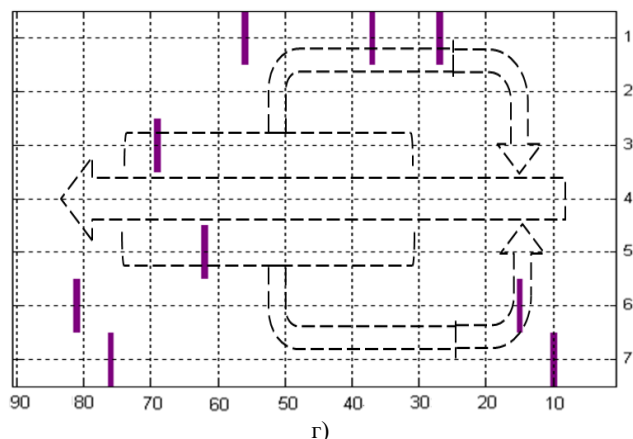
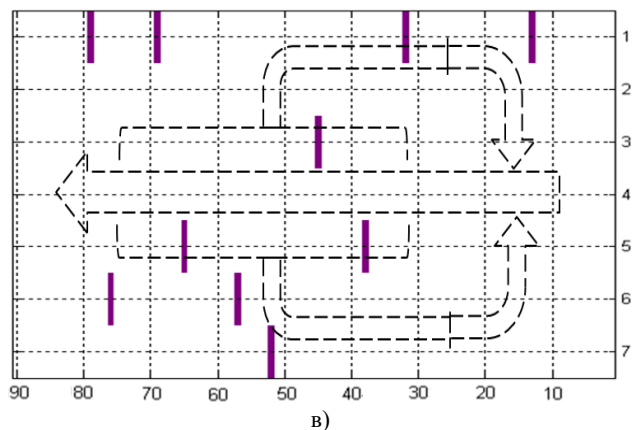


Рис. 8. Примеры состояний первого слоя РНС при рекуррентном вызове сигналов из памяти с непрерывным запоминанием результатов с продвижением СЕИ вдоль слоев справа налево. Состояния на рисунке 8 (а, б, в, г) расположены в порядке очередности

Схемы ассоциативного рекуррентного вызова сигналов из памяти РНС на рисунке 8 показаны штрихпунктирными линиями. Согласно рисунку 8 более ранние сигналы вызывают из памяти следующие за ними СЕИ. Затем они продвигаются по сети без искажений. После прохождения по сети СЕИ покидают ее, но оставляют «след» о своем присутствии. Рисунок 8(б) отличается от рисунка 8(а) тем, что все сигналы сдвинуты влево с генерацией нового

сигнала справа. Аналогичная ситуация с рисунком 8(в, г), только при большем числе сдвигов и вызовах из памяти новых сигналов в виде единичных импульсов.

Эксперименты с РНС, содержащими в каждом слое по 840 и 1050 нейронов показали, что получаемые результаты принципиально не отличаются от рисунков 4–8. В частности, по отношению к рисункам 7, 8 особенности состояли в большем числе последовательных СЕИ, одновременно обрабатываемых в РНС. С увеличением числа одновременно обрабатываемых в сети СЕИ, повышался уровень внимания РНС при рекуррентном вызове сигналов из ассоциативной памяти.

Сравнительная характеристика трех потоковых РНС с числом нейронов в идентичных слоях 630, 840 и 1050 единиц при обработке последовательностей сигналов длиной 1500 СЕИ отражена в таблице 1. В числителях приведены значения $N(t/T_{ssp})$ загрузки РНС единичными импульсами в установившихся режимах, при обработке одинаковых по длине последовательностей СЕИ, но различных по объему самих СЕИ. В знаменателях отражены значения T_{ssp} – длительности шагов обработки СЕИ в РНС, равных 1/2 периода их поступления в секундах. Время рефрактерности нейронов РНС в этом случае, в отличие от примеров на рисунке 4, составляло двум шагам работы сети.

Из анализа таблицы видно с увеличением числа нейронов в каждом слое РНС с 630 единиц до 1050 единиц текущая загрузка сети при обработке последовательности СЕИ (по четыре импульса) возрастает от 48 до 83 импульсов. При этом время T_{ssp} увеличивается в 2,8 раза. С позиции ВІВО устойчивости согласно таблице 1 во всех экспериментах при ограниченных входных воздействиях на потоковые РНС были полученные ограниченные состояния сетей.

Установлено, что важную роль в обеспечении устойчивой обработки СЕИ потоковыми РНС с непрерывным обучением играет соблюдение следующего условия. Сигналы должны подаваться в сеть с периодом меньше времени невосприимчивости нейронов после возбуждения. Последовательности таких совокупностей единичных импульсов могут нести всю информацию о различных физических и других сигналах.

Таблица 1. Сравнительная характеристика потоковых РНС по значениям загрузки в установившихся режимах и времени обработки СЕИ

№ п/п	Число нейронов в каждом слое РНС	Число единичных импульсов в каждой СЕИ входной последовательности			
		2	3	4	7
1	630	27	38	48	79
		0,388	0,412	0,424	0,452
2	840	35	52	65	100
		0,696	0,732	0,760	0,800
3	1050	45	63	83	122
		1,096	1,156	1,192	1,292

Также для устойчивой генерации СЕИ требуется обеспечивать несимметричность ассоциативного вызова сигналов из памяти и изменение порогов возбуждения нейронов в логических полях в зависимости от их текущей загрузки.

При несоблюдении этих условий, суммарные значения весов синапсов могут стремиться к своему пределу с катастрофически забыванием информации с последующим перевозбуждением РНС.

В практическом плане, потоковые РНС, содержащие от нескольких сотен до нескольких тысяч нейронов в слоях, могут реализовываться программным путем. РНС с этими характеристиками применимы для решения ряда задач нейросетевого прогнозирования и нейроуправления трудно формализуемыми процессами с непрерывным обучением. При этом пространство состояний этих процессов на дискретные моменты времени должно быть соизмеримо с пространством входных полей РНС.

В частности, с применением потоковых РНС проводились исследования по оценке точности решения задач прогнозирования скорости потоков транспортных средств на восьми улицах Санкт-Петербурга согласно схеме (22). В интересах этого формировались временные ряды со значениями этих скоростей с интервалами в три часа. Значения скоростей перед вводом в РНС кодировались в виде конкретных СЕИ. Результаты оценивались по показателям средней абсолютной ошибки (MAE – mean absolute error), средней абсолютной процентной ошибки (MAPE – mean absolute percentage error) и среднеквадратической ошибки (RMSE – root mean square error). Полученные оценки сравнивались с аналогичными показателями, полученными при применении известных методов [14, 27, 28]: долгой

краткосрочной памяти (LSTM – long short – term memory) и интегрированной модели авторегрессии – скользящего среднего (ARIMA – autoregressive integrated average). Установлено, что применение потоковых РНС по сравнению с LSTM позволяет снизить MAE на 21,8%, MAPE на 9,6%, RMSE на 15,3%. По отношению к ARIMA выигрыш потоковых РНС составил по MAE – 21,0%, MAPE – 21,3%, RMSE – 11,0%.

7. Заключение. Следуя принципам научной школы члена-корреспондента РАН Р.М. Юсупова, проведен анализ возможных решений по обеспечению устойчивой интеллектуальной обработки сигналов потоковыми импульсными РНС с непрерывным обучением. К наиболее перспективным подходам по обеспечению этой устойчивости отнесены использование неперекрывающихся и ортогональных представлений обрабатываемых в РНС сигналов, а также другие методы на основе регуляризации.

Поставлена и решена задача разработки новых методов устойчивой интеллектуальной обработки сигналов потоковыми РНС с непрерывным обучением. В интересах ее решения РНС формализована относительно конечным операционным автоматом. Предложен новый метод устойчивой интеллектуальной нейросетевой обработки сигналов с непрерывным обучением и реализующая его усовершенствованная потоковая импульсная РНС. К основным особенностям предложенного метода относятся новые правила предобработки и введения сигналов в сеть, условия запоминания и генерации последовательностей СЕИ, а также схемы решения потоковыми РНС прямых и обратных задач анализа и синтеза. Не смотря, что в основе функционирования усовершенствованной РНС лежит решение прямой задачи анализа, за счет применения специальных схем ввода информации в сеть можно решать и другие задачи. Усовершенствования потоковой РНС касаются следующих аспектов. Это введение в сеть дополнительной несимметричности ассоциативной обработки сигналов. Использованы импульсные нейроны с продолжительной рефрактерностью и с динамическими порогами возбуждения, зависящими от уровня загрузки отдельных логических полей слоев сети. Добавлены правила по приоритетности возбуждения нейронов, относящихся к одним и тем же логическим полям.

В отличие от известных методов, лежащих в основе таких систем как Chat GPT, Deep Seek и других, реализованных на основе нейронных сетей трансформеров, предложенный метод предусматривает непрерывное обучение и оперативную адаптацию к изменяющимся внешним условиям. Непрерывное устойчивое

обучение позволяет обойти ограничения, свойственные Chat GPT, Deep Seek и другим языковым моделям. Кроме этого предложенный метод с непрерывным обучением, в отличие от известных решений, может успешно обрабатывать различные, заранее неизвестные сигналы.

Результаты моделирования показали, что предложенный метод и реализующая его усовершенствованная потоковая РНС способны устойчиво непрерывно обучаться без явного катастрофического забывания ранее воспринятых сигналов и устойчиво генерировать последовательности связанных СЕИ. В такой РНС, при выполнении условий согласно предложенному методу, обеспечивается баланс между положительными и отрицательными ассоциативными взаимодействиями сигналов в сети (между запоминанием новой и забыванием устаревшей информации). Эта РНС может обрабатывать сверхдлинные последовательности СЕИ и не перегружаться.

Предложенные новые решения могут найти применение при создании перспективных интеллектуальных роботов с непрерывным обучением и нейропроцессоров нового поколения. В качестве роботов могут выступать те, которые должны функционировать в условиях высокой неопределенности событий, постоянно изменяющихся законах их проявления. Это могут быть беспилотные подводные, наземные, воздушные, космические и другие роботы.

Литература

1. Man N., Osipov V., Zhukova N., Subbotin A., Ignatov D.I. Neural networks for intelligent multilevel control of artificial and natural objects based on data fusion: a survey. *Information Fusion*. 2024. vol. 110.
2. Осипов В.Ю. Устойчивость рекуррентных нейронных сетей с управляемыми синапсами. *Информационные технологии*. 2011. Т. 9. С. 69–73.
3. Osipov V., Osipova M. Space-time signal binding in recurrent neural networks with controlled elements. *Neurocomputing*. 2018. vol. 308. pp. 194–204.
4. Parisi G., Kemker R., Part J., Kanan C., Wermter S. Continual lifelong learning with neural networks: A review. *Neural Networks*. 2019. vol. 113. pp. 54–71.
5. Van de Ven G., Siegelmann H., Tolias A. Brain-inspired replay for continual learning with artificial neural networks. *Nature communications*. 2020. vol. 11. DOI: 10.1038/s41467-020-17866-2.
6. Tercan H., Deibert P., Meisen T. Continual learning of neural networks for quality prediction in production using memory aware synapses and weight transfer. *Journal of Intelligent Manufacturing*. 2022. vol. 33. pp. 283–292.
7. Haykin S. *Neural Networks and Learning Machines*, third ed., Prentice Hall, New York, 2008. 936 p.
8. Ehret B., Henning C., Cervera M., Meulemans A., Oswald J., Grewe B. Continual learning in recurrent neural networks. *ICLR 2021 Conference Paper*. 2021.
9. Schmidgall S., Ziaei R., Achterberg J., Kirsch L., Hajiseyedrazi P., Eshraghian J. Brain-inspired learning in artificial neural networks: A review. *APL Machine Learning*. 2024. vol. 2.

10. Красовский А.А. Справочник по теории автоматического управления. М.: Наука. Гл. ред. физ.-мат. Лит, 1987. 712 с.
11. Rajchakit G., Agarwal P., Ramalingam S. *Stability Analysis of Neural Networks*. Springer, 2021. DOI: /0.1007/978-981-16-6534-9.
12. Xie K., Zhang C., Lee S., He Y., Liu Y. Delay-dependent Lurie-Postnikov type Lyapunov-Krasovskii functionals for stability analysis of discrete-time delayed neural networks. *Neural Networks*. 2024. vol. 173.
13. Vaswani A., Shazeer N., Parmar N., Uszkoreit J., Jones L., Gomez A., Kaiser L., Polosukhin I. Attention is all you need. 31st Conference on Neural Information Processing Systems (NIPS 2017). 2017. pp. 6000–6010.
14. Osipov V. Nikiforov V., Zhukova N., Miloserdov D. Urban traffic flows forecasting by recurrent neural networks with spiral structures of layers. *Neural Computing and Applications*. 2020. vol. 32(18). pp. 14885–14897.
15. Kirkpatrick J., Pascanu R., Rabinowitz N., Veness J., Desjardins G., Rusu A., Milan K., Quan J., Ramalho T., Grabska-Barwinska A., Hassabis D., Clopath C., Kumaran D., Hadsell R. Overcoming catastrophic forgetting in neural networks. arXiv preprint. arXiv 1612.00796v2. 2017.
16. Zenke F., Poole B., Ganguli S. Continual learning through synaptic intelligence. arXiv preprint. arXiv 1703.04200v3. 2017.
17. Coop R., Arel I. Mitigation of catastrophic forgetting in recurrent neural networks using a fixed expansion layer. The 2013 International Joint Conference on Neural Networks (IJCNN). 2013. pp. 1–7. DOI: 10.1109/IJCNN.2013.6707047.
18. Duncker L., Driscoll L., Shenoy K., Sahani M., Sussillo D. Organizing recurrent network dynamics by task-computation to enable continual learning. *Advances in Neural Information Processing Systems 33 (NeurIPS 2020)*. 2020. vol. 33. pp. 14387–14397.
19. Cossu A., Carta A., Lomonaco V., Bacciu D. Continual learning for recurrent neural networks: An empirical evaluation. *Neural Networks*. 2021. vol. 143. pp. 607–627. DOI: 10.1016/j.neunet.2021.07.021.
20. Shi M., Li L., Cao J., Hua L., Abdel-Aty M. Stability analysis of inertial delayed neural network with delayed impulses via dynamic event-triggered impulsive control. *Neurocomputing*. 2025. vol. 626. DOI: 10.1016/j.neucom.2025.129573.
21. He Z., Li C., Wu H., Nie L., Yu Z.. Exponential stability of continuous-time and discrete-time neural networks with saturated impulses. *Neurocomputing*. 2025. vol. 623. DOI: 10.1016/j.neucom.2025.129400.
22. Wang W., Zeng W., Chen W. New sufficient conditions on the global exponential stability of delayed inertial neural networks. *Neurocomputing*. 2025. vol. 622. DOI: 10.1016/j.neucom.2024.12930.
23. Brogan W.L. *Modern control theory*. 2nd edition. Englewood Cliffs. NJ: Prentice-Hall, 1985. 509 p.
24. Auge D., Hille J., Mueller E., Knoll A. A survey of encoding techniques for signal processing in spiking neural networks. *Neural Processing Letters*. 2021. vol. 53. pp. 4693–4710.
25. Осипов В.Ю. Интеллектуальная нейросетевая машина с функциями мышления. *Информатика и автоматизация*. 2024. Т. 23(4). С. 1077–1109.
26. Линник Ю.В. *Метод наименьших квадратов и основы математико-статистической теории обработки наблюдений*. 2-е изд. М., 1962. 336 с.
27. Айвазян С.А. *Прикладная статистика. Основы эконометрики*. Том 2. М.: Юнити-Дана, 2001. 432 с.
28. Yang B, Sun S, Li J, Lin X, Tian Y. Traffic flow prediction using LSTM with feature enhancement. *Neurocomputing*. 2019. vol. 332. pp. 320–327. DOI: 10.1016/j.neucom.2018.12.016.

Осипов Василий Юрьевич — д-р техн. наук, профессор, директор, Санкт-Петербургский институт информатики и автоматизации Российской академии наук, Федеральное государственное бюджетное учреждение науки «Санкт-Петербургский Федеральный исследовательский центр Российской академии наук» (СПб ФИЦ РАН). Область научных интересов: математическое моделирование, интеллектуальные системы, нейронные сети, информационная безопасность. Число научных публикаций — 170. osipov_vasiliy@mail.ru; 14-я линия В.О., 39, 199178, Санкт-Петербург, Россия; р.т.: +7(812)508-3311(1221).

V. OSIPOV

**STABILITY OF INTELLIGENT SIGNAL PROCESSING
BY STREAMING RECURRENT NEURAL NETWORKS WITH
CONTINUOUS LEARNING**

Osipov V. Stability of Intelligent Signal Processing by Streaming Recurrent Neural Networks with Continuous Learning.

Abstract. Intelligent neural network processing of various signals with continuous learning is of great scientific and practical interest. For such processing, along with other solutions, streaming pulse recurrent neural networks (RNNs) with advanced functionality are used. However, for these RNNs, the issues of ensuring their stable operation have not been studied in many respects. The goal is to increase the stability of streaming pulse RNNs by developing new methods of intelligent signal processing with continuous learning. For this purpose, the capabilities of these RNNs are clarified and approaches to ensuring their stability during training and signal generation are analyzed. The streaming RNN is formalized as a relatively finite operational automaton. A new method for stable intelligent signal processing by an improved RNN with continuous learning is proposed. Schemes for solving various intelligent problems of analysis and synthesis by the proposed method are considered. The modeling results are presented, confirming the operability of the proposed method and the possibility of increasing the stability of RNNs during continuous training and signal generation. The achievability of a stable balance between memorized and gradually forgotten information in the RNN with prompt adaptation to changing external conditions is shown.

Keywords: neural network, intelligence, stability, continuous learning, signal generation.

References

1. Man N., Osipov V., Zhukova N., Subbotin A., Ignatov D.I. Neural networks for intelligent multilevel control of artificial and natural objects based on data fusion: a survey. *Information Fusion*. 2024. vol. 110.
2. Osipov V.Yu. [Robustness of recurrent neural networks with controlled synapses]. *Informacionnye tekhnologii – Information Technology*. 2011. vol. 9. pp. 69–73. (In Russ.).
3. Osipov V., Osipova M. Space-time signal binding in recurrent neural networks with controlled elements. *Neurocomputing*. 2018. vol. 308. pp. 194–204.
4. Parisi G., Kemker R., Part J., Kanan C., Wermter S. Continual lifelong learning with neural networks: A review. *Neural Networks*. 2019. vol. 113. pp. 54–71.
5. Van de Ven G., Siegelmann H., Tolias A. Brain-inspired replay for continual learning with artificial neural networks. *Nature communications*. 2020. vol. 11. DOI: 10.1038/s41467-020-17866-2.
6. Tercan H., Deibert P., Meisen T. Continual learning of neural networks for quality prediction in production using memory aware synapses and weight transfer. *Journal of Intelligent Manufacturing*. 2022. vol. 33. pp. 283–292.
7. Haykin S. *Neural Networks and Learning Machines*, third ed., Prentice Hall, New York, 2008. 936 p.
8. Ehret B., Henning C., Cervera M., Meulemans A., Oswald J., Grewe B. Continual learning in recurrent neural networks. *ICLR 2021 Conference Paper*. 2021.
9. Schmidgall S., Ziaei R., Achterberg J., Kirsch L., Hajiseyedrazi P., Eshraghian J. Brain-inspired learning in artificial neural networks: A review. *APL Machine Learning*. 2024. vol. 2.

10. Krasovskij A.A. Spravochnik po teorii avtomaticheskogo upravleniya [Handbook of Automatic Control Theory]. M.: Nauka. Gl. red. fiz.-mat. Lit., 1987. 712 p. (In Russ.).
11. Rajchakit G., Agarwal P., Ramalingam S. Stability Analysis of Neural Networks. Springer, 2021. DOI: 10.1007/978-981-16-6534-9
12. Xie K., Zhang C., Lee S., He Y., Liu Y. Delay-dependent Lurie-Postnikov type Lyapunov-Krasovskii functionals for stability analysis of discrete-time delayed neural networks. *Neural Networks*. 2024. vol. 173.
13. Vaswani A., Shazeer N., Parmar N., Uszkoreit J., Jones L., Gomez A., Kaiser L., Polosukhin I. Attention is all you need. 31st Conference on Neural Information Processing Systems (NIPS 2017). 2017. pp. 6000–6010.
14. Osipov V. Nikiforov V., Zhukova N., Miloserdov D. Urban traffic flows forecasting by recurrent neural networks with spiral structures of layers. *Neural Computing and Applications*. 2020. vol. 32(18). pp. 14885–14897.
15. Kirkpatrick J., Pascanu R., Rabinowitz N., Veness J., Desjardins G., Rusu A., Milan K., Quan J., Ramalho T., Grabska-Barwinska A., Hassabis D., Clopath C., Kumaran D., Hadsell R. Overcoming catastrophic forgetting in neural networks. *arXiv preprint. arXiv 1612.00796v2*. 2017.
16. Zenke F., Poole B., Ganguli S. Continual learning through synaptic intelligence. *arXiv preprint. arXiv 1703.04200v3*. 2017.
17. Coop R., Arel I. Mitigation of catastrophic forgetting in recurrent neural networks using a fixed expansion layer. The 2013 International Joint Conference on Neural Networks (IJCNN). 2013. pp. 1–7. DOI: 10.1109/IJCNN.2013.6707047.
18. Duncker L., Driscoll L., Shenoy K., Sahani M., Sussillo D. Organizing recurrent network dynamics by task-computation to enable continual learning. *Advances in Neural Information Processing Systems 33 (NeurIPS 2020)*. 2020. vol. 33. pp. 14387–14397.
19. Cossu A., Carta A., Lomonaco V., Bacciu D. Continual learning for recurrent neural networks: An empirical evaluation. *Neural Networks*. 2021. vol. 143. pp. 607–627. DOI: 10.1016/j.neunet.2021.07.021.
20. Shi M., Li L., Cao J., Hua L., Abdel-Aty M. Stability analysis of inertial delayed neural network with delayed impulses via dynamic event-triggered impulsive control. *Neurocomputing*. 2025. vol. 626. DOI: 10.1016/j.neucom.2025.129573.
21. He Z., Li C., Wu H., Nie L., Yu Z.. Exponential stability of continuous-time and discrete-time neural networks with saturated impulses. *Neurocomputing*. 2025. vol. 623. DOI: 10.1016/j.neucom.2025.129400.
22. Wang W., Zeng W., Chen W. New sufficient conditions on the global exponential stability of delayed inertial neural networks. *Neurocomputing*. 2025. vol. 622. DOI: 10.1016/j.neucom.2024.12930.
23. Brogan W.L. Modern control theory. 2nd edition. Englewood Cliffs. NJ: Prentice-Hall, 1985. 509 p.
24. Auge D., Hille J., Mueller E., Knoll A. A survey of encoding techniques for signal processing in spiking neural networks. *Neural Processing Letters*. 2021. vol. 53. pp. 4693–4710.
25. Osipov V. [Intelligent Neural Network Machine with Thinking Functions]. *Informatics and Automation*. 2024. vol. 23(4). pp. 1077–1109. (In Russ.).
26. Linnik Yu.V. Metod naimen'shikh kvadratov i osnovy matematiko-statisticheskoy teorii obrabotki nablyudenij. 2-e izd. [The Method of Least Squares and Fundamentals of Mathematical-Statistical Theory of Observation Processing. 2nd ed.]. M., 1962. 336 p. (In Russ.).
27. Ajvazyan S.A. Prikladnaya statistika. Osnovy ekonometriki. Tom 2 [Applied Statistics. Fundamentals of Econometrics. Volume 2]. M.: Yuniti-Dana, 2001. 432 p. (In Russ.).

28. Yang B, Sun S, Li J, Lin X, Tian Y. Traffic flow prediction using LSTM with feature enhancement. *Neurocomputing*. 2019. vol. 332. pp. 320–327. DOI: 10.1016/j.neucom.2018.12.016.

Osipov Vasily — Ph.D., Dr.Sci., Professor, Director, St. Petersburg Institute for Informatics and Automation of the Russian Academy of Sciences, St. Petersburg Federal Research Center of the Russian Academy of Sciences (SPC RAS). Research interests: mathematical modeling, intelligent systems, neural networks, information security. The number of publications — 170. osipov_vasily@mail.ru; 39, 14th line V.O., 199178, St. Petersburg, Russia; office phone: +7(812)508-3311(1221).

М.Д. Долгушин, А.А. Карпов
**АНАЛИТИЧЕСКИЙ ОБЗОР РЕЧЕВЫХ
И МНОГОМОДАЛЬНЫХ МЕТОДОВ РАСПОЗНАВАНИЯ
КОГНИТИВНЫХ НАРУШЕНИЙ ЛЮДЕЙ**

Долгушин М.Д., Карпов А.А. Аналитический обзор речевых и многомодальных методов распознавания когнитивных нарушений людей.

Аннотация. В течение последнего десятилетия наблюдается рост количества публикаций научно-технической и медицинской направленности, посвящённых автоматическому определению на основании речевого и визуального материала таких когнитивных нарушений людей, которые возникают при таких нейродегенеративных заболеваниях, как деменция, болезнь Альцгеймера, болезнь Паркинсона и другие нарушения. Хотя данные заболевания имеют высокую степень распространения и являются одной из основных причин смертности и ранней инвалидизации людей, способов их эффективного лечения на данном этапе развития медицины отсутствуют или ограничены. В связи с этим ранняя диагностика симптомов и их облегчение вызывают значительный интерес исследователей. В фокусе современных исследований находится задача разработки автоматизированных и автоматических систем, основанных на количественных объективных методах, нейросетевых подходах, использующих различные модальности и их объединение, а также на методах интерпретируемого искусственного интеллекта. В статье представлен обзор и анализ основных исследований с 2022 года, посвящённых автоматическому одно- и многомодальному определению когнитивных нарушений людей. Представлены наиболее часто применяющиеся многомодальные корпуса, применяемые в данной задаче, такие как ADReSS, ADReSSo, TAUkADIAL и др. Описаны современные методы, используемые для выявления когнитивных нарушений на основе данных различных модальностей, представленные на международном соревновании TAUkADIAL-2024, и вне соревнований. По результатам соревнований, лучшими методами распознавания когнитивных нарушений являются ансамбли вероятностных моделей, обученные на интерпретируемых экспертных признаках и нейросетевых признаках текстов и аудио. Рассмотрены многомодальные подходы, использующие визуальную модальность для обучения глубоких нейросетевых моделей. Рассмотрено новое направление, изучающее применимость больших языковых моделей к анализу медицинских текстов и интерпретируемому предсказанию заболевания. Проведена систематизация методов извлечения информативных признаков и применяемых классификаторов. По результатам проведенного обзора сформулированы требования к системам автоматизированного определения когнитивных нарушений.

Ключевые слова: автоматическое определение когнитивных нарушений, речевые технологии в обеспечении здравоохранения, интерпретируемый искусственный интеллект, машинное обучение.

1. Введение. Сфера интеллектуальных речевых технологий в обеспечении здравоохранения и благополучия людей (healthcare and wellbeing) возникла сравнительно недавно как часть более широкой области применения методов искусственного интеллекта в медицинской сфере, но она уже представляет одно из самых

стремительно развивающихся междисциплинарных направлений науки.

В настоящее время проблема раннего медицинского определения когнитивных нарушений становится особенно актуальной. По оценкам Всемирной организации здравоохранения (ВОЗ) за 2021 год¹, деменцией, крайней стадией развития когнитивных нарушений, страдают более 55 миллионов человек (8,1% женщин и 5,4% мужчин старше 65 лет), а болезнью Паркинсона – свыше 8,5 миллионов человек. Кроме того, специалисты ВОЗ прогнозируют рост данных показателей в ближайшие годы. Методы эффективного лечения данных заболеваний на данный момент отсутствуют, однако своевременное определение заболевания и раннее начало терапии способны облегчить симптомы и улучшить качество жизни пациентов.

Несмотря на большой интерес научных сообществ, многие задачи в области определения когнитивных нарушений до сих пор остаются не решенными в силу значительной сложности точной диагностики [1]. Специалисты определяют наличие когнитивных нарушений путем различных опросников и тестирований в формате интервью, но эффективность многих из них зависит от множества факторов, поэтому без использования инвазивных средств точное обнаружение заболевания крайне затруднено.

Активное применение в исследованиях по автоматическому определению когнитивных нарушений находят многомодальные методы [2], то есть использующие несколько модальностей одновременно, в том числе естественную речь и звуки, артикуляцию лица, жесты рук и тела, направление взгляда, общее поведение, магнитно-резонансную томографию, дополнительную медицинскую информацию и т.д. Заболевания, вызывающие когнитивные нарушения, оказывают влияние на весь комплекс поведенческих признаков, но отдельные наборы данных, в основном, одномодальны [3], то есть содержат только речевой материал, видеок cadры или снимки МРТ.

В данной работе решено сосредоточиться на речевых наборах, содержащих аудио или аудио и видео, поскольку разговорная речь выступает одним из индикаторов психоэмоционального состояния, например, в работах [4, 5], и, как следствие, может быть использована для автоматизированного обнаружения когнитивных отклонений. Речь многомодальна и соответственно позволяет использовать методы анализа аудио, текста и, при наличии, видео. Также данный метод имеет ряд преимуществ: неинвазивность (отсутствие необходимости

¹ <https://www.who.int/ru/news/item/02-09-2021-world-failing-to-address-dementia-challenge>

проникать в организм при исследовании) и доступность, связанная с широкой распространенностью средств аудио- и видеозаписи. Подобные методы позволяют, например, использовать телемедицинские технологии для исследования состояния пациента, что может быть крайне важным в случаях ограниченной мобильности, географической удалённости или по иным причинам.

Данный анализ проведен с целью продемонстрировать значимость существующих проблем в области автоматического определения когнитивных нарушений, описать используемые в настоящее время подходы и провести обзор существующих систем определения когнитивных нарушений.

В разделе 2 приводятся определения в области когнитивных расстройств и обсуждается влияние заболеваний на вербальные и невербальные аспекты речевой коммуникации. Раздел 3 посвящён описанию существующих одно- и многомодальных речевых баз данных по когнитивным нарушениям людей. Одно- и многомодальные системы, представленные в работах, опубликованных с 2022 года, описаны в разделах 4 и 5. В разделе 6 рассматриваются работы, изучающие применение больших языковых моделей для извлечения текстовых признаков и определения когнитивных нарушений. Раздел 7 содержит классификацию наиболее эффективных методов для построения автоматических систем определения когнитивных нарушений. Итоговые выводы по результатам проведённого исследования представлены в разделе 8.

2. Основные определения и проявления когнитивных нарушений в речи. В ходе определения ключевых понятий в качестве основного ориентира использованы рекомендации Министерства здравоохранения РФ. Из рекомендаций [6] следует, что когнитивное расстройство – это субъективное и/или объективно выявляемое ухудшение когнитивных функций (внимания, памяти, речи, восприятия и т.д.) по сравнению с исходным индивидуальным и/или среднестатистическим возрастным и образовательным уровнем. Распространенность умеренных когнитивных нарушений (УКН), по данным международных исследований, среди лиц старше 60 лет варьируется от 12 до 18% [7, 8]. Кроме того, распространенность УКН увеличивается с возрастом: с 10% в возрасте 70–79 лет до 25% у лиц в возрасте 80–89 лет [9]. В России масштабные эпидемиологические исследования когнитивных расстройств и деменции не проводились, однако в одном из последних исследований [10] отмечается, что среди пациентов амбулаторно-поликлинического звена 60 лет и старше

широко распространены как когнитивные расстройства, так и деменция: частота УКН составила 49,6%, деменции – 7,8%.

Крайнее проявление когнитивных нарушений – деменция. Деменция [6] определяется как нейропсихиатрический синдром, в структуре которого лежит приобретенное, длительное, клинически значительное ухудшение когнитивных функций, проявляющееся в виде тяжелых когнитивных нарушений и обуславливающее утрату привычного функционирования различной степени тяжести.

Потенциальными рисками для развития УКН и деменции являются болезнь Альцгеймера, болезнь Паркинсона, депрессия, посттравматическое стрессовое расстройство (ПТСР), хронический стресс и тревожность, боковой амиотрофический склероз, рассеянный склероз и др. [11]. По данным ВОЗ, депрессия выступает одним из наиболее распространенных психических расстройств человека. При этом она также считается одним из важных немоторных симптомов болезни Паркинсона и Альцгеймера, раннее обнаружение и своевременное лечение которых может значительно улучшить качество жизни пациентов [12].

Поскольку когнитивные расстройства оказывают разрушительное воздействие на мышление и моторику, расхождения на различных лингвистических уровнях языка и в паралингвистических проявлениях становятся заметны, что обуславливает возможность их автоматического обнаружения посредством анализа речевого сигнала [13].

На рисунке 1 представлена схема, иллюстрирующая связь некоторых нервно-психических заболеваний, вызывающих когнитивные нарушения с механизмами, через которые они оказывают влияние на речевой сигнал. Схема также сопровождается примерами акустических, лингвистических и поведенческих признаков, позволяющих обнаруживать подобные отклонения. Эта схема основана на систематизации из работы [14] и дополнена признаками визуальной модальности (мимики), рассмотренными в исследовании [15]. Более полные медицинские обзоры изменений дискурса и голосовых признаков при различных нервно-психических патологиях представлены в работах [16, 17]. Подробный обзор автоматически извлекаемых лексических и акустических признаков, применяемых для определения когнитивных нарушений и формального расстройства мышления, представлен в исследовании [18].

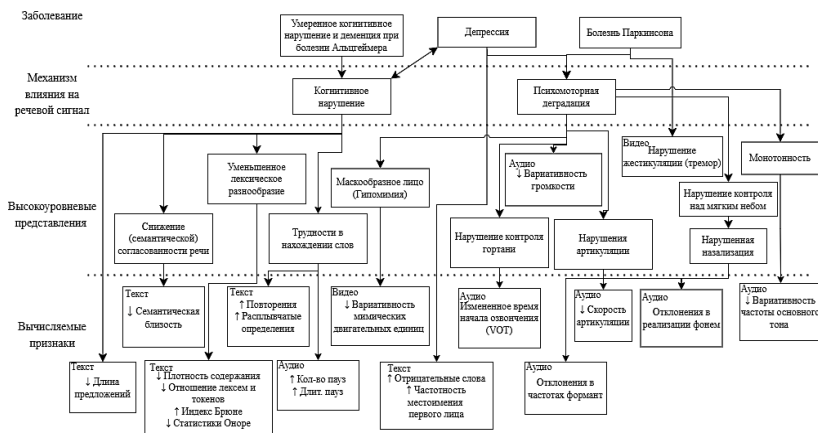


Рис. 1. Систематизация когнитивных заболеваний, нарушений и информативных признаков

Из представленной схемы видно, что влияние некоторых болезней на речевой сигнал частично пересекается. Например, депрессия и болезнь Паркинсона ассоциированы с психомоторной деградацией и, следовательно, имеют схожие проявления в речевом сигнале. Помимо этого, некоторые признаки речевых отклонений могут одновременно характеризовать различные диагнозы, поэтому использование отдельных признаков часто не является достаточным и эффективным для автоматического определения конкретных болезней. Например, депрессия может быть фактором риска для развития деменции при болезнях Альцгеймера и Паркинсона [19]. Их влияние на речевой сигнал также может быть сходным [20]. В работе [21] отмечается, что у пациентов на поздних стадиях развития болезни Альцгеймера сильно выражена апатия, что влияет на снижение общей эмоциональной выразительности мимики.

Особый интерес на схеме представляют высокоуровневые представления (согласованность, беглость речи, эмоциональное состояние и другие). С одной стороны, некоторые из этих представлений очевидны даже для неспециалистов и могут быть установлены, например, родственниками пациента. С другой стороны, большинство данных признаков, в отличие от вычисляемых признаков, не поддаются оценке с помощью количественных показателей. В связи с этим представляется перспективной задача автоматического извлечения данных признаков, например, посредством глубоких нейросетевых методов, в дополнение к уже известным

количественным признакам. Такой подход мог бы упростить интерпретацию полученных результатов, поскольку признак сам по себе был бы достаточно ясным для интерпретации.

Исходя из вышесказанного, задачи автоматического определения отдельного заболевания, являющегося причиной когнитивных нарушений, а также интерпретации автоматически обнаруженных нарушений посредством выявленных корреляций с высокоуровневыми представлениями нарушений, являются актуальными направлениями исследований.

3. Многомодальные базы данных для исследования когнитивных нарушений. Для создания автоматической системы определения возможного заболевания необходимо опираться на речевые данные как пациентов с установленным диагнозом, так и контрольной группы (КГ) без диагноза. В работах, которые рассмотрены в данном разделе, представлены одно- и многомодальные корпуса, содержащие устную речь. Большинство баз данных содержат записи пациентов с УКН и/или деменцией одного определенного типа, чаще всего – болезни Альцгеймера. Однако во многих наборах данных конкретное заболевание, являющееся причиной когнитивных нарушений, не указывается.

Во многих из представленных наборов содержится информация о показателях психологических тестов когнитивных способностей, таких как:

- Краткая шкала оценки психического статуса (Mini Mental State Examination, MMSE²).
- Монреальская шкала оценки когнитивных функций (Montreal Cognitive Assessment, MoCA³).
- Унифицированная шкала оценки болезни Паркинсона Международного общества расстройств движений (MDS-UPDRS⁴).
- Шкала депрессии А.Бека (beck depression inventory, BDI⁵).

В таблице 1 представлена характеристика наиболее распространённых наборов данных, используемых в исследованиях когнитивных нарушений людей.

DementiaBank [22] – крупнейшая опубликованная многомодальная база данных речевых записей пациентов, страдающих деменцией (в основном – УКН и БА, также присутствуют данные по

² <https://mmse.neurol.ru/ru/index.html?ysclid=m92u2aqpfj497146286>

³ <https://mocacognition.com>

⁴ https://mds.movementdisorders.org/UserFiles/file/MDS-UPDRS_Russian.pdf

⁵ <https://www.apa.org/pi/about/publications/caregivers/practice-settings/assessment/tools/beck-depression>

БП и другим нарушениям). База данных содержит несколько наиболее часто встречающихся в исследованиях англоязычных корпусов и корпусы на других языках: испанский, китайский и т.д. Большинство представленных корпусов содержат аудиозаписи и текстовые транскрипции.

Pitt [23] – один из крупнейших англоязычных корпусов речевых записей пациентов с УКН и БА из базы DementiaBank. В ходе работы над созданием корпуса респондентам предлагалось выполнить задачи по описанию изображения, оценке владения языком и пересказу истории. Записи, входящие в состав корпуса, не сбалансированы по полу и возрасту, а также характеризуются различными акустическими условиями записи.

Таблица 1. Сравнение корпусов, содержащих данные людей с когнитивными нарушениями

Набор данных	Язык данных	Статистика Информантов **	Тип данных ***	Оценка	Доступность
Pitt	Англ.	104 – КГ, 208 – УКН/БА, 85 – НД	А, Т, М	MMSE	По запросу [23]
ADReSS	Англ.	78 – Д, 78 – КГ	А, Т, М	MMSE	По запросу [24]
ADReSSo for AD classification	Англ.	122 – Д, 115 – КГ	А, М	MMSE	По запросу [26]
TAUKADIAL	Кит.	98 – УКН/Д, 102 – КГ	А, М	MMSE	По запросу [27]
	Англ.	123 – УКН/Д, 63 – КГ			
NCMMSC's AD dataset	Кит.	26 – БА, 44 – КГ, 54 – УКН	А, М	Метка диагноза	Открыт [28]
GAADRД	Греч.	25 – КГ, УКН и БА	А, В, М	Метка диагноза	По запросу [29]
PROMPT	Япон.	60 часов 117 КГ, Депр., БРЛ, Д	А, В, ТД	MINI, MMSE	По запросу [30]
I-CONNECT	Англ.	92 – УКН, Д	А, В	MMSE	По запросу [31]
Набор ВШЭ	Русск.	100 – УКН и КГ	А, Т, М	MoCA, MMSE	Закрыт [32]

* Англ. – английский, Кит. – китайский, Русск. – русский, Греч. – греческий, Япон. – японский. ** БА – болезнь Альцгеймера, БП – болезнь Паркинсона, БРЛ – биполярное расстройство личности, Д – деменция, Депр. – депрессия, КГ – контрольная группа, НД – неизвестный диагноз, РС – рассеянный склероз, УКН – умеренное когнитивное нарушение. ** А – аудио, В – видео, М – метаданные, Т – Текст, ТД – трекинг движений

ADReSS (Alzheimer's Dementia Recognition through Spontaneous Speech) Challenge Dataset [24] – это стандартизированный набор данных, являющийся подкорпусом корпуса Pitt. ADReSS содержит речевые записи и транскрипции речи устных описаний изображений участниками. Аннотация была проведена с использованием системы кодирования CHAT [25]. Для всех записей была произведена очистка от шума и нормализация по громкости, поскольку многие из записей были получены в разных акустических условиях. Набор был собран с целью решения двух задач: классификации речи для определения наличия или отсутствия заболевания и регрессионной задачи предсказания результата MMSE по аудио и тексту. В 2021 была представлена расширенная версия данного набора – ADReSSo [26] (Alzheimer's Dementia Recognition through Spontaneous Speech Only), содержащая аудиоданные без ручной расшифровки, а также дополнительную выборку, которую можно использовать для предсказания когнитивного снижения с течением времени.

TAUKADIAL Challenge Dataset [27] – сбалансированный многоязычный набор данных, содержащий аудиозаписи спонтанной речи на английском и китайском когнитивно здоровых респондентов и пациентов с УКН. Набор сбалансирован по полу и возрасту с целью минимизации возможных искажений.

NCMMSC's (National Conference on Man-Machine Speech Communication) AD dataset [28] – набор данных, использованный в соревновании NCMMSC2021 AD Recognition Challenge в 2021 году. Набор содержит речь 124 носителей китайского языка, среди которых есть пациенты с когнитивными отклонениями и здоровые люди. Респонденты выполняли задания по описанию изображений и оценке владения языком, а также вели свободный диалог с интервьюером.

GAARD (Greek Alzheimer's Association for Dementia and Related Disorders) [29] – коллекция наборов данных, собранных в рамках проекта Dem@Care, содержащих аудио, видео и физиологические показатели пациентов с диагностированной деменцией, являющихся носителями греческого языка. Записи производились в Греции в лабораторных или домашних условиях. В проекте приняли участие 25 респондентов с различной степенью тяжести когнитивных или поведенческих нарушений. Участникам предлагалось выполнять обычные повседневные дела, такие как чтение, разговор по телефону и т.д., пока их движения записывались на видео, а состояние отслеживалось с помощью различных датчиков.

PROMPT (Project for Objective Measures using computational Psychiatry Technology) dataset⁶ [30] – аудиовизуальный речевой корпус на японском языке, содержащий записи людей старше 20 лет. Среди респондентов есть лица с депрессивными эпизодами и когнитивными нарушениями. Респонденты из контрольной группы и пациенты проходили международное нейропсихиатрическое мини-интервью (The Mini-International Neuropsychiatric Interview, M.I.N.I.) и оценивание по шкале MMSE. Видеозаписи представляют собой десятиминутные интервью респондента с медицинским специалистом на общие темы. Интервью с одними и теми же респондентами проводились до десяти раз в течение нескольких лет для отслеживания возможных изменений с течением времени. В результате суммарная длительность записей набора составила примерно 60 часов.

Internet-Based Conversational Engagement Clinical Trial (I-CONNECT) data set⁷ [31] – аудиовизуальный речевой корпус на английском языке, содержащий записи людей старше 75 лет с УКН. Для исследования на основании критерия качества записи были отобраны 92 видео, в которых респонденты говорят на одну из трех тем: лето, праздники и забота о себе. Набор сбалансирован, половина респондентов имеет умеренные когнитивные нарушения.

Набор ВШЭ [32] содержит 100 спонтанных рассказов, полученных от 100 русскоязычных пожилых людей с УКН и без них. В записях присутствует речь интервьюера. Записи сопровождаются полной текстовой разметкой, выполненной вручную, а также разметкой речевых сбоев и пауз хезитации. Каждому респонденту сопоставлена оценка по шкалам MMSE и MoCA, утверждённая врачами-специалистами. Данный набор является одним из немногих наборов на русском языке по умеренным когнитивным нарушениям, однако он закрыт для публичного использования и крайне несбалансирован по полу и возрасту, что потенциально усложняет применение методов машинного обучения.

Существуют также наборы данных по заболеваниям, связанным с когнитивными изменениями, таким как БП, депрессия и боковой амиотрофический склероз. Это, например, открытый сбалансированный аудиовизуальный корпус In-the-wild Speech Medical Corpus (WSM) [33], содержащий подкорпусы по болезни Паркинсона и депрессии. Данный корпус был собран из видеозаписей, выложенных на сайте YouTube, и аннотирован с помощью краудсорсинга. Другой пример – PC GITA [34], сбалансированный медицинский речевой

⁶ <https://www.i2lab.info/>

⁷ <https://www.i-connect.org/request-data>

корпус на испанском языке, посвящённый БП и часто используемый в исследованиях, а также корпуса Distress Analysis Interview Corpus (DAIC) [35] и Extended Distress Analysis Interview Corpus (E-DAIC) [36] – многомодальные коллекции клинических интервью с пациентами с депрессией и посттравматическим синдромом.

Кроме того, следует упомянуть о наборе голосовых данных пациентов с боковым амиотрофическим склерозом [37], записанном в БГУИР, Минск, Беларусь. Он полностью открыт для использования⁸ и хорошо изучен в контексте применения методов обработки сигналов для точного определения речевых нарушений. Однако применение данного набора для изучения когнитивных нарушений ограничено в связи с его небольшим размером и отсутствием разговорной речи.

Подводя итоги раздела, можно выделить несколько проблем речевых наборов по когнитивным нарушениям:

1) Многие речевые наборы данных являются закрытыми, либо предоставляют только отдельные поведенческие сигналы, часто только аудиосигналы, что приводит к необходимости использовать нейросетевые методы для автоматического извлечения дополнительных модальностей, например, автоматическое распознавание речи.

2) Крайне малочисленны наборы данных, которые можно использовать для автоматического определения сразу нескольких заболеваний.

3) Малое число наборов на естественных языках, отличных от английского и китайского, что ограничивает исследование кросс-языковых методов.

4) Отсутствуют единообразные протоколы сбора речевых и аудиовизуальных данных по когнитивным нарушениям, что объясняется значительно различающимися условиями записи.

5) Существующие наборы данных, за редким исключением, имеют дисбаланс в количестве экземпляров в классах обучающих данных (редки данные с когнитивными нарушениями), а также могут содержать посторонние шумы, голос интервьюера и т.д.

Все это значительно затрудняет создание систем, основанных на глубоких нейросетевых моделях и аудиовизуальных многомодальных подходах.

4. Обзор работ по многоязычному определению когнитивных нарушений в речи, представленных на соревновании TAUkADIAL. В опубликованных ранее аналитических обзорах [38, 39] были рассмотрены работы по неинвазивному

⁸ <https://github.com/Mak-Sim/Troparion/tree/master/SPA2019>

автоматическому одно- и бимодальному определению когнитивных нарушений по аудио и тексту. Авторы данных обзоров подробно проанализировали основные признаки, методы и модели, позволившие получить наилучшие результаты на международных соревнованиях ADReSS, ADReSSo, NCMMSC's и ADReSS-M. Однако наиболее актуальное соревнование TAUkADIAL по многоязычному определению когнитивных нарушений, проводившееся в рамках международной конференции INTERSPEECH-2024, не было рассмотрено.

В ходе соревнования TAUkADIAL 2024 [27] участникам предлагалось разработать систему для автоматического кросс-языкового определения когнитивных нарушений на основе аудиоданных на английском и китайском языках. Данным на обоих языках предоставлялась метка диагноза для автоматической классификации и метка по значению шкалы MMSE для решения регрессионной задачи. Аудиозаписи не были сопровождаемы текстовыми транскрипциями. Лучшие результаты работы систем, представленных на соревновании, отображены в таблице 2.

Таблица 2. Сравнение результатов систем на соревновании TAUkADIAL 2024

Работа	Задача*	Модальность	Набор признаков	Модель	Значения показателей val/test
Luz S. et al. базовая работа [27]	Класс.	Аудио	wav2vec и eGeMAPS	MLP	UAR = 50,9/59,2 %
	Регр.	Текст (Google ASR)	Лингвистические	MLP	RMSE = 2,86/2,89
Duan J. et al. [41]	Класс.	Аудио и текст (Whisper large v3)	Prosody UnitY2 и XLM-RoBERTa-base	Трансформер	UAR = -/77,6 %
Favaro A. et al. [42]	Класс.	Аудио и текст (Whisper large v2 с промптом)	text2vec-base-multilingual, LaBSE, Lealla, Whisper	Ансамбль – XGC, SVC, BAGC, MLP	UAR = 70,0/75,0 %
	Регр.		Лингвистические, темпоральные, Multilingual-E5-Large, XLM-Roberta-Large-Vit-L-14	Ансамбль – XGR, SVR, BAGR, MLP	RMSE = 3,06/2,44
Hoang B. et al. [45]	Регр.	Аудио и текст (Whisper large v2 с промптом)	MFCC, BERT с аугментацией, метка языка Whisper	SVR	RMSE = 2,705/2,578
Pérez-Toro P.A. et al. [47]	Класс.	Аудио	Whisper и темпоральные	MLP	UAR = 80,0/83,3 %
	Регр.	Аудио и текст (Amazon AWS ASR)	BERT, темпоральные и Whisper	SVR	RMSE = 2,34/1,87

* Класс. – классификация КН; Регр. – Регрессия MMSE

Основным показателем точности классификации выступала невзвешенная средняя полнота (unweighted average recall, UAR). Результаты регрессии оценивались с использованием показателя корня из среднеквадратической ошибки (root mean squared error, RMSE).

В базовой системе [25], представленной на соревновании TAUKADIAL 2024, в первую очередь проверялась гипотеза возможности автоматического многоязычного или языконезависимого определения когнитивных нарушений. На рисунке 2 представлена схема универсальной архитектуры, предложенной в данной работе. Из аудиосигнала авторы выделяли 88 акустических признаков с помощью стандартизированного набора признаков eGeMAPS, включенного в программный инструмент OpenSMILE⁹. Также авторы извлекали вектор нейросетевых представлений из аудио с помощью предобученной глубокой нейросетевой модели Wav2Vec [40]. Набор лингвистических признаков формировался на основе результатов автоматического распознавания речи и автоматической разметки частей речи и включал следующие, нормализованные по z-мере, признаки: количество лексем и слов различных частей речи, коэффициент лексического разнообразия, плотность (соотношение слов различных частей речи к суммарному числу лексем) и др. Для классификации и регрессии использовался многослойный перцептрон. В результате лучшая точность классификации метки диагноза по показателю UAR составила 50,9% на валидационной выборке и 59,2% на тестовой выборке при объединении классификаторов, обученных на признаках wav2vec и eGeMAPS. Наивысшая точность предсказания значения MMSE по показателю RMSE оставила 2,86 и 2,89 на валидационной и тестовой выборках соответственно при использовании только лингвистических признаков. Другие сочетания признаков также рассматривались в исследовании, но не оказали положительного эффекта на результирующую точность классификации или регрессии.

В работе [41], в отличие от остальных методов, достигших высоких результатов, использовалось многомодальное объединение нейронных представлений посредством 6-слойной модели трансформера. Также авторы использовали автоматическое сопоставление аудио и текста посредством самообучения (self-supervised learning, SSL) для аугментации и объединения модальностей. На этапе предобработки авторы использовали автоматическое определение языка с помощью Whisper large-v3 и распознавали текст с использованием этой же модели, предварительно

⁹ <https://www.audeering.com/research/opensmile/>

указывая промпт для сохранения речевых сбоев в записях. Для извлечения текстовых признаков авторы использовали предобученную нейросетевую модель XLM-RoBERTa-base. Для извлечения аудиопризнаков было предложено использовать модель перевода речи в латентное пространство с учетом интонации Prosody UnitY2¹⁰. Модель, использующая объединение модальностей и аугментацию сопоставлением аудио и текста, достигла UAR = 77,6%, продемонстрировав тем самым один из лучших результатов на соревновании TAUkADIAL.



Рис. 2. Архитектура базовой системы определения когнитивных нарушений с помощью языконезависимых признаков по речи [27]

¹⁰ https://github.com/facebookresearch/seamless_communication

В исследовании [42] авторы сделали значительный акцент на поиск оптимальных сочетаний интерпретируемых и неинтерпретируемых признаков. В базовый процесс предобработки были внесены следующие изменения: аудиозаписи приводились к частоте дискретизации 16 кГц и нормализовались по корню из среднеквадратической громкости (root mean squared loudness); для автоматического определения языка использовалась нейросетевая модель Whisper large-v2 [43], которая также использовалась для автоматического распознавания, но с добавлением промптов на английском и китайском языках для сохранения значимых для медицинского контекста речевых сбоев. Авторы исследовали следующие интерпретируемые признаки:

- признаки темпа на основе детектора речевой активности (voice activity detection, VAD) из инструментария py-webrtcvad и библиотеки анализа интонационных признаков DigiPsych Prosody;
- акустические признаки из библиотеки Simple Speech Features¹¹;
- текстовые признаки, извлекаемые с помощью инструментария Linguistic Feature Toolkit¹² [44].

К неинтерпретируемым рассмотренным признакам относятся:

- мел-частотные кепстральные коэффициенты (Mel-frequency Cepstral Coefficients, MFCC), включенные в набор признаков ComParE библиотеки OpenSMILE;
- нейросетевые векторы акустических моделей Wav2vec и Whisper, усредненные по временной оси для каждой записи для получения одномерного вектора;
- нейросетевые векторы больших языковых моделей (Large Language Model, LLM).

Лучшие результаты классификации были получены с применением только моделей, обученных на нейросетевых признаках текстовых моделей и Whisper, и составили 70% и 75% по показателю UAR на валидационной и тестовой выборках соответственно. Лучшие результаты при решении задачи регрессии показал ансамбль классификаторов на основе бэггинга, градиентного бустинга и метода опорных векторов, обученных на экспертных и нейросетевых признаках, составив 3,06 и 2,44 по RMSE на валидационной и тестовой выборках соответственно. Использование только интерпретируемых лингвистических и темпоральных признаков с теми же регрессорами,

¹¹ <https://github.com/uzaymacar/simple-speech-features>

¹² <https://lftk.readthedocs.io/en/latest/>

что и в лучшем подходе, также показало улучшение результатов по сравнению с базовыми значениями, незначительно уступающее лучшему результату на тестовой выборке – 3,06 и 2,62 RMSE для валидационной и тестовой выборок соответственно.

В исследовании [45] авторы предположили, что можно успешно применить методы аугментации текста посредством автоматического перевода для решения проблемы ограниченности размеров обучающей выборки и обеспечения универсальности модели по двум языкам. Для аугментации язык записи определялся автоматически с применением модели Whisper large v2 и автоматически распознавался этой же моделью. Для перевода авторы использовали многоязычную нейросетевую модель M2M100 [46], после чего из данных извлекались нейросетевые признаки с помощью англоязычной или многоязычной модели BERT. Также авторы извлекали акустические признаки из аудиозаписей с помощью набора признаков eGeMAPS из библиотеки OpenSMILE. Для уменьшения размерности признаков использовался метод LASSO. Для регрессии авторы использовали – регрессию методом опорных векторов и случайного леса. Полученные результаты предсказания метки RMSE превзошли базовые и составили 2,705 и 2,578 RMSE на валидационной и тестовых выборках соответственно.

В работе [47] были продемонстрированы наилучшие результаты в рамках соревнования TAUADIAL как по задаче классификации когнитивных нарушений, так и по задаче предсказания шкалы MMSE. На этапе предобработки данных авторы обнаружили, что многие англоязычные записи могут содержать речь интервьюера, а не только респондента, из-за чего было принято решение использовать функции VAD, определения говорящего, определения языка и автоматического распознавания речи, предлагаемые коммерческим решением Amazon web services (AWS) transcribe¹³. В качестве признаков использовались нейросетевые признаки, многоязычных языковых моделей RoBERTa и BERT, а также векторы акустических моделей Whisper и Wav2Vec XLSR-53. Также авторы предположили значимость использования данных по темпу в данной задаче, поэтому в дополнение к нейросетевым признакам рассчитывались 16 темпоральных признаков на основе VAD, такие как число пауз и речевых сегментов в секунду, соотношение числа речевых сегментов и пауз, средняя длительность пауз и речевых сегментов и т.д. Для решения задачи классификации авторы рассматривали многослойный персептрон и метод опорных

¹³ <https://www.amazonaws.cn/en/transcribe/>

векторов. Для задачи предсказания значения MMSE рассматривались линейная регрессия и метод опорных векторов. Наилучшие результаты классификации были получены при использовании совмещенного вектора темпоральных признаков и скрытого слоя Whisper для обучения многослойного персептрона, а также использования неполного набора данных для обучения, включавшего только задачи на описание изображений. Значения UAR составили 80,0% и 83,3% на валидационной и тестовой выборках соответственно. Наилучшие результаты по регрессии составили 2,34 и 1,87 по показателю RMSE и были получены при использовании совмещенного вектора BERT, темпоральных признаков и скрытого слоя Whisper для обучения регрессора методом опорных векторов на неполном наборе данных, содержащим только первую подвыборку описаний изображений.

Следует отметить, что в большинстве работ, представленных на соревнованиях, были использованы неглубокие нейросетевые классификаторы или вероятностные модели (деревья решений, метод опорных векторов, градиентный бустинг). Это обусловлено сравнительно малыми размерами исходных данных. Также отмечается интерес к поиску новых сочетаний интерпретируемых, в частности статистических характеристик темпа, и неинтерпретируемых признаков. Несмотря на то, что во многих случаях методы, основанные на нейросетевых неинтерпретируемых признаках, показывают более высокие результаты по сравнению с интерпретируемыми аналогами, отставание интерпретируемых методов по точности не настолько значительно.

5. Обзор работ по визуальному и многомодальному определению когнитивных нарушений. В этом разделе предлагается анализ исследований, выполненных вне соревнований и включающих методы, основанные на визуальной модальности, менее распространенные в данной области в силу ограниченного числа аудиовизуальных корпусов. В таблице 3 представлены основные работы, выполненные вне соревнований и включающие исследования по применению видео и многомодальных (аудио, текст, видео и др.) методов. Точность моделей в таблице приводится преимущественно по показателю точности (Assuracy) в целях обеспечения сопоставимости результатов, также в одном случае используется значение показателя площади под кривой (area under curve, AUC) для оценки модели, поскольку авторы не привели значение точности.

Таблица 3. Сравнение результатов систем, основанных на видео и многомодальных признаках

Работа	Набор данных	Модальность	Набор признаков	Модель	Точность
Zheng C. et al. 2023 [48]	PROMPT	Видео	Коды лицевых движений	SVM	0,71
			Гистограммы направленных градиентов	LSTM	0,79
Okunishi T. et al. 2025 [49]	PROMPT	Видео	Коды лицевых движений, классы эмоций, валентность возбуждения и лицевые эмбединги	Дерево решений	0,93 (AUC)
Mu X. et al. 2024 [50]	I-CONNECT	Аудио, видео, текст (Whisper)	LLAMA-65B, RoBERTa Sentiment, WavLM, DINOv2, лицевые эмоции, коды лицевых движений, rPPG, акустические и демографические признаки	Позднее объединение детерминированных моделей голосованием большинством	0,58
		Аудио, текст (Whisper)	LLAMA-65B, RoBERTa Sentiment, WavLM, акустические и демографические признаки	Позднее объединение детерминированных моделей голосованием большинством	0,63

В работе [48] представлено одно из первых систематических исследований по автоматическому определению когнитивных нарушений с использованием видеомодальности. Авторы исследовали использование кодов лицевых движений с классификацией методом опорных векторов SVM, а также главные компоненты по лицевым полигональным сеткам и гистограммам направленных градиентов при обучении модели с долгой краткосрочной памятью (long-short term memory, LSTM) в применении к набору данных PROMPT на японском языке. Наилучший результат классификации с точностью 78% был получен при использовании гистограмм направленных компонент, хотя авторы отмечают, что данный подход может быть подвержен влиянию освещения на изображении. Использование кодов лицевых движений также продемонстрировало относительно высокую точность, составившую 71% процент. Кроме того, установлено, что удаление незначимых кодов лицевых движений не оказало

существенного влияния на результат. В целом исследование подтвердило возможность использования видеомодальности для выявления когнитивных нарушений.

В работе [49] в качестве признаков для обучения дерева решений авторы использовали коды лицевых движений и автоматически определенные классы эмоций, вычисленные с помощью библиотеки OpenFace, а также метки валентности-включенности и нейросетевые эмбединги. Результирующее значение $AUC = 0,93$. При этом использование дерева решений позволило сохранить интерпретируемость результатов относительно значимых признаков.

В работе [50] представлен многомодальный подход к определению когнитивных нарушений и предсказанию оценок психологического благополучия на материале корпуса I-CONNECT. Архитектура предложенной системы представлена на рисунке 3. В исследовании используется позднее объединение модальностей [51]. Признаки различных модальностей, в том числе демографические (раса, пол, возраст), обрабатываются отдельными моделями: с помощью бинарных классификаторов, логистической регрессии и/или классификаторов с градиентным усилением для классификации шкал, после чего производится объединение на уровне меток классификаторов посредством голосования большинством. Лучшая точность определения КН составила 63% без использования данных лицевой модальности и сердечно-сосудистых признаков. При объединении всех признаков точность снизилась до 58%. В задаче предсказания значения шкалы MoCA модель также показала не самые высокие результаты, достигнув максимальной точности в 58% при использовании всех модальностей. Однако в задаче предсказания значений шкалы клинической деменции (CDR, Clinical Dementia Rating) точность модели составила 74%.

Следует отметить, что изучение видеомодальности для автоматического определения когнитивных нарушений представлено меньшим количеством работ в сравнении с аудио и текстовыми подходами. В рассмотренных нами работах точность в среднем уступает результатам, полученным на аудио и текстовом материале. Однако исследование видеомодальности представляется перспективным направлением в контексте разработки языково-независимых методов. Особый интерес представляют предварительные исследования по извлечению промежуточных нейросетевых признаков, таких как эмоциональное состояние и автоматически распознанные физиологические признаки на основе удаленной фотоплетизмографии и иные.

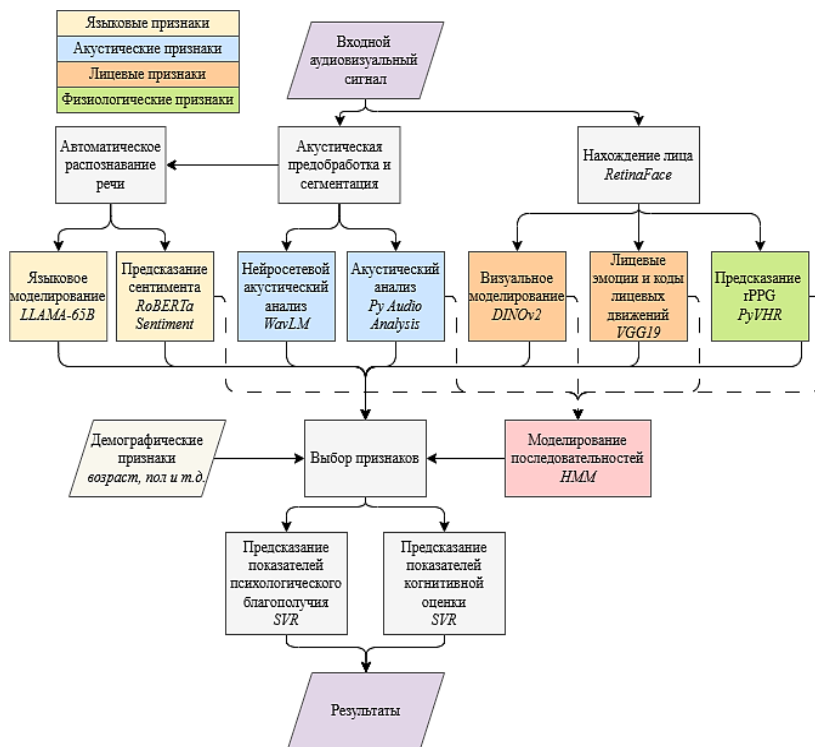


Рис. 3. Архитектура многомодальной системы предсказания показателей когнитивной оценки и психологического благополучия [50] (на рисунке курсивом указаны признаки, инструментарии и модели, использовавшиеся в реализации отдельных модулей системы)

6. Применение языковых макродескрипторов для определения когнитивных нарушений. Как упоминалось в разделе 3, некоторые из симптомов заболеваний, хотя они и естественно отмечаются специалистами и в некоторых случаях даже не специалистами, сравнительно трудно отразить с помощью интерпретируемых количественных лингвистических или акустических признаков. Для того, чтобы решить данную проблему, некоторые авторы последних работ предлагают использовать большие языковые модели для извлечения промежуточных высокоуровневых признаков (макродескрипторов), формализация которых затруднительна.

Впервые данный метод был использован в работе [52], где предлагалось использовать LLM для оценки беглости речи людей.

Предложенная в работе архитектура системы представлена на рисунке 4. Полученные метки беглости речи использовались для обучения нейросетевой модели, вместе с векторными нейросетевыми представлениями аудио и текста, что лишило результирующую систему возможности интерпретации полученных результатов. Наилучшее значение по показателю F-меры составило 87,25% при кросс-валидации на материале набора ADReSSo.

Авторы работы [53] предложили развитие идеи, сравнив эффективность использования больших языковых моделей в качестве прямой системы, определяющей наличие болезни Альцгеймера у пациента, с использованием данных моделей для выделения промежуточных высокоуровневых текстовых признаков и последующего обучения интерпретируемых стохастических моделей классификации. Исследование проводилось на наборе данных ADReSS, содержащем текстовые расшифровки, но рассматривался также вариант, при котором текстовые расшифровки отсутствовали, в результате чего применялись различные системы автоматического распознавания речи Whisper-large, Wav2Vec-large или иные. Наименьшую ошибку распознавания слов в речи WER = 26,9% продемонстрировал Whisper, хотя он склонен опускать заполненные паузы-хезитации, важные для определения заболеваний. Рассматривались такие LLM, как Mistral 7B¹⁴, Mixtral 8x7B¹⁵ и GPT-3.5¹⁶. Наилучшие и наиболее стабильные результаты были достигнуты при использовании LLM для извлечения промежуточных признаков и предсказания болезни с предоставлением модели информации о работе в медицинской области. Значение F1-меры = 79,2% при обучении классификатора на основе случайного леса, и 81,3% при обучении классификатора на основе опорных векторов с использованием макроописателей и предсказания заболевания модели Mistral 7B. Интересно отметить, что языковые модели большего размера демонстрировали в среднем более низкие и менее стабильные результаты.

¹⁴ <https://mistral.ai/news/announcing-mistral-7b>

¹⁵ <https://huggingface.co/blog/mixtral>

¹⁶ <https://openai.com/index/introducing-chatgpt-and-whisper-apis/>

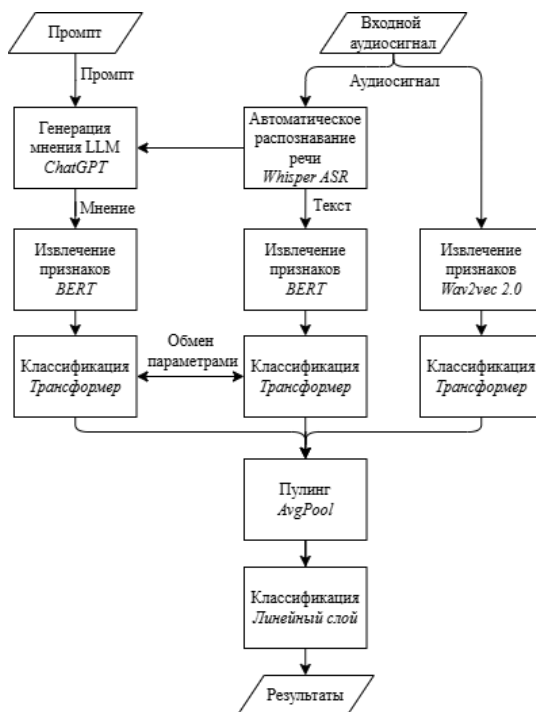


Рис. 4. Архитектура многомодальной системы предсказания показателей когнитивной оценки с применением LLM [52] (на рисунке курсивом указаны признаки, инструментарии и модели, использовавшиеся в реализации отдельных модулей системы)

В работе [14] авторы предложили многомодальный и интерпретируемый подход для определения нескольких заболеваний: слабых когнитивных нарушений, деменции при БА и БП. Эксперименты проводились на следующих наборах данных: Pitt по данным людей с когнитивными нарушениями на английском, PC GITA по данным людей с болезнью Паркинсона на испанском и немедицинском наборе CLAC [54] по данным оценки владения английским для расширения контрольной группы здоровых носителей. Текстовая транскрипция из речи извлекалась автоматически с помощью Whisper-large. При этом использовались различные интерпретируемые акустические, связанные с качеством голоса и формой речевого тракта, и темпоральные признаки, а также лингвистические количественные, например пропорции различных частей речи в тексте, плотность содержания и другие из библиотеки

BlaBla¹⁷ и, основанные на нейросетевых моделях, промежуточные признаки, например связность текста на основе косинусной близости векторов нейросетевой модели All-Mpnet-Base-v2¹⁸, сентимент-анализ моделью TextBlob¹⁹ и анализ цепочек кореферентности, последовательностей слов в тексте, которые относятся к одному и тому же обозначаемому, на основе нейросетевых языковых моделей из инструментария wl-coref²⁰ [55].

Для снижения размерности вектора признаков в [14] использовали корреляционный анализ, то есть удалялись сильно коррелированные признаки, если они оказывались незначимы для речи в выборке людей с когнитивными нарушениями и в выборке без когнитивных нарушений. Признаки группировались в кластеры с помощью иерархической кластеризации, основанной на корреляции Пирсона между всеми парами признаков. Для повышения устойчивости к шумам в разных наборах данных, эталонные признаки выбирались на основе сходства стандартного отклонения их распределений в выборках по здоровым людям и пациентам. Признак с наиболее схожим стандартным отклонением в обеих группах в каждом кластере был определен как эталонный. Средние значения не учитывались, так как их можно скорректировать путем добавления константы.

В качестве классификатора используется нейронная аддитивная модель (Neural Additive Model, NAM) [56], идея которой заключается в том, чтобы применить линейную аппроксимацию функций, представляемых в виде нейронных сетей с одним входом, соответствующих одному признаку, а затем обучить нейронные сети совместно с обратным распространения ошибки. В результате для каждого предсказания можно восстановить вклад входных признаков и связать наиболее значимые акустические или лингвистические признаки с соответствующими симптомами нарушений, понятными специалисту. Точность предложенной модели составила 84,3% на тестовой выборке набора по БА и 75% на тестовой выборке набора по БП. В обоих случаях признаки для обучения нормализовались с помощью нормализации с помощью размаха (MinMax). При этом предложенное решение сохраняет интерпретируемость и демонстрирует сопоставимые результаты с наиболее современными

¹⁷ <https://github.com/novoic/blabla>

¹⁸ <https://huggingface.co/sentence-transformers/all-mpnet-base-v2>

¹⁹ <https://textblob.readthedocs.io/en/dev/>

²⁰ <https://github.com/vdobrovolskii/wl-coref>

решениями, основанными на нейросетевых признаках для данной задачи.

Необходимо также отметить, что существуют несколько эталонных наборов данных для оценки применимости LLM относительно широкого спектра медицинских задач, в частности автоматических ответов на вопросы и проверки полноты данного ответа, в том числе и на вопросы, связанные с когнитивными нарушениями. Несколько наиболее популярных наборов для оценки возможностей генеративных моделей по решению медицинских задач:

- MedQA [57], MedMCQA [58] – англоязычные наборы данных только по ответам на вопросы (question-answering, QA);
- HealthBench [59] – англоязычный набор данных с различными типами задач: QA, генерация, классификация;
- RuMedBench [60] – русскоязычный набор данных с различными типами задач: QA, генерация, классификация.

В результате появления доступных материалов для обучения начали возникать и специализированные медицинские LLM. Далее рассмотрим несколько примеров подобных моделей.

Одна из первых работ, предлагающая создание многозадачной и многомодальной медицинской LLM, – Med-PaLM M [61] на основе собственного набора MultiMedBench. Результаты предложенной системы значительно превосходили результаты неспециализированной PaLM-E [62] во всех задачах (ответы на вопросы, автоматическое реферирование, классификация изображений и др.), хотя показатели специализированных систем эта модель не превзошла в задачах на основе текстовой модальности, либо превзошла только незначительно в задачах на основе визуальной модальности. Также авторы отмечают, что модель плохо справляется с адаптацией на основе небольшого количества примеров (few-shot learning, FSL).

В работе [63] предлагается создание серии специализированных медицинских нейросетевых языковых моделей с открытым доступом и количеством обучаемых параметров менее 10 миллиардов: Meerkat 7B и 8B. Результаты продемонстрировали, что относительно небольшое количество параметров и тщательный подбор данных для обучения цепочкам рассуждений на основе медицинских учебников позволяют значительно превзойти результаты многих специализированных и неспециализированных больших языковых моделей на медицинских эталонных данных. Однако, представленные модели все еще значительно отстают по количественным показателям от коммерческих LLM, таких как GPT-4o²¹ и o1²².

²¹ <https://openai.com/index/hello-gpt-4o/>

В перспективе, применение специализированных медицинских LLM могло бы показать лучшие результаты по сравнению с неспециализированными моделями, ранее исследованными в задаче по определению когнитивных нарушений.

Применение идеи макро дескрипторов на основе глубоких нейронных сетей представляется одной из наиболее интересных в контексте разработки технологий интерпретируемого искусственного интеллекта для здравоохранения. При этом предложенные решения на данный момент не рассматривали макро дескрипторы для анализа многоязычных данных и трудноизмеримых качественных характеристик аудио и видео, например, для анализа эмоционального состояния, что оставляет пространство для будущих исследований.

7. Систематизация информативных признаков и классификаторов, используемых при разработке многомодальных систем автоматического определения когнитивных нарушений. В ходе проведенного обзора был выявлено несколько основных тенденций в выборе признаков и моделей для классификации.

Для учета дисбаланса в работах часто используются показатели точности моделей, устойчивые к дисбалансу. Возможно также применение методик аугментации данных и кросс-корпусных методов. Также важен выбор наиболее информативных признаков. В рассмотренных работах упоминались следующие: метод главных компонент, лассо-регрессия и др.

Для извлечения признаков использовались методы, которые можно классифицировать по модальностям, на основе которых они извлекаются: аудио, видео и текст, среди которых можно выделить разделение на нейросетевые методы и программные инструментарии, позволяющие извлекать экспертные или нейросетевые признаки, они представлены на рисунке 5.

Несмотря на то, что изменения в поведении людей с когнитивными нарушениями могут проявляться в аудио в форме отклонений тона голоса, артикуляции, беглости речи и др. и в видео в форме отклонений в проявлении эмоции и снижения мимической активности, наиболее чувствительной к нарушениям во многих работах оказывались признаки текстовой модальности. С обработкой теста во многих наборах тесно связана задача автоматического распознавания речи и речи с нарушениями, поскольку в наборах ручная транскрипция может отсутствовать. Лучшие модели распознавания речи показывают значительное снижение точности работы на данных с когнитивными нарушениями. В исследовании [64],

²² <https://openai.com/index/introducing-openai-o1-preview/>

рассматривались значения WER для предобученных моделей на основе нейросетевых архитектур Wav2Vec и Whisper, полученные на наборе ADReSS. Whisper-large-v3 продемонстрировал WER=40,07% на выборке речи людей с когнитивными нарушениями и 28,47% на контрольной группе, тогда как wav2vec2-large-960h-lv60-self²³ показал 50,48% и 35,97% на речи людей с когнитивными нарушениями и контрольной группе соответственно. В исследовании [66] отмечается заметное влияние точности распознавания речи на автоматическое определение когнитивных нарушений, поэтому необходимо рассматривать также задачу дообучения моделей на медицинских данных.

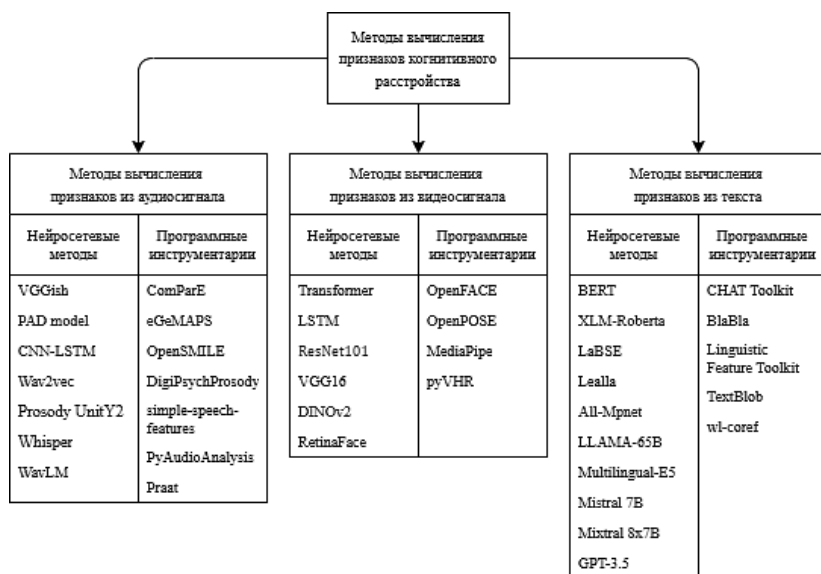


Рис. 5. Систематизация методов для извлечения информативных признаков из многомодальных данных

Среди рассмотренных работ можно выделить регрессионные и классификационные системы, в которых используются как детерминированные машинные классификаторы, так и нейросетевые классификаторы. На рисунке 6 представлены основные, использовавшиеся в работах классификаторы. Важно отметить, что традиционные детерминированные классификаторы не теряют

²³ <https://huggingface.co/facebook/wav2vec2-large-960h-lv60-self>

популярности, несмотря на активное развитие нейросетевых методов. Данная тенденция связана с малыми размерами данных для обучения и ориентацией на использование интерпретируемых моделей в медицинских задачах. Также можно заметить, что в работах совсем недавно стали изучаться LLM для извлечения интерпретируемых признаков.

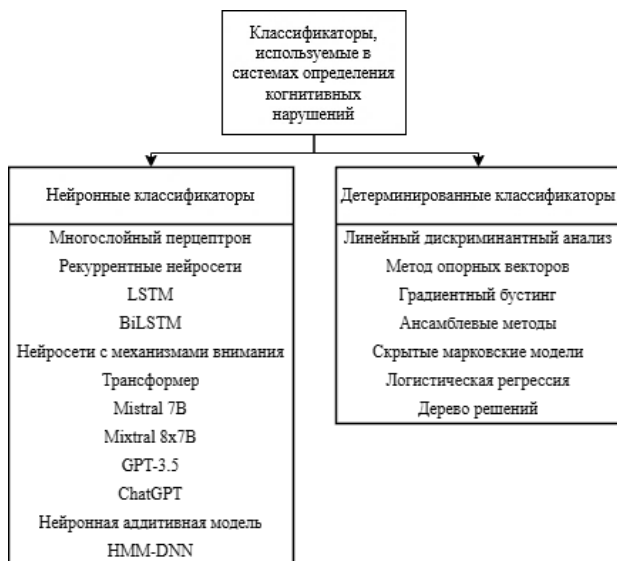


Рис. 6. Систематизация методов машинной классификации, используемых при определении когнитивных нарушений

Важно отметить, что исследование применения современных интерпретируемых подходов, таких как: нейронные аддитивные модели; модели анализа выживаемости, подходящие для предсказания когнитивного снижения через время [66]; интерпретируемые графовые нейронные модели [67] – находится на начальном этапе.

Находят свое применение инструменты интерпретации предсказаний трудноинтерпретируемых моделей. Например, в работе [68], метод SHAP применяется для анализа влияния признаков на результат предсказания для одного человека, полученный с помощью модели XGBoost²⁴. Данный метод позволяет получить индивидуальную оценку риска заболевания, что важно при диагностике конкретного пациента.

²⁴ <https://xgboost.readthedocs.io/en/latest/>

8. Заключение. В статье представлен аналитический обзор современных научных исследований, посвященных разработке автоматических систем определения когнитивных нарушений людей.

В результате выполненного аналитического обзора можно сделать несколько основных выводов:

1) Объединение текстовой модальности с невербальными акустическими, например, учетом пауз, темпа, или лицевыми признаками, например, эмоциональными признаками, позволяет улучшить результаты, но в силу отсутствия текста во многих наборах требуется учитывать качество моделей распознавания речи.

2) Анализ лексической составляющей позволяет обнаружить различные симптомы, например, снижение лексического разнообразия, трудности в нахождении слов и др.

3) Конкурентоспособные решения представляются, в том числе, с применением экспертных признаков и линейных статистических моделей, что потенциально может обеспечить интерпретируемость полученных результатов.

На основе проведенного анализа можно сформулировать потенциальные требования, которые могут быть выдвинуты к разрабатываемым автоматическим системам определения когнитивных нарушений:

1) Необходимо использовать средства нормализации и борьбы со статистическими выбросами и шумами. Большинство речевых и аудиовизуальных наборов по когнитивным нарушениям отличаются малыми объемами, несбалансированностью, неоднобразными условиями записи и высокой внутридикторской вариативностью, из-за чего данные могут быть зашумлены и модели могут переобучаться на отдельных наборах данных.

2) Необходимо отдать предпочтение интерпретируемым моделям классификации, экспертным наборам признаков, а также инструментам интерпретации (например, SHAP) нейросетевых моделей: глубоких нейронных сетей, больших языковых моделей и др. Применение моделей типа «черный ящик» может потенциально привести к ложным пропускам и срабатываниям, т.е. отсутствию своевременного медицинского вмешательства или наоборот к вмешательству в здоровом случае. Интерпретация используемых моделей позволяет специалисту отсеять незначимые случайные закономерности, которые могли возникнуть из-за ограничений обучающих данных.

3) Рекомендуется использовать и интегрировать максимально возможное количество модальностей. Отдельная модальность (текст,

аудио, видео и иные) зачастую не отражают всех особенностей поведения человека, что может привести автоматические средства к ошибкам первого или второго рода (ложным срабатываниям или пропуску нарушений).

В дальнейших исследованиях мы планируем разработать собственный прототип автоматической системы определения когнитивных нарушений по речи, учитывающий разработанные требования, и исследовать на практике акустические, лексические и визуальные лицевые признаки когнитивных нарушений, выявленные в ходе данного обзора и методы их аугментации.

Мы хотели бы посвятить эту работу памяти Рафаэля Мидхатовича Юсупова, который всегда активно поддерживал направления исследований и разработок лаборатории речевых и многомодальных интерфейсов СПИИРАН (СПб ФИЦ РАН). Рафаэль Мидхатович также являлся одним из основоположников становления и развития в России новой научной области – интеллектуальных многомодальных человеко-машинных интерфейсов и систем [69]. Светлая ему память.

Литература

1. Парфенов В.А. Ведение пациентов с когнитивными нарушениями // Неврология, нейропсихиатрия, психосоматика. 2023. Т. 15(1). С. 97–102.
2. Jonell P. et al. Multimodal Capture of Patient Behaviour for Improved Detection of Early Dementia: Clinical Feasibility and Preliminary Results // *Frontiers in Computer Sciences*. 2021. vol. 3. DOI: 10.3389/fcomp.2021.642633.
3. Javeed A., Dallora A.L., Berglund J.S., et al. Machine Learning for Dementia Prediction: A Systematic Review and Future Research Directions. *Journal of Medical Systems*. 2023. vol. 47(17). DOI: 10.1007/s10916-023-01906-7.
4. Ляко Е.Е., Фролова О.В., Гречаный С.В., Матвеев Ю.Н., Верхоляк О.В., Карпов А.А. Голосовой портрет ребенка с типичным и атипичным развитием // СПб.: Издательско-полиграфическая ассоциация высших учебных заведений. 2020. 204 с.
5. Величко А.Н., Карпов А.А. Аналитический обзор систем автоматического определения депрессии по речи // *Информатика и автоматизация*. 2021. № 20(3). С. 497–529. DOI: 10.15622/ia.2021.3.1.
6. Боголепова А.Н. и др. Клинические рекомендации «Когнитивные расстройства у пациентов пожилого и старческого возраста» // *Журнал неврологии и психиатрии им. С.С. Корсакова*. 2022. Т. 121(10-3). С. 6–137.
7. Ganguli M., Chang C.C., Snitz B.E., Saxton J.A., Vanderbilt J., Lee C.W. Prevalence of mild cognitive impairment by multiple classifications: The Monongahela-Youghiogheny Healthy Aging Team (MYHAT) project // *American Journal of Geriatric Psychiatry*. 2010. vol. 18(8). pp. 674–683. DOI: 10.1097/JGP.0b013e3181cdee4f.
8. Larrieu S., Letenneur L., Orgogozo J.M., Fabrigoule C., Amieva H., Le Carret N., Barberger-Gateau P., Dartigues J.F. Incidence and outcome of mild cognitive impairment in a population-based prospective cohort // *Neurology*. 2002. vol. 59(10). pp. 1594–1599. DOI: 10.1212/01.wnl.0000034176.07159.f8.

9. Roberts R.O., Geda Y.E., Knopman D.S., Cha R.H., Pankratz V.S., Boeve B.F., Ivnik R.J., Tangalos E.G., Petersen R.C., Rocca W.A. The Mayo Clinic Study of Aging: design and sampling, participation, baseline measures and sample characteristics // *Neuroepidemiology*. 2008. vol. 30(1). pp. 58–69. DOI: 10.1159/000115751.
10. Чердак М.А. и др. Распространенность когнитивных расстройств у пациентов старшего возраста в Российской Федерации // *Журнал неврологии и психиатрии им. С.С. Корсакова*. Спецвыпуски. 2024. Т. 124(4-2). С. 5–11.
11. Wallensten J., Ljunggren G., Nager A., Wachtler C., Bogdanovic N., Petrovic P., Carlsson A.C. Stress, depression, and risk of dementia – a cohort study in the total population between 18 and 65 years old in Region Stockholm // *Alzheimer's Research & Therapy*. 2023. vol. 15(161). DOI: 10.1186/s13195-023-01308-4.
12. Локшина А.Б., Гришина Д.А. Терапия некогнитивных нервно-психических расстройств при болезни Альцгеймера // *Неврология, нейропсихиатрия, психосоматика*. 2021. Т. 13(6). С. 132–138. DOI: 10.14412/2074-2711-2021-6-132-138.
13. Nestor P.J., Scheltens P., Hodges J.R. Advances in the early detection of Alzheimer's disease // *Nature Medicine*. 2004. vol. 10. pp. S34–S41. DOI: 10.1038/nm1433.
14. Botelho C., Abad A., Schultz T., Trancoso I. Speech as a biomarker for disease detection // *IEEE Access*. 2024. vol. 12. pp. 184487–184508.
15. Yamada Y., Shinkawa K., Ishikawa T., Nishimura M., Nemoto M., Tsukada E., Ota M., Nemoto K., Arai T. Multimodal behavioral analysis for early detection of Alzheimer's disease: A preliminary result: Neuropsychiatry and behavioral neurology/Assessment/Measurement of neuropsychiatric/Behavioral and psychological symptoms // *Alzheimer's & Dementia*. 2020. vol. 16. DOI: 10.1002/alz.042897.
16. Boschi V., Catricala E., Consonni M., Chesi C., Moro A., Cappa S.F. Connected speech in neurodegenerative language disorders: a review // *Frontiers in psychology*. 2017. vol. 8. DOI: 10.3389/fpsyg.2017.00269.
17. Hecker P., Steckhan N., Eyben F., Schuller B.W., Arnrich B. Voice analysis for neurological disorder recognition—a systematic review and perspective on emerging trends // *Frontiers in Digital Health*. 2022. vol. 4. DOI: 10.3389/fdgth.2022.842301.
18. Voleti R., Liss J.M., Berisha V. A review of automated speech and language features for assessment of cognitive and thought disorders // *IEEE journal of selected topics in signal processing*. 2019. vol. 14(2). pp. 282–298. DOI: 10.1109/jstsp.2019.2952087.
19. Byers A.L., Yaffe K. Depression and risk of developing dementia // *Nature Reviews Neurology*. 2011. vol. 7(6). pp. 323–331. DOI: 10.1038/nrneurol.2011.60.
20. Braun F., Bayerl S.P., Perez-Toro P.A., Honig F., Leheld H., Hillemacher T., Noth E., Bocklet T., Riedhammer K. Classifying Dementia in the Presence of Depression: A Cross-Corpus Study // *Proceedings of Interspeech* 2023. vol. 2023. pp. 2308–2312. DOI: 10.21437/Interspeech.2023-1997.
21. Jonell P., et al. Multimodal Capture of Patient Behaviour for Improved Detection of Early Dementia: Clinical Feasibility and Preliminary Results // *Frontiers in Computer Science*. 2021. vol. 3. DOI: 10.3389/fcomp.2021.642633.
22. Lanzi A.M., Saylor A.K., Fromm D., Liu H., MacWhinney B., Cohen M.L. DementiaBank: Theoretical Rationale, Protocol, and Illustrative Analyses // *American Journal of Speech-Language Pathology*. 2023. vol. 32(2). pp. 426–438. DOI: 10.1044/2022_AJSLP-22-00281.
23. Becker J.T., Boller F., Lopez O.L., Saxton J., McGonigle K.L. The natural history of Alzheimer's disease: description of study cohort and accuracy of diagnosis // *Archives of Neurology*. 1994. vol. 51(6). pp. 585–594. DOI: 10.1001/archneur.1994.00540180063015.
24. Luz S., Haider F., Fuente S.d.l., Fromm D., MacWhinney B. Alzheimer's Dementia Recognition through Spontaneous Speech: The ADReSS Challenge // *Proceedings Interspeech*. 2020. pp. 2172–2176. DOI: 10.21437/Interspeech.2020-2571.

25. MacWhinney B. The CHILDES Project: Tools for Analyzing Talk // *Child Language Teaching and Therapy*. 1992. vol. 8(2). pp. 217–218. DOI: 10.1177/026565909200800211.
26. Luz S., et al. Alzheimer's dementia recognition through spontaneous speech // *Frontiers in Computer Science*. 2021. vol. 3.
27. Luz S., et al. Connected Speech-Based Cognitive Assessment in Chinese and English // *Proceedings Interspeech*. 2024. pp. 947–951. DOI: 10.21437/Interspeech.2024-1807.
28. Chen X., Zhang W., Ma Y. Raw Waveform-Based End-to-End Alzheimer's Disease Detection Method // *Acta Electronica Sinica*. 2023. vol. 51. no. 12. pp. 3582–3590.
29. Псарёва Н.Н. Беглость спонтанной речи как предиктор мягкого когнитивного снижения // ВКР по программе бакалавриата. М.: ВШЭ, 2024.
30. Karakostas A., Briassouli A., Avgerinakis K., Kompatsiaris I., Tsolaki M. The dem@care experiments and datasets: a technical report // *arXiv preprint arXiv:1701.01142*. 2016.
31. Kishimoto T. et al. PROMPT collaborators. The project for objective measures using computational psychiatry technology (PROMPT): Rationale, design, and methodology // *Contemporary Clinical Trials Communications*. 2020. vol. 19. DOI: 10.1016/j.conctc.2020.100649.
32. Poor F.F. et al. Prediction of Mild Cognitive Impairment Using a Hybrid Audio-Visual Approach: An I_CONECT Study // *Alzheimer's & Dementia Journal*. 2023. vol. 19. DOI: 10.1002/alz.074808.
33. Correia J., Teixeira F., Botelho C., Trancoso I., Raj B. The in-the-Wild Speech Medical Corpus // 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). 2021. pp. 6973–6977. DOI: 10.1109/ICASSP39728.2021.9414230.
34. Orozco J.R. et al. New Spanish speech corpus database for the analysis of people suffering from Parkinson's disease // *International Conference on Language Resources and Evaluation (LREC)*. 2014. pp. 342–347.
35. Gratch J. et al. The Distress Analysis Interview Corpus of human and computer interviews // *International Conference on Language Resources and Evaluation (LREC)*. 2014. pp. 3123–3128.
36. DeVault D. et al. SimSensei kiosk: A virtual human interviewer for healthcare decision support // *Proceedings of the 13th International Conference on Autonomous Agents and Multiagent Systems (AAMAS'14)*. 2014. pp. 1061–1068.
37. Vashkevich M., Rushkevich Yu., Petrovsky A. Bulbar ALS Detection Based on Analysis of Voice Perturbation and Vibrato // *Proceedings of inter. conf. Signal Processing: Algorithms, Architectures, Arrangements, and Applications (SPA)*. 2019. pp. 267–272. DOI: 10.23919/SPA.2019.8936657.
38. Qi X., Zhou Q., Dong J., Bao W. Noninvasive automatic detection of Alzheimer's disease from spontaneous speech: a review // *Frontiers in Aging Neuroscience*. 2023. vol. 15.
39. Ding K., Chetty M., Noori Hoshyar A., Bhattacharya T., Klein B. Speech based detection of Alzheimer's disease: a survey of AI techniques, datasets and challenges // *Artificial Intelligence Review*. 2024. vol. 57. no. 12.
40. Babu A. et al. XLS-R: Self-supervised cross-lingual speech representation learning at scale // *Proceedings of Interspeech*. 2022. pp. 2278–2282. DOI: 10.21437/Interspeech.2022-143.
41. Duan J., Wei F., Li H.D., Liu J. Pre-trained Feature Fusion and Matching for Mild Cognitive Impairment Detection // *Proceedings of Interspeech*. 2024. pp. 962–966.
42. Favaro A., Cao T., Dehak N., Moro-Velazquez L. Leveraging Universal Speech Representations for Detecting and Assessing the Severity of Mild Cognitive Impairment Across Languages // *Proceedings of Interspeech*. 2024. pp. 972–976. DOI: 10.21437/Interspeech.2024-2030.

43. Radford A., Kim J.W., Xu T., Brockman G., McLeavey C., Sutskever I. Robust speech recognition via large-scale weak supervision // International conference on machine learning. PMLR 2023. 2023. pp. 28492–28518.
44. Lee B.W., Lee J. LFTK: Handcrafted features in computational linguistics // Proceedings of the 18th Workshop on Innovative Use of NLP for Building Educational Applications (BEA 2023). 2023. pp. 1–19. DOI: 10.18653/v1/2023.bea-1.1.
45. Hoang B., Pang Y., Dodge H., Zhou J. Translingual Language Markers for Cognitive Assessment from Spontaneous Speech // Proceedings of Interspeech. 2024. 977–981. DOI: 10.21437/Interspeech.2024-1422.
46. Fan A. et al. Beyond English-centric multilingual machine translation // Journal of Machine Learning Research. 2021. vol. 22. no. 107. pp. 1–48.
47. Pérez-Toro P.A., Arias-Vergara T., Klumpp P., Weise T., Schuster M., Noeth E., Orozco-Arroyave J.R., Maier A.. Multilingual Speech and Language Analysis for the Assessment of Mild Cognitive Impairment: Outcomes from the Taukadiad Challenge // Proceedings of Interspeech. 2024. pp. 982–986. DOI: 10.21437/Interspeech.2024-2115.
48. Zheng C. et al. Detecting Dementia from Face-Related Features with Automated Computational Methods // Bioengineering (Basel). 2023. vol. 10(7). DOI: 10.3390/bioengineering10070862.
49. Okunishi T., Zheng C., Bouazizi M., Ohtsuki T., Kitazawa M., Horigome T., Kishimoto T. Dementia and MCI Detection Based on Comprehensive Facial Expression Analysis from Videos During Conversation // IEEE Journal of Biomedical and Health Informatics. 2025. vol. 29(5). pp. 3537–3548. DOI: 10.1109/JBHI.2025.3526553.
50. Mu X. et al. Detecting Cognitive Impairment and Psychological Well-being among Older Adults Using Facial, Acoustic, Linguistic, and Cardiovascular Patterns Derived from Remote Conversations // arXiv preprint arXiv:2412.14194. 2024.
51. Jiang Z., Seyed S., Griner E., Abbasi A., Rad A.B., Kwon H., Cotes R.O., Clifford G.D. Multimodal Mental Health Digital Biomarker Analysis from Remote Interviews Using Facial, Vocal, Linguistic, and Cardiovascular Patterns // IEEE Journal of Biomedical and Health Informatics. 2024. vol. 28(3). pp. 1680–1691. DOI: 10.1109/JBHI.2024.3352075.
52. Bang J.-U., Han S.-H., Kang B.-O. Alzheimer’s Disease recognition from spontaneous speech using large language models // ETRI Journal. 2024. vol. 46. no. 1. pp. 96–105. DOI: 10.4218/etrij.2023-0356.
53. Botelho C. et al. Macro-descriptors for Alzheimer’s disease detection using large language models. Proc. Interspeech 2024. 2024. pp. 1975–1979.
54. Haulcy R., Glass J. CLAC: A Speech Corpus of Healthy English Speakers // Proceedings of Interspeech. 2021. pp. 2966–2970. DOI: 10.21437/Interspeech.2021-1810.
55. Dobrovolskii V. Word-Level Coreference Resolution // Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing. 2021. pp. 7670–7675. DOI: 10.18653/v1/2021.emnlp-main.605.
56. Agarwal R., Melnick L., Frosst N., Zhang X., Lengerich B., Caruana R., Hinton G.E. Neural additive models: Interpretable machine learning with neural nets // Advances in Neural Information Processing Systems. 2021. vol. 34. pp. 4699–4711.
57. Jin D. et al. What Disease Does This Patient Have? A Large-Scale Open Domain Question Answering Dataset from Medical Exams // Applied Sciences. 2021. vol. 11(14). DOI: 10.3390/app11146421.
58. Pal A. et al. MedMCQA: A Large-scale Multi-Subject Multi-Choice Dataset for Medical domain Question Answering // Proceedings of Machine Learning Research. 2022. vol. 174. pp. 248–260.
59. Arora R.K. et al. HealthBench: Evaluating Large Language Models Towards Improved Human Health // arXiv preprint arXiv:2505.08775. 2025.

60. Blinov P., Reshetnikova A., Nesterov A., Zubkova G., Kokh V. RuMedBench: A Russian Medical Language Understanding Benchmark // Conference on Artificial Intelligence in Medicine. 2022. pp. 383–392.
61. Tu T. et al. Towards generalist biomedical AI // NEJM Ai. 2024. vol. 1(3).
62. Driess D. et al. PaLM-E: an embodied multimodal language model. Proceedings of the 40th International Conference on Machine Learning (ICML'23). 2023. vol. 202. pp. 8469–8488.
63. Kim H. et al. Small language models learn enhanced reasoning skills from medical textbooks // NPJ Digital Medicine. 2025. vol. 8(1). DOI: 10.1038/s41746-025-01653-8.
64. Wang Y. et al. Exploiting prompt learning with pre-trained language models for Alzheimer's Disease detection // IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). 2023. pp. 1–5. DOI: 10.1109/ICASSP49357.2023.10095993.
65. Balagopalan A., Shkaruta K., Novikova J. Impact of ASR on Alzheimer's Disease Detection: All Errors are Equal, but Deletions are More Equal than Others // Proceedings of the Sixth Workshop on Noisy User-generated Text (W-NUT 2020). 2020. pp. 159–164. DOI: 10.18653/v1/2020.wnut-1.21.
66. Utkin L.V., Konstantinov A.V., Eremenko D.Yu., Zaborovsky V.S., Muliukha V.A. Interpretation methods for machine learning models in the framework of survival analysis with censored data: a brief over view // Computing, Telecommunications and Control. 2024. vol. 17. no. 3. pp. 22–31. DOI: 10.18721/JCSTCS.17302.
67. Alkhatib A. et al. Interpretable Graph Neural Networks for Tabular Data // Proceedings of the 27th European Conference on Artificial Intelligence (ECAI 2024). 2024. vol. 392. pp. 1848–1855.
68. Ren H., Zheng Y., Li C., Jing F., Wang Q., Luo Z., Li D., Liang D., Tang W., Liu L., Cheng W. Using Machine Learning to Predict Cognitive Decline in Older Adults From the Chinese Longitudinal Healthy Longevity Survey: Model Development and Validation Study // JMIR Aging. 2025. vol. 8. DOI: 10.2196/67437.
69. Karpov A.A., Yusupov R.M. Multimodal Interfaces of Human-Computer Interaction. Herald of the Russian Academy of Sciences. 2018. vol. 88(1). pp. 67–74. DOI: 10.1134/S1019331618010094.

Долгушин Михаил Дмитриевич — младший научный сотрудник, аспирант, лаборатория речевых и многомодальных интерфейсов, Федеральное государственное бюджетное учреждение науки «Санкт-Петербургский Федеральный исследовательский центр Российской академии наук» (СПб ФИЦ РАН). Область научных интересов: речевые технологии, автоматическое распознавание речи, интерпретируемый искусственный интеллект, компьютерная лингвистика. Число научных публикаций — 12. dolgushin.m@ias.spb.su; 14-я линия В.О., 39, 199178, Санкт-Петербург, Россия; р.т.: +7(812)328-0421.

Карпов Алексей Анатольевич — д-р техн. наук, профессор, руководитель лаборатории, лаборатория речевых и многомодальных интерфейсов, Федеральное государственное бюджетное учреждение науки «Санкт-Петербургский Федеральный исследовательский центр Российской академии наук» (СПб ФИЦ РАН). Область научных интересов: речевые технологии, автоматическое распознавание речи, обработка аудиовизуальной речи, многомодальные человеко-машинные интерфейсы, компьютерная паралингвистика. Число научных публикаций — 450+. karpov@ias.spb.su; 14-я линия В.О., 39, 199178, Санкт-Петербург, Россия; р.т.: +7(812)328-0421.

Поддержка исследований. Данное исследование выполнено при финансовой поддержке Российского научного фонда, проект № 25-11-00319, кроме раздела № 6, выполненного в рамках бюджетной темы № FFZF-2025-0003.

M. DOLGUSHIN, A. KARPOV
**ANALYTICAL REVIEW OF SPEECH AND MULTIMODAL
METHODS FOR COGNITIVE IMPAIRMENTS RECOGNITION
IN PEOPLE**

Dolgushin M., Karpov A. Analytical Review of Speech and Multimodal Methods for Cognitive Impairments Recognition in People.

Abstract. Over the past decade, there has been a noticeable increase in the number of scientific, technical, and medical publications dedicated to the automatic detection of cognitive impairments in humans based on speech and visual data. These impairments are often associated with neurodegenerative diseases such as dementia, Alzheimer's disease, Parkinson's disease, and other disorders. Despite the high prevalence of these conditions and their significant contribution to mortality and early disability, effective treatment options remain unavailable or severely limited in current medical practice. Consequently, early diagnosis and symptom alleviation have become areas of considerable research interest. Current studies focus on the development of automated and automatic systems based on quantitative and objective methods, neural network approaches, the integration of various modalities, and explainable artificial intelligence techniques. This paper presents a comprehensive review and analysis of key studies published since 2022 that address the automatic detection of cognitive impairments using unimodal and multimodal approaches. The review includes the most commonly used multimodal datasets in this domain, such as ADReSS, ADReSSo, and TAUkADIAL. It discusses state-of-the-art methods for detecting cognitive impairments from various modalities, including those presented in international competitions such as TAUkADIAL-2024, as well as methods developed outside of such events. According to competition results, the most effective approaches to recognizing cognitive impairments are ensemble probabilistic models trained on explainable hand-crafted features and neural features extracted from text and audio data. The review also explores multimodal approaches that incorporate visual modalities for training deep neural networks. A new direction in the field is examined, namely, the applicability of large language models to the analysis of medical texts and interpretable disease prediction. The paper systematizes methods for extracting informative features and the classifiers employed. Based on the review, key requirements for systems aimed at the automated detection of cognitive impairments are formulated.

Keywords: automatic detection of cognitive impairment, speech technologies in healthcare, explainable artificial intelligence, machine learning.

References

1. Parfenov VA. [Management of patients with cognitive impairment]. *Nevrologiya, neiropsikhiatriya, psikhosomatika – Neurology, Neuropsychiatry, Psychosomatics*. 2023. vol. 15(1). pp. 97–102. (In Russ.).
2. Jonell P. et al. Multimodal Capture of Patient Behaviour for Improved Detection of Early Dementia: Clinical Feasibility and Preliminary Results. *Frontiers in Computer Sciences*. 2021. vol. 3. DOI: 10.3389/fcomp.2021.642633.
3. Javeed A., Dallora A.L., Berglund J.S., et al. Machine Learning for Dementia Prediction: A Systematic Review and Future Research Directions. *Journal of Medical Systems*. 2023. vol. 47(17). DOI: 10.1007/s10916-023-01906-7.
4. Lyakso E.E., Frolova O.V., Grechanyj S.V., Matveev Ju.N., Verholjak O.V., Karpov A.A. Golosovoj portret rebenka s tipichnym i atipichnym razvitiem [Vocal profile of a

- child with typical and atypical development]. SPb.: Izdatel'sko-poligraficheskaya associaciya vysshih uchebnyh zavedenij. 2020. 204 p. (In Russ.).
5. Velichko A., Karpov A. [Analytical Review of Automatic Systems for Depression Detection by Speech]. *Informatics and Automation*. 2021. vol. 20(3). pp. 497–529. DOI: 10.15622/ia.2021.3.1. (In Russ.).
6. Bogolepova A.N., et al. [Clinical Guidelines for Cognitive Disorders in Elderly and Older Patients]. *Zhurnal nevrologii i psikiatrii im. S.S. Korsakova – S.S. Korsakov Journal of Neurology and Psychiatry*. 2022. vol. 121(10-3). pp. 6–137. (In Russ.).
7. Ganguli M., Chang C.C., Snitz B.E., Saxton J.A., Vanderbilt J., Lee C.W. Prevalence of mild cognitive impairment by multiple classifications: The Monongahela-Youghiogheny Healthy Aging Team (MYHAT) project. *American Journal of Geriatric Psychiatry*. 2010. vol. 18(8). pp. 674–683. DOI: 10.1097/JGP.0b013e3181cdee4f.
8. Larrieu S., Letenneur L., Orgogozo J.M., Fabrigoule C., Amieva H., Le Carret N., Barberger-Gateau P., Dartigues J.F. Incidence and outcome of mild cognitive impairment in a population-based prospective cohort. *Neurology*. 2002. vol. 59(10). pp. 1594–1599. DOI: 10.1212/01.wnl.0000034176.07159.f8.
9. Roberts R.O., Geda Y.E., Knopman D.S., Cha R.H., Pankratz V.S., Boeve B.F., Ivnik R.J., Tangalos E.G., Petersen R.C., Rocca W.A. The Mayo Clinic Study of Aging: design and sampling, participation, baseline measures and sample characteristics. *Neuroepidemiology*. 2008. vol. 30(1). pp. 58–69. DOI: 10.1159/000115751.
10. Cherdak M.A. et al. [Prevalence of cognitive impairment in older adults in the Russian Federation]. *Zhurnal nevrologii i psikiatrii im. S.S. Korsakova. Specvypuski – S.S. Korsakov Journal of Neurology and Psychiatry. Special issues*. 2024. vol. 124(4-2). pp. 5–11. (In Russ.).
11. Wallensten J., Ljunggren G., Nager A., Wachtler C., Bogdanovic N., Petrovic P., Carlsson A.C. Stress, depression, and risk of dementia – a cohort study in the total population between 18 and 65 years old in Region Stockholm. *Alzheimer's Research & Therapy*. 2023. vol. 15(161). DOI: 10.1186/s13195-023-01308-4.
12. Lokshina A.B., Grishina D.A. [Treatment of noncognitive neuropsychiatric disorders in Alzheimer's disease]. *Nevrologiya, neiropsikhiatriya, psikhosomatika – Neurology, Neuropsychiatry, Psychosomatics*. 2021. vol. 13(6). pp. 132–138. DOI: 10.14412/2074-2711-2021-6-132-138. (In Russ.).
13. Nestor P.J., Scheltens P., Hodges J.R. Advances in the early detection of Alzheimer's disease. *Nature Medicine*. 2004. vol. 10. pp. S34–S41. DOI: 10.1038/nrn1433.
14. Botelho C., Abad A., Schultz T., Trancoso I. Speech as a biomarker for disease detection. *IEEE Access*. 2024. vol. 12. pp. 184487–184508.
15. Yamada Y., Shinkawa K., Ishikawa T., Nishimura M., Nemoto M., Tsukada E., Ota M., Nemoto K., Arai T. Multimodal behavioral analysis for early detection of Alzheimer's disease: A preliminary result: *Neuropsychiatry and behavioral neurology/Assessment/Measurement of neuropsychiatric/Behavioral and psychological symptoms. Alzheimer's & Dementia*. 2020. vol. 16. DOI: 10.1002/alz.042897.
16. Boschi V., Catricala E., Consonni M., Chesi C., Moro A., Cappa S.F. Connected speech in neurodegenerative language disorders: a review. *Frontiers in psychology*. 2017. vol. 8. DOI: 10.3389/fpsyg.2017.00269.
17. Hecker P., Steckhan N., Eyben F., Schuller B.W., Arnrich B. Voice analysis for neurological disorder recognition – a systematic review and perspective on emerging trends. *Frontiers in Digital Health*. 2022. vol. 4. DOI: 10.3389/fdgth.2022.842301.
18. Voleti R., Liss J.M., Berisha V. A review of automated speech and language features for assessment of cognitive and thought disorders. *IEEE journal of selected topics in signal processing*. 2019. vol. 14(2). pp. 282–298. DOI: 10.1109/jstsp.2019.2952087.

19. Byers A.L., Yaffe K. Depression and risk of developing dementia. *Nature Reviews Neurology*. 2011. vol. 7(6). pp. 323–331. DOI: 10.1038/nrneuro.2011.60.
20. Braun F., Bayerl S.P., Perez-Toro P.A., Honig F., Lehfeld H., Hillemacher T., Noth E., Bocklet T., Riedhammer K. Classifying Dementia in the Presence of Depression: A Cross-Corpus Study. *Proceedings of Interspeech 2023*. vol. 2023. pp. 2308–2312. DOI: 10.21437/Interspeech.2023-1997.
21. Jonell P., et al. Multimodal Capture of Patient Behaviour for Improved Detection of Early Dementia: Clinical Feasibility and Preliminary Results. *Frontiers in Computer Science*. 2021. vol. 3. DOI: 10.3389/fcomp.2021.642633.
22. Lanzi A.M., Saylor A.K., Fromm D., Liu H., MacWhinney B., Cohen M.L. DementiaBank: Theoretical Rationale, Protocol, and Illustrative Analyses. *American Journal of Speech-Language Pathology*. 2023. vol. 32(2). pp. 426–438. DOI: 10.1044/2022_AJSLP-22-00281.
23. Becker J.T., Boller F., Lopez O.L., Saxton J., McGonigle K.L. The natural history of Alzheimer's disease: description of study cohort and accuracy of diagnosis. *Archives of Neurology*. 1994. vol. 51(6). pp. 585–594. DOI: 10.1001/archneur.1994.00540180063015.
24. Luz S., Haider F., Fuente S.d.l., Fromm D., MacWhinney B. Alzheimer's Dementia Recognition through Spontaneous Speech: The ADReSS Challenge. *Proceedings Interspeech*. 2020. pp. 2172–2176. DOI: 10.21437/Interspeech.2020-2571.
25. MacWhinney B. The CHILDES Project: Tools for Analyzing Talk. *Child Language Teaching and Therapy*. 1992. vol. 8(2). pp. 217–218. DOI: 10.1177/026565909200800211.
26. Luz S., et al. Alzheimer's dementia recognition through spontaneous speech. *Frontiers in Computer Science*. 2021. vol. 3.
27. Luz S., et al. Connected Speech-Based Cognitive Assessment in Chinese and English. *Proceedings Interspeech*. 2024. pp. 947–951. DOI: 10.21437/Interspeech.2024-1807.
28. Chen X., Zhang W., Ma Y. Raw Waveform-Based End-to-End Alzheimer's Disease Detection Method. *Acta Electronica Sinica*. 2023. vol. 51. no. 12. pp. 3582–3590.
29. Псарёва Н.Н. Беглость спонтанной речи как предиктор мягкого когнитивного снижения. ВКР по программе бакалавриата. М.: ВШЭ, 2024.
30. Karakostas A., Briassouli A., Avgerinakis K., Kompatsiaris I., Tsolaki M. The dem@care experiments and datasets: a technical report. *arXiv preprint arXiv:1701.01142*. 2016.
31. Kishimoto T. et al. PROMPT collaborators. The project for objective measures using computational psychiatry technology (PROMPT): Rationale, design, and methodology. *Contemporary Clinical Trials Communications*. 2020. vol. 19. DOI: 10.1016/j.conctc.2020.100649.
32. Poor F.F. et al. Prediction of Mild Cognitive Impairment Using a Hybrid Audio-Visual Approach: An I CONECT Study. *Alzheimer's & Dementia Journal*. 2023. vol. 19. DOI: 10.1002/alz.074808.
33. Correia J., Teixeira F., Botelho C., Trancoso I., Raj B. The in-the-Wild Speech Medical Corpus. 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). 2021. pp. 6973–6977. DOI: 10.1109/ICASSP39728.2021.9414230.
34. Orozco J.R. et al. New Spanish speech corpus database for the analysis of people suffering from Parkinson's disease. *International Conference on Language Resources and Evaluation (LREC)*. 2014. pp. 342–347.
35. Gratch J. et al. The Distress Analysis Interview Corpus of human and computer interviews. *International Conference on Language Resources and Evaluation (LREC)*. 2014. pp. 3123–3128.

36. DeVault D. et al. SimSensei kiosk: A virtual human interviewer for healthcare decision support. Proceedings of the 13th International Conference on Autonomous Agents and Multiagent Systems (AAMAS'14). 2014. pp. 1061–1068.
37. Vashkevich M., Rushkevich Yu., Petrovsky A. Bulbar ALS Detection Based on Analysis of Voice Perturbation and Vibrato. Proceedings of inter. conf. Signal Processing: Algorithms, Architectures, Arrangements, and Applications (SPA). 2019. pp. 267–272. DOI: 10.23919/SPA.2019.8936657.
38. Qi X., Zhou Q., Dong J., Bao W. Noninvasive automatic detection of Alzheimer's disease from spontaneous speech: a review. *Frontiers in Aging Neuroscience*. 2023. vol. 15.
39. Ding K., Chetty M., Noori Hoshyar A., Bhattacharya T., Klein B. Speech based detection of Alzheimer's disease: a survey of AI techniques, datasets and challenges. *Artificial Intelligence Review*. 2024. vol. 57. no. 12.
40. Babu A. et al. XLS-R: Self-supervised cross-lingual speech representation learning at scale. *Proceedings of Interspeech*. 2022. pp. 2278–2282. DOI: 10.21437/Interspeech.2022-143.
41. Duan J., Wei F., Li H.D., Liu J. Pre-trained Feature Fusion and Matching for Mild Cognitive Impairment Detection. *Proceedings of Interspeech*. 2024. pp. 962–966.
42. Favaro A., Cao T., Dehak N., Moro-Velazquez L. Leveraging Universal Speech Representations for Detecting and Assessing the Severity of Mild Cognitive Impairment Across Languages. *Proceedings of Interspeech*. 2024. pp. 972–976. DOI: 10.21437/Interspeech.2024-2030.
43. Radford A., Kim J.W., Xu T., Brockman G., McLeavey C., Sutskever I. Robust speech recognition via large-scale weak supervision. *International conference on machine learning. PMLR* 2023. pp. 28492–28518.
44. Lee B.W., Lee J. LFTK: Handcrafted features in computational linguistics. *Proceedings of the 18th Workshop on Innovative Use of NLP for Building Educational Applications (BEA 2023)*. 2023. pp. 1–19. DOI: 10.18653/v1/2023.bea-1.1.
45. Hoang B., Pang Y., Dodge H., Zhou J. Translingual Language Markers for Cognitive Assessment from Spontaneous Speech. *Proceedings of Interspeech*. 2024. 977–981. DOI: 10.21437/Interspeech.2024-1422.
46. Fan A. et al. Beyond English-centric multilingual machine translation. *Journal of Machine Learning Research*. 2021. vol. 22. no. 107. pp. 1–48.
47. Pérez-Toro P.A., Arias-Vergara T., Klumpp P., Weise T., Schuster M., Noeth E., Orozco-Arroyave J.R., Maier A.. Multilingual Speech and Language Analysis for the Assessment of Mild Cognitive Impairment: Outcomes from the Taukadi Challenge. *Proceedings of Interspeech*. 2024. pp. 982–986. DOI: 10.21437/Interspeech.2024-2115.
48. Zheng C. et al. Detecting Dementia from Face-Related Features with Automated Computational Methods. *Bioengineering (Basel)*. 2023. vol. 10(7). DOI: 10.3390/bioengineering10070862.
49. Okunishi T., Zheng C., Bouazizi M., Ohtsuki T., Kitazawa M., Horigome T., Kishimoto T. Dementia and MCI Detection Based on Comprehensive Facial Expression Analysis from Videos During Conversation. *IEEE Journal of Biomedical and Health Informatics*. 2025. vol. 29(5). pp. 3537–3548. DOI: 10.1109/JBHI.2025.3526553.
50. Mu X. et al. Detecting Cognitive Impairment and Psychological Well-being among Older Adults Using Facial, Acoustic, Linguistic, and Cardiovascular Patterns Derived from Remote Conversations. *arXiv preprint arXiv:2412.14194*. 2024.
51. Jiang Z., Seyed S., Griner E., Abbasi A., Rad A.B., Kwon H., Cotes R.O., Clifford G.D. Multimodal Mental Health Digital Biomarker Analysis from Remote Interviews

- Using Facial, Vocal, Linguistic, and Cardiovascular Patterns. *IEEE Journal of Biomedical and Health Informatics*. 2024. vol. 28(3). pp. 1680–1691. DOI: 10.1109/JBHI.2024.3352075.
52. Bang J.-U., Han S.-H., Kang B.-O. Alzheimer's Disease recognition from spontaneous speech using large language models. *ETRI Journal*. 2024. vol. 46. no. 1. pp. 96–105. DOI: 10.4218/etrij.2023-0356.
53. Botelho C. et al. Macro-descriptors for Alzheimer's disease detection using large language models. *Proc. Interspeech 2024*. 2024. pp. 1975–1979.
54. Haulcy R., Glass J. CLAC: A Speech Corpus of Healthy English Speakers. *Proceedings of Interspeech*. 2021. pp. 2966–2970. DOI: 10.21437/Interspeech.2021-1810.
55. Dobrovolskii V. Word-Level Coreference Resolution. *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*. 2021. pp. 7670–7675. DOI: 10.18653/v1/2021.emnlp-main.605.
56. Agarwal R., Melnick L., Frosst N., Zhang X., Lengerich B., Caruana R., Hinton G.E. Neural additive models: Interpretable machine learning with neural nets. *Advances in Neural Information Processing Systems*. 2021. vol. 34. pp. 4699–4711.
57. Jin D. et al. What Disease Does This Patient Have? A Large-Scale Open Domain Question Answering Dataset from Medical Exams. *Applied Sciences*. 2021. vol. 11(14). DOI: 10.3390/app11146421.
58. Pal A. et al. MedMCQA: A Large-scale Multi-Subject Multi-Choice Dataset for Medical domain Question Answering. *Proceedings of Machine Learning Research*. 2022. vol. 174. pp. 248–260.
59. Arora R.K. et al. HealthBench: Evaluating Large Language Models Towards Improved Human Health. *arXiv preprint arXiv:2505.08775*. 2025.
60. Blinov P., Reshetnikova A., Nesterov A., Zubkova G., Kokh V. RuMedBench: A Russian Medical Language Understanding Benchmark. *Conference on Artificial Intelligence in Medicine*. 2022. pp. 383–392.
61. Tu T. et al. Towards generalist biomedical AI. *NEJM Ai*. 2024. vol. 1(3).
62. Driess D. et al. PaLM-E: an embodied multimodal language model. *Proceedings of the 40th International Conference on Machine Learning (ICML'23)*. 2023. vol. 202. pp. 8469–8488.
63. Kim H. et al. Small language models learn enhanced reasoning skills from medical textbooks. *NPJ Digital Medicine*. 2025. vol. 8(1). DOI: 10.1038/s41746-025-01653-8.
64. Wang Y. et al. Exploiting prompt learning with pre-trained language models for Alzheimer's Disease detection. *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. 2023. pp. 1–5. DOI: 10.1109/ICASSP49357.2023.10095993.
65. Balagopalan A., Shkaruta K., Novikova J. Impact of ASR on Alzheimer's Disease Detection: All Errors are Equal, but Deletions are More Equal than Others. *Proceedings of the Sixth Workshop on Noisy User-generated Text (W-NUT 2020)*. 2020. pp. 159–164. DOI: 10.18653/v1/2020.wnut-1.21.
66. Utkin L.V., Konstantinov A.V., Eremenko D.Yu., Zaborovsky V.S., Muliukha V.A. Interpretation methods for machine learning models in the framework of survival analysis with censored data: a brief over view. *Computing, Telecommunications and Control*. 2024. vol. 17. no. 3. pp. 22–31. DOI: 10.18721/JCSTCS.17302.
67. Alkhatib A. et al. Interpretable Graph Neural Networks for Tabular Data. *Proceedings of the 27th European Conference on Artificial Intelligence (ECAI 2024)*. 2024. vol. 392. pp. 1848–1855.
68. Ren H., Zheng Y., Li C., Jing F., Wang Q., Luo Z., Li D., Liang D., Tang W., Liu L., Cheng W. Using Machine Learning to Predict Cognitive Decline in Older Adults

From the Chinese Longitudinal Healthy Longevity Survey: Model Development and Validation Study. JMIR Aging. 2025. vol. 8. DOI: 10.2196/67437.

69. Karpov A.A., Yusupov R.M. Multimodal Interfaces of Human-Computer Interaction. Herald of the Russian Academy of Sciences. 2018. vol. 88(1). pp. 67–74. DOI: 10.1134/S1019331618010094.

Dolgushin Mikhail — Junior researcher, postgraduate student, Speech and multimodal interfaces laboratory, St. Petersburg Federal Research Center of the Russian Academy of Sciences (SPC RAS). Research interests: speech technology, automatic speech recognition, explainable artificial intelligence, computational linguistics. The number of publications — 12. dolgushin.m@iias.spb.su; 39, 14th Line of V.O., 199178, St. Petersburg, Russia; office phone: +7(812)328-0421.

Karpov Alexey — Ph.D., Dr.Sci., Professor, Head of the laboratory, Speech and multimodal interfaces laboratory, St. Petersburg Federal Research Center of the Russian Academy of Sciences (SPC RAS). Research interests: speech technology, automatic speech recognition, audio-visual speech processing, multimodal human-computer interfaces, and computational paralinguistics. The number of publications — 450+. karpov@iias.spb.su; 39, 14th Line of V.O., 199178, St. Petersburg, Russia; office phone: +7(812)328-0421.

Acknowledgements. This research is financially supported by the Russian Science Foundation, project No. 25-11-00319, with the exception of Section 6, which was conducted under the state research No. FFZF-2025-0003.

С.Н. КАРПОВИЧ, А.В. СМИРНОВ, Н.Н. ТЕСЛЯ
**МЕТОДОЛОГИЯ ПОСТРОЕНИЯ ЭТАЛОННОГО ТЕСТА
ДЛЯ ОЦЕНКИ РАБОТЫ LLM С ЧИСЛИТЕЛЬНЫМИ**

Карпович С.Н., Смирнов А.В., Тесля Н.Н. Методология построения эталонного теста для оценки работы LLM с числительными.

Аннотация. В статье представлена методология разработки эталонного теста для оценки навыков числового мышления в больших языковых моделях (Large Language Models, LLM). Под числовым мышлением в контексте LLM понимается способность модели корректно интерпретировать, обрабатывать и использовать числовую информацию в тексте – включая понимание значений чисел, их соотношений, выполнение арифметических операций, а также корректную генерацию числительных в ответах. Предложенная методология основана на декомпозиции прикладных задач и позволяет оценивать отдельные аспекты числового мышления на примере задач с числительными. Особое внимание уделяется способу представления чисел в текстовых инструкциях к LLM, поскольку это напрямую влияет на качество итогового ответа. Необходимость точной оценки числового мышления LLM обусловлена тем, что эта способность критически важна для широкого спектра прикладных задач работы с текстами, в том числе для автоматизированного составления кратких изложений, генерации аналитических отчётов, извлечения и интерпретации количественных данных, а также для диалоговых систем, работающих с финансовой, научной или технической информацией. На основе анализа современных подходов к оценке LLM сформулированы основные принципы построения эталонных тестов с упором на универсальность и применимость в реальных сценариях. В соответствии с предложенной методологией разработан эталонный тест MUE (Math Understanding Evaluation), включающий пять наборов тестовых заданий, каждый из которых предназначен для оценки отдельного аспекта числового мышления LLM. Проведена сравнительная оценка качества популярных LLM, определены лидеры, а также выявлены сильные и слабые стороны их числового мышления. Полученные результаты могут использоваться разработчиками LLM для улучшения архитектур и стратегий обучения, а также конечными пользователями и интеграторами для выбора оптимальной модели в прикладных проектах.

Ключевые слова: методология, большие языковые модели (LLM), эталонный тест LLM, обработка естественного языка (NLP), числительные.

1. Введение. В настоящее время большие языковые модели (LLM), такие как GPT [1], Gemini [2], Mistral [3], LLaMA [4], GigaChat [5], YandexGPT [6], Qwen, DeepSeek [7, 8] успешно применяются при создании цифровых продуктов в образовании, здравоохранении, юриспруденции, информационных технологиях, маркетинге, научных исследованиях, государственном секторе. Под большими языковыми моделями (Large Language Models, LLM) понимаются такие языковые модели, масштаб которых и архитектура позволяют им демонстрировать эмерджентные свойства – функциональные возможности, не наблюдаемые в меньших по размеру

моделях. Как правило, LLM содержат от сотен миллионов до десятков миллиардов параметров и обучены на обширных корпусах текстов, что обеспечивает способность решать широкий спектр задач без специализированного дообучения. Основные прикладные задачи, решаемые с помощью LLM, включают генерацию текста, перевод, редактирование и улучшение текста, анализ и классификация, ответы на вопросы и диалоги, обработка естественного языка, извлечение именованных сущностей.

Для таких продуктов и задач важно обеспечить высокое качество и стабильность ответов. LLM галлюцинируют (генерируют правдоподобную, но фактически неверную информацию), генерируют ошибки, дают некорректные или неожиданные ответы. Это требует внедрения дополнительных механизмов валидации и контроля качества создаваемого контента.

Одним из способов повышения надежности является оптимизация формулировок текстовых инструкций – промптов. Промпт-инженеры экспериментально подбирают конфигурации, обеспечивающие стабильные и точные ответы. Однако этот процесс осложняется большим количеством языков, наличием синонимов и различных формулировок для одной и той же задачи.

Особую сложность для LLM представляют задачи, связанные с обработкой и интерпретацией числовых данных [8 – 15]. Например, может быть поставлена задача сгенерировать текст, ограниченный заданным количеством символов, или выполнить арифметические операции, встроенные в текстовый запрос. Несмотря на то, что современные LLM демонстрируют впечатляющие результаты в обработке текста, их способность к точным математическим расчетам и интерпретации чисел остается недостаточно изученной и, зачастую, ограниченной.

Основной целью исследования является разработка методологии построения эталонного теста (бенчмарка) для систематического тестирования LLM в части работы с числительными. Методология должна определить принципы оценки способности моделей интерпретировать, обрабатывать и генерировать числовую информацию в различных форматах и задачах. В рамках исследования, представленного в данной статье, были проанализированы существующие подходы и эталонные тесты, выявлена их применимость к анализу числительных; разработана методология построения эталонного теста с набором заданий, критериями и процедурой тестирования; проведено сравнительное тестирование популярных LLM (GPT, Gemini, Llama, Claude и др.); проанализированы результаты,

выделены сильные и слабые стороны моделей, классифицированы наиболее частые ошибки и даны практические рекомендации по формулировке запросов с числительными.

Материал статьи структурирован следующим образом. Во втором разделе обосновывается важность разработки методологии для оценки работы больших языковых моделей (LLM) с акцентом на обработку числительных. В третьем разделе представлен обзор существующих подходов к тестированию LLM, который включает два направления: эталонные тесты для оценки математических способностей (3.1) и эталонные тесты для анализа общего понимания, знаний и рассуждений (3.2). Четвертый раздел содержит формальную постановку задачи (4.1), раскрывает предложенную методологию построения эталонного теста, включая основные принципы (4.2) и детальное описание этапов разработки (4.3), ориентированных на оценку точности, устойчивости и интерпретируемости моделей при работе с числительными. Пятый раздел демонстрирует практическое применение методологии: описывается структура разработанного набора данных (5.1), и представлен эталонный тест MUE (5.2), предназначенный для анализа работы LLM с числовыми данными. Шестой раздел содержит результаты экспериментов, подтверждающих эффективность предложенного подхода. В заключении сформулированы выводы о научной и практической ценности методологии, а также обозначены перспективы дальнейших исследований.

2. Актуальность разработки методологии построения эталонного теста для оценки работы LLM с числительными. При подготовке входных данных к обработке каждая LLM использует два последовательных шага: токенизацию и векторизацию. Токенизация – это процесс разбиения текста на отдельные части (токены), которые служат минимальными единицами анализа для последующей обработки в задачах NLP. Эти токены могут быть словами, подсловами, символами, числами, знаками препинания или даже фразами, в зависимости от выбранного метода. Токенизация является ключевым этапом в обработке текста, который определяет, как модель будет интерпретировать данные. Современные методы токенизации включают алгоритм Byte Pair Encoding (BPE) [16], который используется в GPT и RoBERTa, алгоритм WordPiece [17], применяемый в моделях BERT, и алгоритм SentencePiece [18], используемый в T5 и ALBERT. Векторизация – это процесс преобразования токенизированных текстовых данных в числовые векторы (эмбеддинги), которые могут быть обработаны алгоритмами

машинного обучения и глубокого обучения, с сохранением семантических и синтаксических свойств текста.

Разные подходы к токенизации и векторизации в LLM приводят к различным векторным представлениям одного и того же текста. Это актуально для числительных. В то время как для человека различные формы записи числа (текстовые, на разных языках – «сто двадцать три», «one hundred twenty-three», «cent vingt-trois»; цифровые – «СХХIII», «123») имеют одинаковое значение, для модели они могут представлять существенно разные последовательности токенов. Это приводит к соответствующим различиям в векторных представлениях, степень семантической близости которых определяется, например, косинусным расстоянием. Такие различия могут влиять на способность LLM корректно интерпретировать числовые данные и выполнять соответствующие задачи.

В качестве примера рассмотрим косинусные расстояния между векторными представлениями числа «123», полученными с помощью API большой языковой модели GigaChat (таблица 1), рассчитанное по формуле 1, где V_1 и V_2 – векторные представления токенов.

$$\bar{V}_1 = (v_{1x}, v_{1y}); \bar{V}_2 = (v_{2x}, v_{2y}),$$

$$Sim_{\cos}(\bar{V}_1, \bar{V}_2) = \frac{(\bar{V}_1, \bar{V}_2)}{|\bar{V}_1| * |\bar{V}_2|} = \frac{v_{1x} * v_{2x} + v_{1y} * v_{2y}}{\sqrt{v_{1x}^2 + v_{1y}^2} * \sqrt{v_{2x}^2 + v_{2y}^2}}. \quad (1)$$

Таблица 1. Пример косинусных расстояний между векторными представлениями числа 123 в различных форматах записи

	Русский	Английский	Французский	Китайский	Цифры
Русский	1,00	0,95	0,93	0,91	0,87
Английский		1,00	0,95	0,90	0,89
Французский			1,00	0,91	0,87
Китайский				1,00	0,87
Цифры					1,00

Данные таблицы 1 свидетельствуют о влиянии формы представления числительного на его внутренне представление и интерпретацию в LLM. При схожем семантическом содержании текстовых форм на русском, английском и французском языках (косинусное расстояние близко к 1), представление на китайском уже заметно отличается (0,91). Наибольшее расхождение наблюдается между текстовыми и числовыми формами (например, 0,87 для

русского), что может вызывать различия в реакциях LLM на инструкции.

При разработке методологии построения эталонного теста необходимо определить принципы отбора задач для оценки числового мышления на пересечении математики и языкового понимания. Методология должна быть ориентирована на специалистов, использующих LLM в прикладных задачах, охватывающих аспекты числового мышления (восприятие чисел – идентификация и извлечение числовых значений из текста, выполнение базовых арифметических действий, сравнение числовых значений, логические рассуждения – поиск закономерностей в последовательностях чисел) и языкового понимания (контекстуальное и лексическое понимание, причинно-следственные связи, анализ текста, ответы на вопросы на основе общих знаний).

3. Обзор существующих эталонных тестов для анализа больших языковых моделей. Эталонные тесты LLM представляют собой наборы задач и методик оценки для сравнения больших языковых моделей по заранее установленным метрикам. Они могут включать тесты на понимание и генерацию текста, перевод, краткое изложение, ответы на вопросы и другие задачи, требующие обработки естественного языка. Наиболее известные международные коллекции эталонных тестов представлены на платформе Hugging Face [19], где собраны такие популярные тестовые наборы как Instruction-Following Evaluation (IFEval) [20], Big Bench Hard (BBH) [21], Mathematics Aptitude Test of Heuristics (MATH) [8], Graduate-Level Google-Proof Q&A (GPQA) [22], Multistep Soft Reasoning (MuSR) [23] и Massive Multitask Language Understanding–Professional (MMLU-Pro) [24]. Создаются специализированные сравнительные таблицы, объединяющие результаты оценки моделей по нескольким эталонным тестам, такие как, например, MERA, для оценки русскоязычных моделей [25]. В целом, основной задачей эталонных тестов является стандартизация процесса оценки LLM, предоставляя единый подход к оценке посредством единых задач и метрик для сравнения и отслеживания прогресса моделей.

Далее рассмотрим основные эталонные тесты, используемые для оценки языковых моделей.

3.1. Эталонные тесты для оценки математических способностей. Эталонный тест HARDMATH [8] сфокусирован на задачах прикладной математики. Он включает задания на асимптотический анализ, работу с полиномами (нормализация,

нахождение корней), решение нелинейных дифференциальных уравнений, вычисление интегралов и задачи со сложным контекстом.

Набор задач DeepMind [9] разработан для оценки математического рассуждения LLM и состоит из последовательных текстовых вопросов, охватывающих различные области математики: от арифметики до анализа.

Набор данных MATH [10] содержит 12 500 сложных математических задач уровня соревнований с пошаговыми решениями, классифицированных по пяти уровням сложности. Он предназначен для углублённой оценки способности моделей в математическом рассуждении.

Мультимодальный эталонный тест MATHVISTA [11], предназначен для оценки способностей математического мышления ИИ в визуальных контекстах. Он содержит 6 141 пример, объединяющий задачи из различных областей математики и визуальные данные (из 28 существующих и трех новых наборов данных: IQTest, FunctionQA, PaperQA).

Эталонный тест FrontierMath [12] состоит из оригинальных и ранее не публиковавшихся математических задач исследовательского уровня (теория чисел, алгебраическая геометрия и др.), требующих творческого подхода.

NumericBench [13] оценивает базовые навыки числового мышления, необходимые для реальных приложений, таких как анализ финансовых данных или прогнозирование погоды, включающие в себя распознавание чисел, арифметические операции, контекстное извлечение, сравнение, суммирование, логическое рассуждение.

Эталонный тест NUPA [14] охватывает четыре основных представления чисел (целые, вещественные числа, дроби и научная нотация) и 17 различных типов задач, связанных с их обработкой.

Эталонный тест Numberland [15] оценивает способности к числовому мышлению, абстрактному пониманию чисел и их взаимосвязей («числовой интуиции»), которое позволяет решать математические задачи с ограниченными вычислительными ресурсами.

3.2. Эталонные тесты для оценки общего понимания, знаний и рассуждений. Набор данных MuSR (Multi-Step Reasoning) [23] разработан для оценки способности LLM к многошаговым «мягким» рассуждениям на основе текстовых нарративов (детективные истории, задачи на размещение объектов и другие).

Эталонный тест MMLU (Massive Multitask Language Understanding) [24], представленный на конференции ICLR 2021, оценивает многозадачное понимание и знания LLM на 57

разнообразных темах разного уровня сложности (от элементарной математики и истории до права и информатики). Оценка проводится в режимах «Zero-Shot» (промпты без примеров результата работы LLM) и «Few-Shot» (промпты с одним и более примерами результата работы LLM).

Для комплексной оценки LLM на русском языке предложен открытый эталонный тест MERA (Multimodal Evaluation of Russian-language Architectures) [25]. Он включает в себя 21 задачу, являющиеся сложными для фундаментальных моделей: вопросы охватывают знания о мире, логику, причинно-следственные связи, этику ИИ и многое другое.

SuperGLUE [26] – это расширенный и усложненный эталонный тест, пришедший на смену GLUE, для оценки общего понимания языка системами NLP. Включает в себя 8 задач (BoolQ, CommitmentBank, COPA, MultiRC, ReCoRD, RTE, WiC, WSC), требующих глубокого языкового понимания и способностей к рассуждению для решения разнообразных задач.

RussianSuperGLUE [27] является адаптированным для русского языка аналогом эталонного теста SuperGLUE. Включает 9 задач (LiDiRus, RUSSE, PARus, TERRa, RCB, RWSD, MuSeRC, RuCoS, DaNetQA), каждая из которых оценивает различные аспекты понимания русского языка, такие как разрешение языковых неоднозначностей, понимание причинно-следственных связей, распознавание текстового следствия и машинное чтение.

BIG-Bench Hard (BBH) [21] представляет собой поднабор из 23 особо сложных задач, выбранных из более широкого эталонного теста BIG-Bench. Эти задачи были отобраны, так как предыдущие LLM не могли превзойти на них средний человеческий результат, и они требуют продвинутых способностей к многошаговому рассуждению (включая такие виды, как алгоритмическое, арифметическое, на основе знаний о мире и понимания языка).

IFEval [20] – это эталонный тест, разработанный для объективной и автоматизированной оценки способности моделей LLM следовать инструкциям, заданным на естественном языке.

Эталонный тест GPQA (Graduate-Level Google-Proof Q&A) [22] состоит из 448 сложных вопросов с множественным выбором по биологии, физике и химии, созданных и проверенных экспертами. Вопросы спроектированы так, чтобы ответ на них было трудно найти с помощью поиска («Google-Proof»).

В таблице 2 приведён систематизированный перечень рассмотренных эталонных тестов с указанием их функциональной

направленности (математические задачи или оценка языкового понимания), а также характеристик: уровень сложности тестовых заданий, частота обновления набора данных, возможность расширения пользовательскими элементами. Знаком вопроса обозначено отсутствие информации о возможностях эталонного теста.

Таблица 2. Сравнение эталонных тестов по типам задач

Название эталонного теста	Оценка уровня сложности	Обновляемость набора заданий	Возможность расширения
Математические задачи			
HARDMATH [8]	Высокий	–	+
Набор задач DeepMind [9]	От элементарного до высокого	+	+
MATH [10]	Высокий	–	?
MATHVISTA [11]	Высокий	–	?
FrontierMath [12]	Исключительно высокий	–	–
NumericBench [13]	Высокий	–	?
NUPA [14]	От элементарного до высокого	–	?
Numberland [15]	От элементарного до высокого	–	–
Языковое понимание			
MuSR [23]	Высокий	+	+
MMLU [24]	Высокий	–	?
MERA [25]	Высокий	–	–
SuperGLUE [26]	Высокий	–	–
RussianSuperGLUE [27]	Высокий	–	–
BIG-Bench Hard [21]	Высокий	–	–
IFEval [20]	Средний	–	+
GPQA [22]	Высокий	–	–

Анализ таблицы 2 показывает, что существующие эталонные тесты охватывают широкий диапазон тестовых сценариев – от прикладных вычислительных и мультимодальных задач до комплексных проверок предметных знаний, логических рассуждений и многошагового вывода. При этом сохраняются области, в которых оценка возможностей LLM носит неполный или фрагментарный характер. Заметен недостаток эталонных тестов, предназначенных для комплексного тестирования числового мышления и языковой компетентности, имеющих ключевое значение в ряде прикладных задач. Установлено, что большинство рассмотренных наборов заданий имеют высокий уровень сложности, что целесообразно в

академической среде, однако не обеспечивает репрезентативной оценки производительности LLM в условиях простых прикладных задач. Кроме того, подавляющее число эталонных тестов не обновляются, что повышает вероятность предобученности моделей на исходных данных и не позволяет объективно оценить их способность к обобщению. Небольшое число эталонных тестов допускает добавление новых тестовых заданий, что существенно ограничивает их применимость в динамично изменяющихся условиях. Разработка эталонных тестов, обеспечивающих комплексную оценку нескольких аспектов работы LLM, включающих задания различной сложности, обладающих регулярной обновляемостью и возможностью интеграции пользовательских тестов, представляется перспективным направлением, способным устранить указанные недостатки и расширить функциональную применимость систем оценки.

4. Методология построения эталонного теста для оценки работы LLM с числительными. Несмотря на наличие множества подходов для оценки LLM (например, CheckList [28]), сохраняется потребность в инструментах, направленных на систематическое исследование способности моделей инвариантно воспринимать числовые данные в различных формах записи. Это имеет значение при проверке их эффективности в прикладных задачах, требующих точной обработки числовой информации. Разработка методологии для построения кастомизированных эталонных тестов представляет собой важную стратегическую задачу, позволяя каждому практику и исследователю самостоятельно проводить оценку LLM и выполнять проверку качества информационных систем.

4.1. Постановка задачи. Пусть N множество чисел. Для каждого числа $n \in N$ определим множество его текстовых и символьных представлений $R(n)$. Это множество включает:

- Цифровую запись: $r_{\text{dig}}(n)$, например, 123;
- Запись римскими цифрами: $r_{\text{rom}}(n)$, например, CXXII;
- Текстовую запись на различных языках.

Пусть $L = \{l_1, l_2, l_3, \dots, l_k\}$ – множество естественных языков (например, l_1 – русский, l_2 – английский). Тогда $r_{\text{text},l}(n)$ – это текстовое представление числа n на языке $l \in L$. Для текстовой записи возможны два варианта, когда задание и числительные написаны на одном языке и на разных, например задание дано на русском языке, а другой язык использован только для числительного.

Таким образом множество всех представлений числа n можно определить как:

$$R(n) = \{r_{\text{dig}}(n), r_{\text{rom}}(n)\} \cup \{r_{\text{text}, l}(n) \mid l \in L\}.$$

Пусть M – это большая языковая модель, которую рассматриваем как функцию, преобразующую множество входных текстовых промптов P в всех возможных ответов модели O :

$$M : P \rightarrow O.$$

Определим тестовое задание T как кортеж (упорядоченный набор) эталонных ответов $a_i^{\text{ref}} \in A$, которые могут быть числом, текстом или другой структурой, соответствующих промпту $p_i \in P$. При этом будем считать, что часть возможных ответов модели входит в множество эталонных ответов:

$$T_i = (p_i, a_i^{\text{ref}}), \exists a_i^{\text{ref}} : a_i^{\text{ref}} \in O, A \cap O \neq \emptyset.$$

Задача заключается в создании эталонного теста B , представляющего собой множество тестовых заданий T_i – кортежей, содержащих промпт $p_i \in P$ и эталонный ответ $a_i^{\text{ref}} \in A$, таких, чтобы они содержали все возможные представления чисел из $R(n)$:

$$B = \{T_1, T_2, \dots, T_m\}.$$

Одно из основных требований к эталонному тесту заключается в том, что он должен содержать группы задач, построенные на одной и той же логической основе, но с использованием разных представлений чисел. В соответствии с этим, определим G_j как j -ю группу задач. Все задачи в этой группе имеют общую семантическую основу и единый эталонный ответ $a_i^{\text{ref}} \in A$, но используют разные вариации промпта. Промпты $p_{ij} \in P$ создаются путем подстановки различных представлений $r \in R(n)$ одних и тех же чисел n в базовый шаблон промпта ψ_j :

$$G_j = \left\{ (p_{i1}, a_j^{\text{ref}}), (p_{i2}, a_j^{\text{ref}}), \dots, (p_{ik}, a_j^{\text{ref}}) \right\},$$

где $p_{ij} = \psi_j(r_1, r_2, \dots)$ с различными $r_k \in R(n_k)$.

Для оценки результата введем функцию сравнения EM , которая определяет, является ли вывод модели o_i эквивалентным эталонному ответу a_i^{ref} :

$$EM(o_i, a_i^{\text{ref}}) \rightarrow \{0, 1\},$$

где 1 означает правильный ответ, а 0 – неправильный. Функция EM устойчива к формату, например, $EM(\text{cmo}, 100) = 1$.

В данной постановке задачи обеспечивается оценка робастности, устойчивости к формату представления чисел – способность модели давать одинаково правильный результат на семантически эквивалентных задачах, независимо от формы записи чисел. Для этого в каждой группе задач G_j вычисляется доля правильных ответов. Идеально робастная модель будет иметь $Acc(M, G_j) = 1$ для всех j , где она знает правильный ответ.

4.2. Основные принципы методологии. По результатам анализа существующих наборов данных и эталонных тестов были сформулированы следующие принципы методологии построения эталонного теста для оценки работы LLM.

1) Ориентация на задачу. Полученный эталонный тест должен гарантировать, что критерии оценки и метрики непосредственно связаны с реальными целями применения модели.

2) Изоляция тестирования отдельных навыков модели. Многие задачи требуют комплексного подхода при их решении и могут быть разделены на отдельные подзадачи. Принцип изоляции отдельных навыков позволяет точно определить сильные и слабые стороны модели, избегая смешения эффектов при проверке сложных многокомпонентных задач. Для упрощения сложных задач до более простых, атомарных используется подход «Downward Evolution» [29].

3) Репрезентативность и валидность метрик в общей и частных задачах. Метрики, используемые в эталонном тесте, должны отражать как общую эффективность модели по задаче, так и адекватно оценивать каждый навык в отдельности. Валидность в данном принципе означает, что полученные оценки действительно соответствуют заявленным аспектам поведения модели.

4) Репрезентативность набора данных для тестирования. Набор тестовых примеров должен покрывать значимые сценарии и вариативность входных данных, чтобы результаты были применимы к реальным условиям.

5) Воспроизводимость результатов тестирования. Процедуры запуска эталонных тестов, настройки окружения и расчёта метрик документируются так, чтобы при повторном проведении (иными исследователями) получались идентичные выводы.

4.3. Процесс построения эталонного теста. Исходя из сформулированных принципов, процесс построения эталонного теста в рамках методологии разделен на этапы, представленные на рисунке 1. Рассмотрим подробнее методы и подходы, используемых на каждом этапе методологии построения эталонного теста для оценки работы LLM на примере задачи краткого изложения текста.

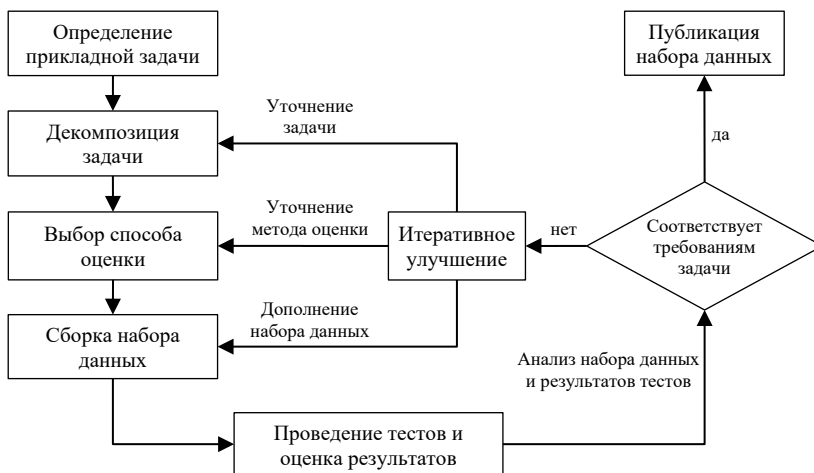


Рис. 1. Ключевые этапы процесса создания эталонного теста

Определение прикладной задачи. На первом этапе необходимо четко определить задачу для тестирования, где корректная работа LLM является важным компонентом. Чем детальнее будет определена задача и требования к ее выполнению на данном этапе, тем легче будет определить последующие шаги по разработке тестов. В задаче краткого изложения языковая модель должна не только провести анализ смысла текста, извлечь сущностную часть и кратко ее изложить, но и учесть инструкции, указывающие на ограничение

количества символов в результате, а также сохранить смысл всех важных фрагментов, выраженных числительными.

Декомпозиция задачи до минимальных простых подзадач. Этот шаг позволяет изолированно тестировать специфические навыки, и необходим для стандартизации тестов, объективизации оценки и унификации результатов.

Применительно к задаче краткого изложения и другим задачам, требующим работы с числительными, можно выделить следующие типы простых подзадач:

Соблюдение количественных ограничений:

- Генерация текста заданной длины (в символах, словах, предложениях, токенах).
- Выбор определенного количества элементов из списка.
- Ответ на вопрос с использованием точного числа.

Понимание и интерпретация числительных:

- Распознавание чисел, записанных цифрами (арабскими, римскими) и словами (на разных языках).
- Сравнение чисел.
- Выполнение простых арифметических инструкций, встроенных в текст.
- Извлечение числовых значений из текста.

Базовое понимание структуры текста (опосредованно влияющее на работу с числами в контексте):

- Определение слова, предложения, знака препинания.
- Понимание частей речи (существительное, прилагательное, глагол), что может быть важно для интерпретации инструкций с числами (например, «найди первые три существительных»).
- Понимание позиции слова в предложении.

Хотя нет возможности предвидеть все варианты корректных свободных ответов на сложную задачу целиком, можно провести тестирование на понимание числительных, используя задания с числами, представленными на разных языках и в разной форме (цифры, текст). Также есть возможность подготовить набор данных с заданным входным текстом и просить модель обрабатывать его или извлекать нужные части этого текста для точной сверки с эталоном.

Выбор способа оценки. На следующем этапе важно определить, каким способом будет оцениваться результат. В различных эталонных тестах используются разнообразные метрики и подходы:

- Точное совпадение (Exact Match, EM) [30, 13 – 15] – актуально для задач с единственно верным числовым ответом или при проверке точного воспроизведения числовых ограничений.

- F1-мера (F1 Score) – используется, когда важны и точность, и полнота (например, при извлечении набора чисел).
- IFEval [20] – оценка следования верифицируемым инструкциям (Instruction-Following Eval).
- BERTScore [32], ROUGE [32] – для оценки семантической близости и совпадения n-грамм, полезны при оценке качества сгенерированного текста, содержащего числа.
- Оценка человеком (Human Judgment) [33, 34] – незаменима для комплексной оценки, особенно для новых или сложных задач.
- Анкетирование и сравнение с эталоном [35] – структурированный сбор мнений экспертов или пользователей.
- Метрика pass@k [36] – часто используется в задачах генерации кода, но может быть адаптирована.
- Подход «LLM-as-a-judge» [37] – использование одной LLM для оценки результатов другой.
- Коэффициент корреляции Мэтьюса (Matthews Correlation Coefficient, MCC) [38].

Для задач, связанных с числительными, важны метрики, способные оценить точность (EM), соблюдение количественных ограничений (например, проверка длины) и корректность извлечения или генерации числовых значений. В таблице 3 представлены способы оценки, применяемые в различных эталонных тестах.

Сборка набора данных. Это самая трудоемкая часть работы. Важно учитывать следующие требования:

- источники данных должны быть релевантны задачам;
- данные должны быть подобраны так, чтобы максимально эффективно тестировать целевые навыки (например, разнообразие форматов числительных, разные диапазоны значений);
- возможна генерация синтетических данных или использование шаблонов для создания тестовых примеров с числами;
- нужна валидация данных на корректность, однозначность;
- набор данных должен быть достаточным по объему для получения статистически значимых результатов.

Проведение тестов и оценка результатов. Этот этап включает запуск эталонного теста на различных моделях и сопоставление результатов. Важно проводить оценку репрезентативности сделанных замеров и анализировать не только количественные показатели, но и качественные аспекты ошибок. Если полученный тест соответствует требованиям к задаче он отправляется на публикацию, иначе принимается решение о необходимости улучшения набора данных.

Итеративное улучшение. Построение эталонного теста представляет собой итеративный процесс, включающий регулярное пополнение набора данных новыми примерами, особенно для пограничных случаев и выявленных слабых мест моделей, а также корректировку формулировки задачи и добавление новых подзадач, что обеспечивает актуальность и релевантность эталонного теста.

Таблица 3. Способы оценки результата, применяемые в эталонных тестах

Название эталонного теста	Метод оценки
HARDMATH	EM, Human Evaluation
Набор задач DeepMind	EM, Accuracy
MATH	EM
MATHVISTA	Accuracy
FrontierMath	Human Evaluation
NumericBench	EM
NUPA	EM
Numberland	EM
MuSR	Human Evaluation
MMLU	Human Evaluation
MERA	EM, pass@k, Accuracy, Human Evaluation
SuperGLUE	Accuracy, EM
RussianSuperGLUE	Accuracy, EM, F1, Matthews Correlation Coefficient (MCC)
BIG-Bench Hard	Human Evaluation
IFEval	IFEval
GPQA	Human Evaluation

5. Пример использования предложенной методологии построение эталонных тестов. В качестве примера применения предложенной методологии рассмотрено построение эталонного теста для оценки «числового мышления» больших языковых моделей (LLM) на задаче краткого изложения текста. Эта задача, помимо требований сжатого и точного воспроизведения содержания, часто включает работу с количественными данными, что позволяет выявить способность LLM корректно интерпретировать и воспроизводить числовую информацию. Для исключения влияния лингвистической вариативности предложена декомпозиция на подзадачи: (1) проверка соблюдения количественных ограничений на длину ответа, включающая тестирование понимания различных форм числовых инструкций (цифрами и словами) и оценку процента успешного соблюдения; (2) проверка точности воспроизведения ключевых числовых данных с использованием

метрики точного совпадения (ЕМ) для чисел. Такой подход обеспечивает изоляцию проверяемых навыков и позволяет выявлять сильные и слабые стороны LLM в аспекте числового мышления.

5.1. Описание структуры набора данных. Для автоматизации тестирования и обеспечения воспроизводимости результаты должны быть представлены в структурированном виде. В методологии предложен стандартизированный формат на основе JavaScript Object Notation (JSON), обеспечивающий четкую структуру и однозначное разделение компонентов тестовых заданий.

Спецификация структуры JSON-объекта, представляющего одно тестовое задание, приведена в Листинге 1.

```
{
  "instruction": "Выполни текстовое задание {inputs}",
  "inputs": "Напиши сто пятнадцать",
  "outputs": ["115", "115.", "сто пятнадцать", "CXV"],
  "meta": {
    "id": 14,
    "task_type": "Кодирование числовых данных",
    "type_input": "arabic_num"
  }
}
```

Листинг 1. Пример структуры JSON-файла элементарного задания на проверку числового мышления

Каждый объект в Листинге 1 включает следующие обязательные поля:

- **instruction**: инкапсулирует базовую инструкцию для оцениваемой LLM, функционально эквивалентную системному промпту (system prompt). Для оценки базовых способностей LLM к числовому мышлению, формулировка инструкции целенаправленно исключает применение продвинутых методик промпт-инжиниринга (например, «Few-Shot» или «Chain-of-Thought»), если только тест не направлен на изучение их влияния на результат.

- **inputs**: содержит текст непосредственной задачи (например, исходный текст для краткого изложения или более конкретный вопрос на понимание числа).

- **outputs**: представляет собой список строк, каждая из которых является одним из допустимых эталонных (правильных) вариантов ответа.

– meta: вспомогательный раздел для хранения метаданных (например, идентификатор задачи, задача, тип записи задачи числовой или текст).

5.2. Описание созданного эталонного теста. По предложенной методологии построен эталонный тест MUE (Math Understanding Evaluation, <https://github.com/cimsweb/MUE>), который состоит из пяти групп тестовых заданий, каждая из которых оценивает отдельную способность LLM, составляющую «числовое мышление»:

– MUE-1 (Кодирование числовых данных). Оценивается способность LLM давать одинаково правильный результат на семантически эквивалентных задачах, несмотря на разную форму записи чисел.

– MUE-2 (Анализ текстовых элементов). Оценивается способность LLM идентифицировать и оперировать элементами текста по их количественным признакам (например, «напиши второе предложение», «сколько слов в этом тексте?», «какой текст короче?»), а также выполнять перестановку и замену элементов.

– MUE-3 (Количественный пересчет). Предназначен для оценки способности LLM выполнять подсчет объектов, указанных в инструкции или тексте (например, «Напиши три точки»).

– MUE-4 (Сравнение числовых значений). Проверяет способность LLM сравнивать численные значения (например, «Какое число больше: А) 1 999 000 или Б) 1 млн?»).

– MUE-5 (Интерпретация и применение формул). Оценивает понимание LLM общеизвестных количественных фактов, свойств объектов и единиц измерения (например, «Сколько секунд в сутках?»), а также способность конвертации единиц измерения и выполнения базовых статистических операций (медиана, среднее).

Каждая группа задач в наборе данных включает в себя 100 заданий, предназначенных для оценки языковых моделей. Количество заданий определяется проверяемой задачей, минимально достаточное количество определяется экспериментальным путем. Эти задания структурированы таким образом, чтобы охватить различные форматы представления числительных и языковые особенности постановки задачи. В частности, числительные записаны в следующих форматах: арабскими цифрами, римскими цифрами, русским текстом, английским текстом в русскоязычной постановке задания, английским текстом в англоязычной постановке задания. Эта многоформатная структура позволяет провести комплексный анализ робастности языковых моделей к вариациям в представлении численных данных и языковых контекстах.

Оценка задач в эталонном тесте MUE осуществляется с использованием метрики Exact Match (EM), в рамках которой для каждого примера присваивается значение 1, если целевая последовательность точно совпадает с предсказанной последовательностью (одним из вариантов ответов, перечисленных в поле outputs) (2). В противном случае присваивается значение 0. Общий результат вычисляется как среднее значение точности по всем последовательностям (3).

$$EM(A, R) = \begin{cases} 1, & A = R \\ 0, & A \neq R \end{cases}, \quad (2)$$

$$EM_{\text{total}} = \frac{1}{N} \sum_{i=1}^N EM(A_i, R_i). \quad (3)$$

При тестировании всех LLM параметр температуры в API устанавливается на низкое значение (0.1) для повышения детерминированности и воспроизводимости результатов. Перед сравнением ответов выполняется их очистка от артефактов форматирования (пробелов, кавычек и др.), затрудняющих оценку. В ходе исследований выявлены характерные особенности генерации, требующие унификации очистки и приведения результатов к единому формату, сопоставимому с эталонными ответами, с целью исключения ложных срабатываний при оценке. Целесообразно удалять пробелы, переносы строк, специальные символы в начале и конце текста. DeepSeek [22] добавляет дополнительные элементы разметки «*think*», в которых описывается процесс рассуждения. В экспериментах ожидается, что модель предоставит точный ответ, строго соответствующий формулировке задания. В статье IFEval [9, 20] предложен подход, позволяющий снизить количество ложно отрицательных ответов LLM, однако для текущего эталонного теста использование такого метода видится избыточным.

Для формирования воспроизводимых и интерпретируемых эталонных тестов, минимизирующих влияние субъективных факторов, необходимо изолированно тестировать отдельные навыки LLM с применением метрики Exact Match (EM) с целью оценки результатов. Предлагаемый эталонный тест MUE обеспечивает возможность комплексной оценки различных аспектов числового мышления LLM. Задача краткого изложения текста служит наглядным примером того, как возможности больших языковых моделей проявляются в прикладных задачах.

6. Результаты экспериментов. Для проведения экспериментов были выбраны следующие модели: GigaChat, GPT, Grok, Gemini, LLaMA, Yandex GPT, Qwen, DeepSeek, Mistral. Версии тестируемых моделей приведены в таблице 4 с результатами экспериментов.

Таблица 4. Результаты экспериментов

Модели	Тип ввода числительного	Тесты					
		MUE-1	MUE-2	MUE-3	MUE-4	MUE-5	AVG
GigaChat-2	arabic_num	0,95	0,05	0,15	0,55	0,75	0,49
	text_en	0,95	0,25	0,30	0,65	0,70	0,57
	roman_num	0,70	0,15	0,05	0,40	0,55	0,37
	text_ru	0,95	0,15	0,15	0,60	0,70	0,51
	text_ru_en	0,90	0,15	0,00	0,65	0,60	0,46
	total	0,89	0,15	0,13	0,57	0,66	0,48
GigaChat-2-Max	arabic_num	1,00	0,30	0,45	0,75	0,90	0,68
	text_en	1,00	0,40	0,55	0,80	0,85	0,72
	roman_num	0,80	0,40	0,35	0,65	0,55	0,55
	text_ru	1,00	0,35	0,55	0,85	0,85	0,72
	total	0,96	0,36	0,48	0,77	0,80	0,67
GigaChat-2-Pro	arabic_num	0,95	0,20	0,45	0,80	0,90	0,66
	text_en	0,95	0,35	0,35	0,80	0,85	0,66
	roman_num	0,80	0,20	0,20	0,65	0,50	0,47
	text_ru	0,95	0,30	0,55	0,90	0,85	0,71
	text_ru_en	0,80	0,25	0,35	0,80	0,80	0,60
	total	0,89	0,26	0,38	0,79	0,78	0,62
YandexGPT	arabic_num	0,90	0,25	0,25	0,65	0,90	0,59
	text_en	0,85	0,40	0,40	0,85	0,85	0,67
	roman_num	0,75	0,30	0,20	0,30	0,55	0,42
	text_ru	0,85	0,45	0,45	0,75	0,90	0,68
	text_ru_en	0,90	0,35	0,30	0,70	0,90	0,63
	total	0,85	0,35	0,32	0,65	0,82	0,60
yandexgpt-lite	arabic_num	0,85	0,10	0,15	0,30	0,90	0,46
	text_en	0,30	0,30	0,20	0,25	0,80	0,37
	roman_num	0,75	0,25	0,05	0,10	0,60	0,35
	text_ru	0,80	0,30	0,20	0,25	0,80	0,47
	text_ru_en	0,80	0,25	0,15	0,10	0,75	0,41
	total	0,70	0,24	0,15	0,20	0,77	0,41

Продолжение Таблицы 4

Модели	Тип ввода числительного	Тесты					
		MUE-1	MUE-2	MUE-3	MUE-4	MUE-5	AVG
gpt-4o	arabic num	1,00	0,15	0,40	0,80	1,00	0,67
	text en	1,00	0,40	0,65	0,65	0,80	0,70
	roman num	0,85	0,20	0,30	0,60	0,70	0,53
	text ru	1,00	0,40	0,60	0,85	0,85	0,74
	text ru en	1,00	0,30	0,45	0,80	0,80	0,67
	total	0,97	0,29	0,48	0,74	0,83	0,66
DeepSeek-R1-Distill-Llama-70B	arabic num	0,95	0,30	0,30	0,75	0,55	0,57
	text en	0,40	0,40	0,20	0,40	0,05	0,29
	roman num	0,60	0,50	0,20	0,85	0,35	0,50
	text ru	0,90	0,50	0,50	0,65	0,40	0,59
	text ru en	0,75	0,45	0,15	0,75	0,50	0,52
	total	0,72	0,43	0,27	0,68	0,37	0,49
Llama-3.3-70B-Instruct	arabic num	0,50	0,20	0,20	0,80	0,85	0,51
	text en	0,65	0,45	0,35	0,75	0,75	0,59
	roman num	0,90	0,30	0,20	0,55	0,50	0,49
	text ru	1,00	0,40	0,35	0,75	0,80	0,66
	text ru en	1,00	0,40	0,25	0,70	0,75	0,62
	total	0,81	0,35	0,27	0,71	0,73	0,57
RuadapTQwen2.5-32B-Pro-Beta	arabic num	0,85	0,20	0,15	0,80	0,90	0,58
	text en	0,80	0,25	0,25	0,60	0,80	0,54
	roman num	0,90	0,30	0,15	0,70	0,65	0,54
	text ru	1,00	0,25	0,40	0,85	0,85	0,67
	text ru en	1,00	0,30	0,40	0,75	0,85	0,66
	total	0,91	0,26	0,27	0,74	0,81	0,60
mistral-large	arabic num	0,75	0,20	0,40	0,75	0,85	0,59
	text en	0,75	0,20	0,35	0,70	0,60	0,52
	roman num	0,80	0,25	0,25	0,35	0,50	0,43
	text ru	1,00	0,40	0,45	0,70	0,75	0,66
	text ru en	0,95	0,40	0,30	0,65	0,75	0,61
	total	0,85	0,29	0,35	0,63	0,69	0,56
grok-3	arabic num	1,00	0,25	0,30	0,80	0,95	0,66
	text en	1,00	0,50	0,45	0,90	0,95	0,76
	roman num	0,95	0,45	0,30	0,50	0,70	0,58
	text ru	1,00	0,55	0,45	0,85	0,95	0,76
	text ru en	1,00	0,55	0,40	0,75	0,95	0,73
	total	0,99	0,46	0,38	0,76	0,90	0,70
gemini-2.5-flash	arabic num	1,00	0,25	0,40	0,85	1,00	0,70
	text en	0,95	0,50	0,55	0,95	0,95	0,78
	roman num	1,00	0,55	0,30	0,75	0,90	0,70
	text ru	1,00	0,65	0,55	0,95	0,95	0,82
	text ru en	1,00	0,65	0,45	0,95	0,95	0,80
	total	0,99	0,52	0,45	0,89	0,95	0,76

В контексте анализа сложных задач для больших языковых моделей (LLM) внимание уделяется обработке текстовых элементов и количественному пересчёту. Выявлено, что кодирование численных данных представляет собой наименее трудоёмкую задачу для большинства LLM (таблица 5).

Таблица 5. Оценка среднего значения ЕМ по типу задачи

Задача	Среднее ЕМ по всем экспериментам (робастность)
MUE-1 (Кодирование числовых данных)	0,88
MUE-2 (Анализ текстовых элементов)	0,33
MUE-3 (Количественный пересчет)	0,33
MUE-4 (Сравнение числовых значений)	0,68
MUE-5 (Интерпретация и применение формул)	0,76

Задача понимания текстовых элементов наряду с количественным пересчётом представляет наибольшую сложность. Задание вида «В ответ напиши только 2 предложения текста» оказалась сложным для многих моделей: LLM возвращали больше, либо меньше текста, чем требовалось, что свидетельствует о поверхностном «понимании» сути задания. Выполнение такого задания не представляет сложности для внимательного человека. В задаче количественного пересчета требуется владеть методами счёта, например, задание «Напиши в одну строку шесть раз слово «яблоко» и в два раза меньше слово «апельсин» содержит очевидную для человека ловушку, но для многих моделей оказывается сложным. Применимо к прикладным задачам следует использовать наиболее простую максимально прямолинейную одноязычную формулировку заданий, избегать сложных конструкций и зависимостей, для числительных использовать либо текстовое написание, либо арабские цифры без сокращений.

Исследование продемонстрировало, что эффективность LLM существенно возрастает при обработке заданий, полностью сформулированных на русском языке. Для повышения стабильности ответов LLM в прикладных задачах целесообразно использовать текстовое написание числительных. При этом инструкции, содержащие римские числа, значительно снижают точность и корректность ответов моделей (таблица 6).

Согласно усреднённой оценке, полученной на данном эталонном тесте, модель семейства Gemini продемонстрировала наивысшую производительность, заняв лидирующую позицию. Второе место в рейтинге эффективности заняла модель семейства Grok, тогда

как модели GigaChat расположились на третьей позиции (таблица 7). Для выявления лидера в прикладных задачах следует проводить тесты разных LLM.

Таблица 6. Оценка среднего значения ЕМ по типу ввода числового значения

Тип ввода числового значения	Среднее ЕМ по всем экспериментам (робастность)
arabic_num	0,60
En	0,60
roman_num	0,49
Ru	0,67
ru_en	0,62

Таблица 7. Оценка среднего значения ЕМ по LLM модели

LLM модель	Среднее ЕМ по всем экспериментам (робастность)
GigaChat-2	0,48
GigaChat-2-Max	0,67
GigaChat-2-Pro	0,62
Yandexgpt	0,60
yandexgpt-lite	0,41
gpt-4o	0,66
DeepSeek-R1-Distill-Llama-70B	0,49
Llama-3.3-70B-Instruct	0,57
RuadaptQwen2.5-32B-Pro-Beta	0,60
mistral-large	0,56
grok-3	0,70
gemini-2.5-flash	0,76

Эталонный тест MUE оценивает пять навыков числового мышления: кодирование числовых данных, анализ текстовых элементов, количественный пересчёт, сравнение числовых значений, а также интерпретацию и применение формул. Экспериментальные результаты выявили ограничения LLM при решении задач, которые для человека являются простыми. Установлено, что на корректность работы модели оказывает влияние способ представления числительных в задании. Основные причины наблюдаемых ошибок связаны, во-первых, с особенностями токенизаторов, формирующих различные векторные представления для разных способов записи чисел, и, во-вторых, с архитектурными ограничениями LLM, затрудняющими решение численных задач. Для повышения точности в прикладных сценариях рекомендуется формулировать инструкции

максимально ясно, избегать сложных синтаксических конструкций, сокращений и противоречивых условий. Стабилизация качества ответов может быть достигнута путём эмпирического подбора наиболее устойчивых формулировок промпта из числа семантически эквивалентных вариантов. Выполнение арифметических операций целесообразно возлагать на специализированные функции, разработанные для решения соответствующих задач.

Заключение. В работе предложена методология построения эталонных тестов для оценки работы больших языковых моделей (LLM) с акцентом на их способность интерпретировать и обрабатывать числовую информацию. В качестве примера подготовлен эталонный тест MUE (Math Understanding Evaluation), включающий наборы данных для изолированной проверки компонентов числового мышления LLM. Экспериментальные исследования, выполненные с использованием MUE, выявили влияние способа записи числительных на результаты, а также подтвердили применимость методологии для объективного сравнения LLM и оценки их пригодности в прикладных сценариях, требующих высокой точности работы с числами.

Результаты исследования подчёркивают необходимость регулярного автоматизированного тестирования LLM, интегрируемых в информационные системы, для обеспечения их надёжности, безопасности и предсказуемости. Приоритетными направлениями развития являются создание токенизаторов, учитывающих специфику числовых данных, формирование целей обучения, ориентированных на численный анализ, и разработка специализированных эталонных тестов для различных прикладных и исследовательских задач. Сравнение результатов в разные периоды позволит отслеживать динамику и фиксировать прогресс в развитии навыков числового мышления у LLM.

Авторы посвящают настоящую статью памяти доктора технических наук, члена-корреспондента Российской академии наук, заслуженного деятеля науки и техники Российской Федерации Юсупова Рафаэля Мидхатовича. Его выдающиеся результаты внесли значительный вклад в развитие отечественной научной школы и формирование современного научного мышления во многих областях информатики, информационных технологий и теории управления. Научное наследие Юсупова Р.М. продолжает служить основой для дальнейших исследований и остаётся важной частью достижений отечественной и мировой науки и техники.

Литература

1. Radford A., Narasimhan K., Salimans T., Sutskever I. Improving language understanding by generative pre-training. 2018.
2. Team G., Anil R., Borgeaud S., Alayrac J.B., Yu J., Soricut R., Blanco L, et al. Gemini: a family of highly capable multimodal models. 2023. arXiv preprint arXiv:2312.11805.
3. Jiang A.Q., Sablayrolles A., Mensch A., Bamford C., Chaplot D.S., Casas D.D., Sayed W.E., et al. Mistral 7B. 2023. arXiv preprint arXiv:2310.06825.
4. Touvron H., Lavril T., Izacard G., Martinet X., Lachaux M.A., Lacroix T., Rozière B., Goyal N., Hambro E., Azhar F., Rodriguez A., Joulin A., Grave E., Lample G. Llama: Open and efficient foundation language models. 2023. arXiv preprint arXiv:2302.13971.
5. Дагаев А.Е., Попов Д.И. Сравнение автоматического обобщения текстов на русском языке. Программные системы и вычислительные методы. 2024. Т. 4. С. 13–22. DOI: 10.7256/2454-0714.2024.4.69474.
6. Tsanda A., Bruches E. Russian-Language Multimodal Dataset for Automatic Summarization of Scientific Papers. 2024. arXiv preprint arXiv:2405.07886.
7. Liu A., Feng B., Wang B., Wang B., Liu B., Zhao C., Dengr C., Ruan C., Dai D., Guo D., Yang D., et al. Deepseek-v2: A strong, economical, and efficient mixture-of-experts language model. 2024. arXiv preprint arXiv:2405.04434.
8. Fan J., Martinson S., Wang E.Y., Hausknecht K., Brenner J., Liu D., Peng N., Wang C., Brenner M.P. HARDMath: A Benchmark Dataset for Challenging Problems in Applied Mathematics. 2024. arXiv preprint arXiv:2410.09988.
9. Saxton D., Grefenstette E., Hill F., Kohli P. Analysing mathematical reasoning abilities of neural models. 2019. arXiv preprint arXiv:1904.01557.
10. Hendrycks D., Burns C., Kadavath S., Arora A., Basart S., Tang E., Song D., Steinhardt J. Measuring mathematical problem solving with the math dataset. 2021. arXiv preprint arXiv:2103.03874.
11. Lu P., Bansal H., Xia T., Liu J., Li C., Hajishirzi H., Cheng H., Chang K.W., Galley M., Gao J. Mathvista: Evaluating mathematical reasoning of foundation models in visual contexts. 2023. arXiv preprint arXiv:2310.02255.
12. Glazer E., Erdil E., Besiroglu T., Chicharro D., Chen E., Gunning A., Olsson C.F., Denain J.S., Ho A., Santos E.D., Järvinen O., et al. Frontiermath: A benchmark for evaluating advanced mathematical reasoning in ai. 2024. arXiv preprint arXiv:2411.04872.
13. Li H., Chen X., Xu Z., Li D., Hu N., Teng F., Li Y., Qiu L., Zhang C.J., Li Q., Chen L. Exposing numeracy gaps: A benchmark to evaluate fundamental numerical abilities in large language models. 2025. arXiv preprint arXiv:2502.11075.
14. Yang H., Hu Y., Kang S., Lin Z., Zhang M. Number cookbook: Number understanding of language models and how to improve it (2024). arXiv preprint arXiv:2411.03766.
15. Rahman R. Large Language Models in Numberland: A Quick Test of Their Numerical Reasoning Abilities. 2025. arXiv preprint arXiv:2504.00226.
16. Sennrich R., Haddow B., Birch A. Neural machine translation of rare words with subword units. 2015. arXiv preprint arXiv:1508.07909.
17. Schuster M., Nakajima K. Japanese and korean voice search. IEEE international conference on acoustics, speech and signal processing (ICASSP). 2012. pp. 5149–5152.
18. Kudo T., Richardson J. Sentencepiece: A simple and language independent subword tokenizer and detokenizer for neural text processing. 2018. arXiv preprint arXiv:1808.06226.

19. Myrzakhan A., Bsharat S.M., Shen Z. Open-LLM-Leaderboard: From Multi-choice to Open-style Questions for LLMs Evaluation, Benchmark, and Arena. 2024. arXiv preprint arXiv:2406.07545.
20. Zhou J., Lu T., Mishra S., Brahma S., Basu S., Luan Y., Zhou D., Hou L. Instruction-following evaluation for large language models. 2023. arXiv preprint arXiv:2311.07911.
21. Suzgun M., Scales N., Schärli N., Gehrmann S., Tay Y., Chung H.W., Chowdhery A., Le Q.V., Chi E.H., Zhou D., Wei J. Challenging big-bench tasks and whether chain-of-thought can solve them. 2022. arXiv preprint arXiv:2210.09261.
22. Rein D., Hou B.L., Stickland A.C., Petty J., Pang R.Y., Dirani J., Michael J., Bowman S.R. Gpqa: A graduate-level google-proof q&a benchmark. 2023. arXiv preprint arXiv:2311.12022.
23. Sprague Z., Ye X., Bostrom K., Chaudhuri S., Durrett G. Musr: Testing the limits of chain-of-thought with multistep soft reasoning. 2023. arXiv preprint arXiv:2310.16049.
24. Hendrycks D., Burns C., Basart S., Zou A., Mazeika M., Song D., Steinhardt J. Measuring massive multitask language understanding. 2020. arXiv preprint arXiv:2009.03300.
25. Fenogenova A., Chervyakov A., Martynov N., Kozlova A., Tikhonova M., Akhmetgareeva A., Emelyanov A., Shevelev D., Lebedev P., Sinev L., Isaeva U., et al. Mera: A comprehensive LLM evaluation in Russian. 2024. arXiv preprint arXiv:2401.04531.
26. Wang A., Pruksachatkun Y., Nangia N., Singh A., Michael J., Hill F., Levy O., Bowman S. Superglue: A stickier benchmark for general-purpose language understanding systems. *Advances in neural information processing systems*. 2019. vol. 32.
27. Shavrina T., Fenogenova A., Emelyanov A., Shevelev D., Artemova E., Malykh V., Mikhailov V., Tikhonova M., Chertok A., Evlampiev A. RussianSuperGLUE: A Russian language understanding evaluation benchmark. 2020. arXiv preprint arXiv:2010.15925.
28. Ribeiro M.T., Wu T., Guestrin C., Singh S. Beyond accuracy: Behavioral testing of NLP models with CheckList. 2020. arXiv preprint arXiv:2005.04118.
29. Luo H., Sun Q., Xu C., Zhao P., Lou J., Tao C., Geng X., Lin Q., Chen S., Zhang D. Wizardmath: Empowering mathematical reasoning for large language models via reinforced evol-instruct. 2023. arXiv preprint arXiv:2308.09583.
30. Zhang T., Kishore V., Wu F., Weinberger K.Q., Artzi Y. Bertscore: Evaluating text generation with bert. 2019. arXiv preprint arXiv:1904.09675.
31. Lin C.Y. Rouge: A package for automatic evaluation of summaries. *Text summarization branches out*. 2004. pp. 74–81.
32. Rajpurkar P., Zhang J., Lopyrev K., Liang P. Squad: 100,000+ questions for machine comprehension of text. 2016. arXiv preprint arXiv:1606.05250.
33. Amigó E., Giménez J., Gonzalo J., Márquez L. MT evaluation: Human-like vs. human acceptable. *Proceedings of the COLING/ACL 2006 Main Conference Poster Sessions*. 2006. pp. 17–24.
34. Celikyilmaz A., Clark E., Gao J. Evaluation of text generation: A survey. 2020. arXiv preprint arXiv:2006.14799.
35. Dong C., Li Y., Gong H., Chen M., Li J., Shen Y., Yang M. A survey of natural language generation. *ACM Computing Surveys*. 2022. vol. 55(8). pp. 1–38.
36. Chen M., Tworek J., Jun H., Yuan Q., Pinto H.P., Kaplan J., Edwards H., Burda Y., Joseph N., Brockman G., Ray A., et al. Evaluating large language models trained on code. 2021. arXiv preprint arXiv:2107.03374.

37. Zheng L., Chiang W.L., Sheng Y., Zhuang S., Wu Z., Zhuang Y., Lin Z., Li Z., Li D., Xing E., Zhang H. Judging llm-as-a-judge with mt-bench and chatbot arena. *Advances in Neural Information Processing Systems*. 2023. vol. 36. 46595–46623.
38. Chicco D., Jurman G. The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. *BMC genomics*. 2020. vol. 21. pp. 1–3.

Карпович Сергей Николаевич — канд. техн. наук, директор по информационным технологиям, Общество с ограниченной ответственностью «Рамблер ДС». Область научных интересов: тематическое моделирование, обработка текстов на естественном языке, информационный поиск, машинное обучение. Число научных публикаций — 13. cims@yandex.ru; Варшавское шоссе, 9/1, 117105, Москва, Россия; р.т.: +7(812)328-8071.

Смирнов Александр Викторович — д-р техн. наук, профессор, главный научный сотрудник, руководитель лаборатории, лаборатория интегрированных систем автоматизации, Санкт-Петербургский Федеральный исследовательский центр Российской академии наук (СПб ФИЦ РАН). Область научных интересов: системы поддержки принятия решений, интеллектуальные системы, интеллектуальное управление конфигурациями виртуальных и сетевых организаций, логистика знаний. Число научных публикаций — 450. smir@iias.spb.su; 14-я линия В.О., 39, 199178, Санкт-Петербург, Россия; р.т.: +7(812)328-8071.

Тесля Николай Николаевич — канд. техн. наук, старший научный сотрудник, лаборатория интегрированных систем автоматизации, Санкт-Петербургский Федеральный исследовательский центр Российской академии наук (СПб ФИЦ РАН). Область научных интересов: управление знаниями, онтологии, обработка текста на естественном языке, языковые модели. Число научных публикаций — 100. teslya@iias.spb.su; 14-я линия В.О., 39, 199178, Санкт-Петербург, Россия; р.т.: +7(812)328-8071.

Поддержка исследований. Работа выполнена в рамках бюджетной темы FFZF-2025-0003.

S. KARPOVICH, A. SMIRNOV, N. TESLYA
**METHODOLOGY FOR CREATING A BENCHMARK
TO EVALUATE LLM PERFORMANCE ON NUMERALS**

Karpovich S., Smirnov A., Teslya N. **Methodology for Creating a Benchmark to Evaluate LLM Performance on Numerals.**

Abstract. The article presents a methodology for designing a benchmark to assess numerical reasoning skills in Large Language Models (LLMs). In the context of LLMs, numerical reasoning is defined as a model's ability to correctly interpret, process, and utilize numerical information in text, including understanding magnitudes and relations between numbers, performing arithmetic operations, and generating numerals accurately in its outputs. The proposed methodology is based on decomposing applied tasks and enables targeted evaluation of specific facets of numerical reasoning using tasks that involve numerals. Particular attention is paid to the representation of numbers in textual prompts to LLMs, as this factor directly affects the quality of the final output. The need for rigorous assessment of LLMs' numerical reasoning stems from its critical role across a wide range of text-centric applications, including automated summarization, generation of analytical reports, extraction and interpretation of quantitative data, and conversational systems operating on financial, scientific, or technical information. Drawing on an analysis of state-of-the-art LLM evaluation approaches, core principles for constructing evaluation benchmarks are formulated with an emphasis on generality and real-world applicability. In accordance with the proposed methodology, the MUE (Math Understanding Evaluation) benchmark is introduced; it comprises five test suites, each designed to assess a distinct aspect of LLM numerical reasoning. A comparative evaluation of popular LLMs is conducted, leading models are identified, and the strengths and weaknesses of their numerical reasoning are characterized. The findings are intended to inform LLM developers in refining architectures and training strategies, and to guide end users and integrators in selecting an optimal model for applied projects.

Keywords: methodology, Large Language Models (LLM), LLM benchmark, Natural Language Processing (NLP), numerals.

References

1. Radford A., Narasimhan K., Salimans T., Sutskever I. Improving language understanding by generative pre-training. 2018.
2. Team G., Anil R., Borgeaud S., Alayrac J.B., Yu J., Soricut R., Blanco L, et al. Gemini: a family of highly capable multimodal models. 2023. arXiv preprint arXiv:2312.11805.
3. Jiang A.Q., Sablayrolles A., Mensch A., Bamford C., Chaplot D.S., Casas D.D., Sayed W.E., et al. Mistral 7B. 2023. arXiv preprint arXiv:2310.06825.
4. Touvron H., Lavril T., Izacard G., Martinet X., Lachaux M.A., Lacroix T., Rozière B., Goyal N., Hambro E., Azhar F., Rodriguez A., Joulin A., Grave E., Lample G. Llama: Open and efficient foundation language models. 2023. arXiv preprint arXiv:2302.13971.
5. Dagaev A.E., Popov D.I. [Comparison of automatic summarization of texts in Russian]. *Programmnye sistemy i vychislitel'nye metody – Software systems and computational methods*. 2024. vol. 4. pp. 13–22. DOI: 10.7256/2454-0714.2024.4.69474. (In Russ.).

6. Tsanda A., Bruches E. Russian-Language Multimodal Dataset for Automatic Summarization of Scientific Papers. 2024. arXiv preprint arXiv:2405.07886.
7. Liu A., Feng B., Wang B., Wang B., Liu B., Zhao C., Dengr C., Ruan C., Dai D., Guo D., Yang D., et al. Deepseek-v2: A strong, economical, and efficient mixture-of-experts language model. 2024. arXiv preprint arXiv:2405.04434.
8. Fan J., Martinson S., Wang E.Y., Hausknecht K., Brenner J., Liu D., Peng N., Wang C., Brenner M.P. HARDMath: A Benchmark Dataset for Challenging Problems in Applied Mathematics. 2024. arXiv preprint arXiv:2410.09988.
9. Saxton D., Grefenstette E., Hill F., Kohli P. Analysing mathematical reasoning abilities of neural models. 2019. arXiv preprint arXiv:1904.01557.
10. Hendrycks D., Burns C., Kadavath S., Arora A., Basart S., Tang E., Song D., Steinhardt J. Measuring mathematical problem solving with the math dataset. 2021. arXiv preprint arXiv:2103.03874.
11. Lu P., Bansal H., Xia T., Liu J., Li C., Hajishirzi H., Cheng H., Chang K.W., Galley M., Gao J. Mathvista: Evaluating mathematical reasoning of foundation models in visual contexts. 2023. arXiv preprint arXiv:2310.02255.
12. Glazer E., Erdil E., Besiroglu T., Chicharro D., Chen E., Gunning A., Olsson C.F., Denain J.S., Ho A., Santos E.D., Järvinen O., et al. Frontiermath: A benchmark for evaluating advanced mathematical reasoning in ai. 2024. arXiv preprint arXiv:2411.04872.
13. Li H., Chen X., Xu Z., Li D., Hu N., Teng F., Li Y., Qiu L., Zhang C.J., Li Q., Chen L. Exposing numeracy gaps: A benchmark to evaluate fundamental numerical abilities in large language models. 2025. arXiv preprint arXiv:2502.11075.
14. Yang H., Hu Y., Kang S., Lin Z., Zhang M. Number cookbook: Number understanding of language models and how to improve it (2024). arXiv preprint arXiv:2411.03766.
15. Rahman R. Large Language Models in Numberland: A Quick Test of Their Numerical Reasoning Abilities. 2025. arXiv preprint arXiv:2504.00226.
16. Sennrich R., Haddow B., Birch A. Neural machine translation of rare words with subword units. 2015. arXiv preprint arXiv:1508.07909.
17. Schuster M., Nakajima K. Japanese and korean voice search. IEEE international conference on acoustics, speech and signal processing (ICASSP). 2012. pp. 5149–5152.
18. Kudo T., Richardson J. Sentencepiece: A simple and language independent subword tokenizer and detokenizer for neural text processing. 2018. arXiv preprint arXiv:1808.06226.
19. Myrzakhan A., Bsharat S.M., Shen Z. Open-LLM-Leaderboard: From Multi-choice to Open-style Questions for LLMs Evaluation, Benchmark, and Arena. 2024. arXiv preprint arXiv:2406.07545.
20. Zhou J., Lu T., Mishra S., Brahma S., Basu S., Luan Y., Zhou D., Hou L. Instruction-following evaluation for large language models. 2023. arXiv preprint arXiv:2311.07911.
21. Suzgun M., Scales N., Schärli N., Gehrmann S., Tay Y., Chung H.W., Chowdhery A., Le Q.V., Chi E.H., Zhou D., Wei J. Challenging big-bench tasks and whether chain-of-thought can solve them. 2022. arXiv preprint arXiv:2210.09261.
22. Rein D., Hou B.L., Stickland A.C., Petty J., Pang R.Y., Dirani J., Michael J., Bowman S.R. Gpqa: A graduate-level google-proof q&a benchmark. 2023. arXiv preprint arXiv:2311.12022.
23. Sprague Z., Ye X., Bostrom K., Chaudhuri S., Durrett G. Musr: Testing the limits of chain-of-thought with multistep soft reasoning. 2023. arXiv preprint arXiv:2310.16049.

24. Hendrycks D., Burns C., Basart S., Zou A., Mazeika M., Song D., Steinhardt J. Measuring massive multitask language understanding. 2020. arXiv preprint arXiv:2009.03300.
25. Fenogenova A., Chervyakov A., Martynov N., Kozlova A., Tikhonova M., Akhmetgareeva A., Emelyanov A., Shevelev D., Lebedev P., Sinev L., Isaeva U., et al. Mera: A comprehensive LLM evaluation in Russian. 2024. arXiv preprint arXiv:2401.04531.
26. Wang A., Pruksachatkun Y., Nangia N., Singh A., Michael J., Hill F., Levy O., Bowman S. Superglue: A stickier benchmark for general-purpose language understanding systems. *Advances in neural information processing systems*. 2019. vol. 32.
27. Shavrina T., Fenogenova A., Emelyanov A., Shevelev D., Artemova E., Malykh V., Mikhailov V., Tikhonova M., Chertok A., Evlampiev A. RussianSuperGLUE: A Russian language understanding evaluation benchmark. 2020. arXiv preprint arXiv:2010.15925.
28. Ribeiro M.T., Wu T., Guestrin C., Singh S. Beyond accuracy: Behavioral testing of NLP models with CheckList. 2020. arXiv preprint arXiv:2005.04118.
29. Luo H., Sun Q., Xu C., Zhao P., Lou J., Tao C., Geng X., Lin Q., Chen S., Zhang D. Wizardmath: Empowering mathematical reasoning for large language models via reinforced evol-instruct. 2023. arXiv preprint arXiv:2308.09583.
30. Zhang T., Kishore V., Wu F., Weinberger K.Q., Artzi Y. Bertscore: Evaluating text generation with bert. 2019. arXiv preprint arXiv:1904.09675.
31. Lin C.Y. Rouge: A package for automatic evaluation of summaries. *Text summarization branches out*. 2004. pp. 74–81.
32. Rajpurkar P., Zhang J., Lopyrev K., Liang P. Squad: 100,000+ questions for machine comprehension of text. 2016. arXiv preprint arXiv:1606.05250.
33. Amigó E., Giménez J., Gonzalo J., Márquez L. MT evaluation: Human-like vs. human acceptable. *Proceedings of the COLING/ACL 2006 Main Conference Poster Sessions*. 2006. pp. 17–24.
34. Celikyilmaz A., Clark E., Gao J. Evaluation of text generation: A survey. 2020. arXiv preprint arXiv:2006.14799.
35. Dong C., Li Y., Gong H., Chen M., Li J., Shen Y., Yang M. A survey of natural language generation. *ACM Computing Surveys*. 2022. vol. 55(8). pp. 1–38.
36. Chen M., Tworek J., Jun H., Yuan Q., Pinto H.P., Kaplan J., Edwards H., Burda Y., Joseph N., Brockman G., Ray A., et al. Evaluating large language models trained on code. 2021. arXiv preprint arXiv:2107.03374.
37. Zheng L., Chiang W.L., Sheng Y., Zhuang S., Wu Z., Zhuang Y., Lin Z., Li Z., Li D., Xing E., Zhang H. Judging llm-as-a-judge with mt-bench and chatbot arena. *Advances in Neural Information Processing Systems*. 2023. vol. 36. 46595–46623.
38. Chicco D., Jurman G. The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. *BMC genomics*. 2020. vol. 21. pp. 1–3.

Karpovich Sergey — Ph.D., Chief Information Officer, LLC “Rambler DC”. Research interests: topic modeling, natural language processing, information retrieval, machine learning. The number of publications — 13. cims@yandex.ru; 9/1, Varshavskoe Hwy, 117105, Moscow, Russia; office phone: +7(812)328-8071.

Smirnov Alexander — Ph.D., Dr.Sci., Professor, Chief researcher, head of the laboratory, Computer-aided integrated systems laboratory, St. Petersburg Federal Research Center of the Russian Academy of Sciences (SPC RAS). Research interests: intelligent configuration management of virtual and networked organizations, knowledge logistics, decision support.

The number of publications — 450. smir@iias.spb.su; 39, 14-th Line V.O., 199178, St. Petersburg, Russia; office phone: +7(812)328-8071.

Teslya Nikolay — Ph.D., Senior researcher, Computer-aided integrated systems laboratory, St. Petersburg Federal Research Center of the Russian Academy of Sciences (SPC RAS). Research interests: ontologies, knowledge bases, knowledge management, natural language processing. The number of publications — 100. teslya@iias.spb.su; 39, 14-th Line V.O., 199178, St. Petersburg, Russia; office phone: +7(812)328-8071.

Acknowledgements. This research is funded by the state research FFZF-2025-0003.

I. KOTENKO, I. SAENKO, O. LAUTA, V. SADOVNIKOV, E. ICHETOVKIN, W. LI
**ANALYSIS OF MODERN RESEARCH ON PROTECTION
AGAINST ADVERSARIAL ATTACKS IN ENERGY SYSTEMS**

Kotenko I., Saenko I., Laut O., Sadovnikov V., Ichetovkin E., Li W. Analysis of Modern Research on Protection against Adversarial Attacks in Energy Systems.

Abstract. Machine learning-based systems, or machine learning systems, are currently attractive targets for attackers, since disruption of such systems can cause crucial consequences for critical infrastructure, in particular, energy systems. Therefore, the number of different types of cyber attacks against machine learning systems, which are called adversarial attacks, is continuously increasing, and these attacks are the subject of study for many researchers. Accordingly, many publications devoted to reviews of adversarial attacks and defense methods against them appear every year. Many types of adversarial attacks and defense methods in these review articles overlap. However, more recent studies contain information about new types of attacks and defense methods. The purpose of this article is to analyze the research conducted over the past six years in highly ranked journals, with an emphasis on review papers. The result of the study is a refined classification of adversarial attacks, characteristics of the most common attacks, as well as a refined classification and characteristics of defense methods against these attacks. The analysis focuses on adversarial attacks that target energy systems. The article concludes with a discussion of the advantages and disadvantages of various adversarial defense methods.

Keywords: cyber attacks, artificial intelligence, machine learning, adversarial attacks, threat model, defense methods, overview, energy systems, classification.

1. Introduction. Artificial intelligence (AI) and machine learning (ML) are areas of computer science that are increasingly used to create decision support systems that can process and analyze information in the same way as humans do. The key idea behind the operation of ML systems is that they are able to learn on their own by analyzing large amounts of empirical data and then use the acquired knowledge to make decisions in new situations. The effectiveness of ML systems is due to their ability to find hidden patterns in data describing various processes or phenomena much faster than humans. Due to this, ML systems are becoming increasingly widespread in many areas of human activity [1]. However, it should not be forgotten that this field, despite a large number of studies, still remains insufficiently studied, and the number of papers in this area is currently growing rapidly [2]. In view of this fact, it is logical to conclude that an equally young area of informatics and computer science is information security of applied artificial intelligence (AI) models [3, 4]. AI is gradually being integrated into business structures and widely implemented in specific applications [5], a trend which undoubtedly presents new challenges for ML models developers [6].

These trends are particularly relevant to energy infrastructure, which is one of the most critical sectors in modern society [7]. The security of energy infrastructure is essential for the well-being of society. Faults in power grids can be caused by various factors, such as, for example, natural phenomena (lightning, floods), contact with the external environment (trees, birds, animals), wear and tear or aging of equipment. All of which can lead to cascading effects that may harm the economy and safety [8]. Detecting and classifying such faults with high accuracy are crucial for the safety of society. For this reason, ML-based methods and approaches are currently used for their detection and classification. ML models, especially those based on deep learning (DL), are widely used in energy system infrastructure due to the huge volumes of data covering energy networks [9].

Simultaneously, smart grids have become prime targets for intruders [10]. The complexity of ML models and methods introduces numerous vulnerabilities which can be used to implement attacks against these models and methods (adversarial attacks). The first cases of implementation of gradient-based adversarial attacks were demonstrated in [11, 12]. These attacks can violate the confidentiality, integrity, and/or availability of smart grids (SG) [13, 14]. Adversarial attacks are implemented using adversarial examples, which are subtle but non-random perturbations designed to force the ML model to produce incorrect output, such as an incorrect classification of an input sample containing data on the state of the power grid [15]. It should be noted that adversarial attacks can be performed with low computational costs. This is especially true for black-box attacks, which do not require internal information about the operation of the ML system. At the same time, defense methods, even against black box attacks, may require different amounts of computing resources, which depends on the type and object of the attack.

Figure 1 shows a general scenario in which a hacker attacks an ML system that includes an intelligent power plant control system (in terms of ML models) and datasets containing parameters of the state of the energy infrastructure that are used for ML. A hacker performs data poisoning and parameter tampering attacks on datasets used to train ML models (neural networks), as well as evasion attacks and various types of adversarial machine learning (AML) attacks that are aimed at changing the operating parameters of ML models.

The hacker's goal is to force the ML model used in the fault classification system of the Power Plant Management System (PPMS) to incorrectly classify the input sample into an erroneous class. To do this, the hacker changes the label of the target class in the training dataset or makes

some other seemingly insignificant changes to the dataset or ML model. As a result, PPMS begins to misclassify faults, which in turn leads to erroneous identification of their causes and erroneous prediction of consequences. Thus, adversarial attacks can cause potentially catastrophic damage to energy infrastructure if they are not detected in time and adequate protective measures are not taken against these attacks [16].

The number of papers addressing adversarial attacks and defense methods against them grows considerably each year.

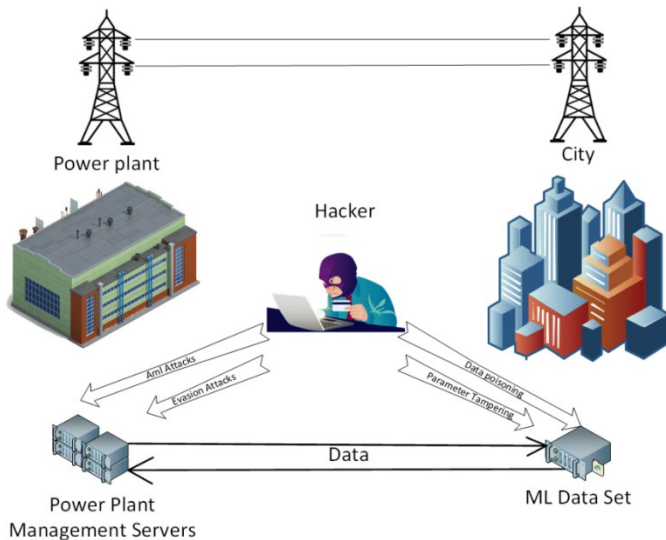


Fig. 1. Impact of adversarial attacks on ML system of energy infrastructure

At the same time, review papers appear every year that clarify the algorithms for implementing such attacks and consider new solutions to counter these attacks. The authors of this paper also proposed approaches and solutions for protecting against attacks on ML systems [17, 18], where an analysis of the existing situation in this area formed the basis of the proposed solutions.

Additionally, it is undoubtedly valuable to conduct a more in-depth analysis of known attacks and defense methods in ML systems. For this purpose, the analysis focused on review papers on adversarial attacks and methods of defense against them that have appeared in recent years. In total, over 30 review papers were selected for analysis for the period 2019-2024. In addition, the papers on adversarial attacks in the field of energy infrastructure for the period 2020-2025 were selected and processed. By

analyzing these papers, one can notice trends in the development of adversarial attacks and defense methods which are very important to know for efficient information protection of the infrastructure of energy systems.

The purpose of this paper is to consider attacks on ML systems and methods of defense against them, based on the analysis of reviews devoted to adversarial attacks as well as research papers devoted to the development and evaluation of attacks and defense methods in the field of energy infrastructure. Therefore, this analysis aims to provide a comprehensive overview of the protection of power systems against attacks on ML systems.

The novelty of the work consists in the synthesis of the latest achievements in the field of defense against attacks on ML systems within one work. At the same time, generalized classification systems for attacks and defense methods are proposed, taking into account known solutions for such classification. In addition, the article provides a description of the most popular adversarial attacks and defense methods. The consideration of the most popular defense methods is performed at a formalized level. Thus, this article serves as a consolidated resource and a practical guide for researchers in the field of information security of ML systems.

The article has the following further structure. Section 2 provides a general description of related papers that were analyzed. Section 3 is devoted to the consideration of the most popular adversarial attacks. Section 4 considers methods of defense against adversarial attacks. The discussion is held in Section 5. Section 6 contains the final conclusions.

2. Analysis of related papers. The analysis of related articles was carried out for two purposes. The first purpose was to familiarize with review papers on the topic of adversarial attacks and protection methods. These studies, as a rule, provided a classification of adversarial attacks and protection methods, and also characterized them. In this case, various areas of application for ML systems affected by adversarial attacks were considered. It should be noted that the classification of adversarial attacks and protection methods in different review papers was different. The second purpose was an attempt to generalize the articles that considered the results of implementing adversarial attacks and protection in the energy field. Here, it was of interest to identify the most popular approaches to performing adversarial attacks and protection methods that were implemented in various areas of managing and monitoring energy networks and their elements.

The selection of review papers for analysis was carried out as follows. The search engines Google.com and Scholar.google.com with the term “Adversarial attack defense review” were used. The search period was set as 2020–2025. As a rule, articles published in journals with quartiles Q1

and Q2 in the index database Scopus were subject to selection (since a large number of papers have been published on the topic, and the high quartile is an indicator of a high level of publication). The quartile was determined based on the data presented in the system <https://www.scimagojr.com>. A total of 30 review articles were selected for analysis in this way. Table 1 provides general information about the selected review articles.

Table 1. General information about the selected review articles

P_1	P_2	P_3	P_4	P_5	P_6	
Huang X. et al. [19]	2020	Q1	203	11	0.05	Deep Learning
Martins N. et al. [20]	2020	Q1	72	18	0.25	Malicious Code
Oseni A. et al. [21]	2020	Q2	139	25	0.18	Machine Learning
Xu H. et al. [22]	2020	Q2	136	12	0.09	Images, Graphics, Text
Ren K. et al. [23]	2020	Q3	115	4	0.03	Deep Learning
Zhou X. et al. [24]	2020	Q3	13	4	0.31	Electricity, Smart Grids
Akhtar N. et al. [25]	2021	Q1	450	339	0.75	Computer Vision
Zhang H. et al. [26]	2021	Q1	152	34	0.22	Electricity, Smart Grids
Chakraborty A. et al. [27]	2021	Q1	78	1	0.01	Deep Learning
Rosenberg I. et al. [28]	2021	Q2	163	38	0.23	Cybersecurity
Tian J. et al. [29]	2021	Q2	65	42	0.65	Electricity, Smart Grids
Kong Z. et al. [30]	2021	Q3	132	75	0.57	Images, Text, Malicious Code
Zhou Sh. et al. [31]	2022	Q1	185	121	0.65	Deep Learning
Khamaiseh S.Y. et al. [32]	2022	Q1	128	45	0.35	Deep Learning
Liang H. et al. [33]	2022	Q1	52	21	0.40	Deep Learning
Tian Q. et al. [34]	2022	Q1	46	26	0.57	Digital Signals
Anastasiou Th. et al. [35]	2022	Q1	34	27	0.79	Deep Learning
Li Y. et al. [36]	2022	Q2	103	48	0.47	Deep Learning
Tian J. et al. [37]	2022	Q2	49	21	0.43	Electricity, Smart Grids
Li H. et al. [38]	2022	Q2	46	7	0.15	Images
Girdhar M. et al. [39]	2023	Q1	246	140	0.57	Autonomous Vehicles
Goyal Sh. et al. [40]	2023	Q1	176	119	0.68	Text
Al-Khassawneh Y.A. [41]	2023	Q1	53	23	0.43	Cybersecurity
He K. et al. [42]	2023	Q1	166	64	0.39	Cybersecurity
Sun L. et al. [43]	2023	Q1	254	134	0.53	Autonomous Vehicles
Qureshi A.U.H. et al. [44]	2023	Q2	179	48	0.27	Graphics
Hoang V.-T. et al. [45]	2024	Q1	173	136	0.78	Deep Learning
Baniecki H. et al. [46]	2024	Q1	179	131	0.73	Deep Learning
Andrade E. de O. et al. [47]	2024	Q1	65	54	0.83	Images
Zhang Ch. Et al. [48]	2025	Q1	130	122	0.93	Images, Text

Table 1 presents the following indicators: P_1 – reference; P_2 – year of publication; P_3 – quartile according to the Scopus database; P_4 – total number of references in the article; P_5 – number of references to sources published since 2019; P_6 – ratio of the share of references to sources

published since 2019 to the total number of references (i.e., $P6 = P5 / P4$); $P7$ – subject area.

Analyzing the data on the $P3$ indicator, we can conclude that review papers on the topic of attacks on ML systems are of great importance and are of undoubted interest. The vast majority of the analyzed papers were published in highly ranked journals of the Q1 and Q2 quartiles. Only two analyzed papers of 2020, [23] and [24], belong to the Q3 quartile.

The $P4$, $P5$, and $P6$ indicators characterize the power of the reviews. In journals of the Q1 and Q2 quartiles, as a rule, the $P4$ indicator takes a value greater than 100. The $P5$ and $P6$ indicators serve to characterize the relevance of the reviews. For papers published in 2020, the $P5$ index has a small value (not exceeding 20) which is expected, since the threshold for the $P5$ assessment was set at 2019. This specific cutoff year was chosen for the purpose of establishing a consistent six-year benchmark for analysis. Publications from 2019 and later represent a six-year window of recent research, which we classify as "fresh". For 2021–2025, the $P5$ index increases, taking its highest value in 2025. The $P6$ index shows the degree of "freshness" of the entire review work. As a rule, with an increase in the year of publication of the article, the value of $P6$ also increases, especially for papers of the Q1 quartile, for which the total number of references $P4$ exceeds 100. The growth of the values of $P5$ and $P6$ with an increase in the year of publication of the article also indicates that, as a rule, in different review papers one can find references to the same original studies, which consider specific cases of the implementation of various types of adversarial attacks and defense methods.

The analysis of the $P6$ indicator allows us to conclude that most reviews have a broad subject area, which is marked as Deep Learning or Machine Learning (11 reviews). At the same time, there are reviews devoted to a specific subject area. The following subject areas are considered: Images, Graphics, Text and/or Malicious Code – 8 articles; Electricity, Smart Grids – 4; Cybersecurity – 3; Autonomous Vehicles – 2; Computer Vision – 1; Digital Signals – 1. Such a wide range of subject areas in the review papers suggests that the topic of adversarial attacks and defense against them currently affects almost all modern intelligent systems.

Each of the review papers considered the classification of attacks and defense measures, and also provided characteristics and examples of their implementation, taken from originals. It should be noted that the classification of attacks and defense measures was updated from year to year.

Information on the studies that consider the implementation of adversarial attacks and defense measures in the energy field is provided in Tables 2 and 3.

Table 2. Studies on adversarial attacks and defenses in smart grids

Studies	Description of attacks	Description of defense measures	Scope of application
[49]	Adversarial evasion and poisoning attacks	Typical countermeasures against adversarial attacks	Smart grids
[50]	The main attack surfaces and a threat analysis	-	Smart grids
[51]	Different levels of adversarial attacks	Industrial Internet of Things (IoT) architecture and a convolutional neural network with a strategy based on continuous wavelet transform	Demand side management in a smart grid
[52]	Generative Adversarial Network (GAN)	A new zero-sum game based on the Generative Adversarial Networks	Energy System Operators
[53, 54]	Evasion attacks by injecting adversarial samples	An anomaly detector combining an autoencoder, a convolutional-recurrent, and a feed-forward neural networks.	Electricity theft detectors
[55]	Evasion attacks: Distillation, No-Adversarial-Sample-Training, and False-Labeling	Federated learning (FL) and Adversarial training	Smart meters
[56]	Adversarial perturbation, backdoor injection, DoS attacks	A fine-tuned Deep Neural Network (DNN) model	Energy usage control systems
[57]	Universal adversarial attack to deceive deep learning (DL) models	Advanced DL models	Power quality disturbances
[58]	Adversarial attack injection with various reality-imitating techniques	DL monitoring method based on robust feature engineering	Phasor measurement units
[59]	Fast Gradient Sign Method (FGSM), Projected Gradient Descent (PGD), and Carlini and Wagner attacks	An artificial neural network to analyze a dataset of IEC 60870-5-104 traffic data	The communication network of SGs
[60]	DoS attacks, backdoor injections, and adversarial perturbations	Integration of ML and DL techniques	Smart grids

Continuation of the Table 2

Studies	Description of attacks	Description of defense measures	Scope of application
[61, 62]	Novel attack method called Ensemble and Transfer Adversarial Attack	DL networks	Demand Response mechanisms
[63]	False data injection	-	Smart grids
[64]	Adversarial attacks with distribution	Deep reinforcement learning with a variational auto-encoder, which constrains the deviation of malicious samples	Dynamic pricing system
[65]	Misleading the prediction of deep neural networks	Deep neural networks	Non-intrusive load monitoring
[66]	Adversarial machine learning attacks	DL models	Distributed demand response
[67]	Deep black box adversarial attacks	Quantum voting ensemble models	Smart Power Grids
[68]	Generating adversarial attacks based on linear optimization	Monitoring the observable variables	Electricity markets
[69]	Different cyber adversaries	Different detection techniques, equipment protection plans, and mitigation strategies	Electricity generation, transmission, and distribution sectors
[70]	A novel adversarial attack GAN-GRID	Security mechanisms against adversarial manipulation	The stability prediction system of a smart grid
[71]	Adversarial false data injection evasion attacks	Unsupervised spatial-temporal graph autoencoder	Data-driven detectors
[72]	Inference and evasion attacks	GAN and an ML classifier	Vehicle-to-Microgrid
[73]	Targeted adversarial false data injection strategy	DL models	Bad data detectors
[74]	Complex white-box adversarial attacks	Bayesian Neural Network framework	Smart grid fault prediction
[75, 76]	Overview of attack methods on ML-based smart grids	Overview of defense methods in ML-based smart grids	IoT-based smart grid
[77]	Different adversarial attack scenarios	Mitigation strategy for adversarial attacks based on causal theory	Measurement data protection

Table 3. Studies on adversarial attacks and defenses in energy systems in general

Studies	Description of attacks	Description of defense measures	Scope of application
[78]	Adversarial learning attack	DL models	Renewable energy forecasting
[79]	Targeted, semi-targeted, and untargeted adversarial attacks	Adversarial training	Wind power forecasting
[80]	White-box attacks and black-box attacks	DL models	Solar power generation forecasting
[81]	Targeted universal adversarial attacks	Adversarial training, defensive distillation, and feature squeezing	Power quality disturbance classification
[82]	A novel attack in Reinforcement learning (RL)	A "robustification" of RL algorithms	Prosumer pricing aggregations
[83]	FGSM, FGSM-variant, PGD, and Stochastic Serial Attacks	Multi-Source Feature Detector to detect adversarial attacks	Classification of power quality
[84]	Limited-memory Broyden-Fletcher-Goldfarb-Shanno (L-BFGS) and the Jacobian-based Saliency Map Attack (JSMA)	DL models	Power System State Estimation
[85]	Civil (or Sybil) and Fast Gradient Sign Method attacks	Random Forest, Extreme Gradient Boosting, Decision Tree, Support Vector Machine, and k-Nearest Neighbors.	Wind power forecasting
[86]	Adversarial algorithm attacks	Investing in an energy-storage system	Power system load forecasting
[87]	Synthetic minority class samples	Generative-adversarial-based semi-supervised learning framework	Diagnostics of attacks and faults in electrical networks
[88]	Various attacking objectives	Bayesian training method	Electric load forecasting
[89]	Momentum Iterative FGSM	Iterative adversarial training	Power quality disturbances classification
[90]	Various naturally perturbed data	Physics-constrained adversarial training	Power grids
[91]	Hidden FGSM attacks	ML models	Power distribution systems
[92]	False data injection attacks with adversarial examples	DL models	Bad data detectors
[93]	FGSM and Zeroth Order Optimization	Industrial IoT platform combined with an ensemble ML strategy	Power transformer

Continuation of the Table 3

Studies	Description of attacks	Description of defense measures	Scope of application
[94]	Manipulating Input Data in DNN	DL models	Base station power distribution
[95 – 97]	FGSM, Momentum Iterative FGSM, PGD attacks	DL models	Power allocation
[98]	Data poisoning techniques: scaling and random noise effects	FL models	Energy Forecasting
[99]	Different adversarial attacks	Reactive power injection	Distributed energy resources
[100]	Adversarial ML attacks	ML models	Solar photovoltaics
[101]	False data injection attacks	An unsupervised adversarial autoencoder	Power distribution grids
[102]	The massive interaction of data inside and outside the demand response management system	The adversarial multi-agent reinforcement learning framework	The demand response management system
[103]	Manipulating distributed energy resources	Mitigation strategies to thwart adversaries	Power distribution grids
[104]	Fast Gradient Sign Method	One-class Support Vector Machine and Local Outlier Factor	Photovoltaic Power Generation Forecasting
[105]	Adversarial ML algorithm powered by black-box optimization	DL models	Demand side management
[106]	An evasive deep black-box adversarial attack	ML models	Cyber–Physical Power Systems

Table 2 presents information related to smart grids. The term “Adversarial attacks defense smart grid” was used to select these studies. Table 3 presents information related to energy systems in general. In this

case, the term “Adversarial attacks defense power energy” was used. As a rule, papers published in the period 2020-2025 were subject to selection.

The characteristics of the articles are carried out according to the following criteria: (1) description of the attacks under consideration; (2) description of the protection methods under consideration; (3) the area or element in the energy system that was affected by the attack.

Analyzing the data presented in Tables 2 and 3, we can draw the following conclusions.

1. Adversarial attacks in energy infrastructures can be aimed at any elements or tasks in these infrastructures. For energy infrastructure elements, the targets of adversarial attacks can be both monitoring and state control tools (smart meters, theft detectors, quality violation detectors, bad data detectors, etc.) and elements related to trade and marketing (demand response and demand management mechanisms, pricing, electricity markets, etc.). For the tasks to be solved, the targets of adversarial attacks specified in the studies considered above are most often forecasting tasks (stability and failures of the smart grid, wind energy, solar energy, photovoltaic energy, electrical load, etc.).

The next most common task is the classification of power quality violations. This fact is explained by the fact that forecasting and classification problems in energy infrastructures are solved using various ML models (shallow, deep, federated, reinforcement), the stability of which is primarily targeted by adversarial attacks.

2. The papers considered above examine both new types of adversarial attacks (Ensemble and Transfer Adversarial Attack, GAN-GRID, etc.) and well-known ones (FGSM, GAN, PGD, Carlini & Wagner attack, False Data Injection, etc.). New types of attacks are proposed in order to draw the attention of developers and researchers to the resilience of energy infrastructure to adversarial attacks and encourage them to develop new protection measures. Known attacks were considered in order to demonstrate the capabilities of the proposed approaches to protection.

3. Among the protection methods, the most popular are measures based on the use of ML and DL models. At the same time, there are also original approaches, such as defensive distillation, feature squeezing, continuous wavelet transforms, quantum voting ensemble models, physics-constrained adversarial training and others.

In the next two sections, we consider the most common types of adversarial attacks and defense methods against them.

3. Attacks against artificial intelligence (machine learning) systems. To classify attacks against ML systems, various features are used in review papers [19 – 48], in particular, the following:

- data area under attack (images, audio, text);
- ML system architecture level where the attack is implemented (input, output, generation, gradient, features, application, authentication, database, memory);
- feature characterizing the attack (data redirection, inappropriate data input, data corruption, weak authentication, overflow);
- attack type (model extraction, poisoning, evasion, special type, digital, physical);
- attack method (attacks on parts of the ML model, direct attacks on data);
- attack target (model compromise, evade of detection, inappropriate data input);
- input data type (continuous, discrete);
- knowledge of the attacked system (“white box”, “black box”, “gray box”).

The meanings of some features (data area under attack, feature characterizing the attack, attack method, and input data type) are obvious. However, the meanings of other features require explanation.

The main types of adversarial attacks are model extraction, data poisoning, and evasion. Model extraction is an attacker's analysis of an ML system in order to extract the data on which it was trained [107]. The damage from such an attack is that the training data or the model itself may be confidential. Data poisoning attacks are one of the main security threats at the training stage of ML models. They compromise the integrity of the model by polluting the training data. As a result, the resulting model may be distorted or even unsuitable for further use [108]. Evasion attacks are a special type of adversarial attacks in which malicious data is delivered to the ML system without being detected by conventional security systems [109]. Attacks of this type are particularly widespread in Smart Power Grids, where they can generate adversarial evasion samples by altering malicious consumption data, tricking detectors into classifying them as benign [110].

ML models are increasingly used in security systems, such as Network Intrusion Detection Systems (NIDS). Therefore, attackers set goals for adversarial attacks related to compromising ML models, evading detection, or introducing unwanted data [111]. Compromising the model leads to a decrease in confidence in the results shown by the NIDS and errors in detecting network attacks. Evading detection or introducing unwanted data allows, as mentioned above, one to introduce malicious or unwanted data into the ML system, respectively.

The feature of “knowledge of the attacked system” is used in attack classifications proposed by almost all review papers. The “white box” value

means that the attacker has full knowledge of both the ML model and the training datasets. The “black box” is the opposite case, when the attacker knows nothing about either the ML model or the datasets. An intermediate option is the “gray box”, when the attacker has partial, incomplete knowledge of the ML model and data sets.

Tables 4, 5, and 6 present the distribution of the most well-known adversarial attacks by various classification features, taking into account whether the attack belongs to the “black”, “white”, or “gray” box, respectively. The full name of the attack, and its abbreviation, a brief description of the attack, the parameters used by the attack, and the values of the classification features of the attack are given.

The last column shows the S indicator, which characterizes the popularity of the attack. The S value for each attack is calculated using the following formula:

$$S = \sum_{i=1}^n A_i, \quad (1)$$

where A_i is a parameter that shows how many times the value of the i -th feature, that the attack has, occurs in Tables 4, 5, and 6.

For example, for the BA attack, the value of A_1 (feature of “Data”) is 18, A_2 (“Level”) is 4, A_3 (“Feature”) is 3, A_4 (“Type”) is 3, A_5 (“Method”) is 12, A_6 (“Target”) is 19, A_7 (“Input type”) is 18. Therefore, for the BA attack, the value of the S indicator is 77.

Note that since the total number of attacks presented in Tables 4, 5, and 6 is 31, and the number of classification features is 7, the maximum possible value of S is 217. However, this value is not achievable in practice. The maximum value of S is the “black box” attack OPA-98. The minimum value of $S = 40$ is the “gray box” attack BOA.

Thus, each attack has an S indicator, which shows how often the classification features corresponding to this attack are encountered. Consequently, it is possible to distribute the list of attacks presented in the tables into groups, including attacks with the most popular features, where the S value is high (greater than 80), attacks with specific features, where S is less than 70, and attacks with intermediate values. In this way, we can identify groups of attacks, in each of which the set of features will be specific or more general. This can help in designing methods of protection against such attacks, since by identifying a group by features, we can design protection against the entire group at once. And if the S indicator is small, then an individual approach to protection is needed for such an attack.

Table 4. Distribution of “black box” attacks by classification features

Name	Description	Parameters	Data	Level	Feature	Type	Method	Target	Input type	S
Boundary Attack (BA)	Finding class boundaries by changing model output values slightly	Step of changing the output value and the number of iterations	Image	Output	Redirection	Extraction	Direct attacks on data	Model compromise	Continuous	77
Zero-Query Attacks (ZQA)	Minimal input data changes without access to the model	Number of model queries and size of changes	Image	Output	Redirection	Extraction	Direct attacks on data	Model compromise	Continuous	77
Generative Adversarial Network (GAN)	Generating new examples using GAN to create attack samples	GAN architecture and number of training epochs	Image	Generation	Inappropriate data input	Special	Attacks on parts of the ML model	Inappropriate data input	Continuous	69
One Pixel Attack (OPA)	Changing a single pixel in an image to achieve a significant impact on the model's output	Pixel selection and change value	Image	Input	Data corruption	Digital	Direct attacks on data	Model compromise	Continuous	98
Zeroth Order Optimization (ZOO)	Estimating the loss function without computing gradients using stochastic methods	Step size and number of loss function evaluations	Image	Output	Redirection	Extraction	Attacks on parts of the ML model	Model compromise	Continuous	73
Natural Evolution Strategies (NES)	Evolutionary algorithms for finding optimal changes to input data	Population size and number of generations	Image	Generation	Inappropriate data input	Special	Attacks on parts of the ML model	Inappropriate data input	Continuous	69
Genetic Algorithms (GA)	Genetic operations for creating new audio samples to introduce unwanted changes	Population size, mutation and crossover probability	Audio	Generation	Inappropriate data input	Special	Attacks on parts of the ML model	Inappropriate data input	Discrete	50
Improved Genetic Algorithm (IGA)	Improved genetic operations for optimizing audio samples	Parameters are similar to GA, but with improved selection and mutation	Audio	Generation	Inappropriate data input	Special	Attacks on parts of the ML model	Inappropriate data input	Discrete	50
Real-World Noise (RWN)	Injecting real noise into audio data to manipulate model output	Noise level and type of sound source	Audio	Input	Data corruption	Poisoning	Direct attacks on data	Model compromise	Continuous	77
Probability Weighted Word Saliency (PWWS)	Using probabilistic word weighting to generate attack texts	Word probabilities and their impact on the model	Text	Generation	Data corruption	Special	Direct attacks on data	Model compromise	Continuous	83
Greedy Search Algorithm (GSA)	A greedy search algorithm to find the weakest points in text data	Search strategy and criteria for selecting words	Text	Generation	Data corruption	Special	Direct attacks on data	Model compromise	Continuous	83
Insertion and Removal of Words (IRW)	Modifying text by adding or removing words to manipulate	Number of words inserted/removed and their context	Text	Generation	Data corruption	Special	Direct attacks on data	Model compromise	Continuous	83

Table 5. Distribution of “white box” attacks by classification features

Name	Description	Parameters	Data	Level	Feature	Type	Method	Target	Input type	S
Fast Gradient Sign Method (FGSM)	Modify pixels based on the gradient of the model loss	The parameter ϵ determines the rate of change	Image	Gradient	Data corruption	Digital	Direct attacks on data	Model compromise	Continuous	97
Iterative Gradient Sign Method (IGSM)	Iteratively modify pixels, accumulating changes for stronger attacks	The number of iterations and the step of change	Image	Gradient	Data corruption	Digital	Direct attacks on data	Model compromise	Continuous	97
Jacobian Saliency Map Attack (JSMA)	Modify pixels using the importance of gradients to select the most vulnerable pixels	Determining the importance of each pixel	Image	Gradient	Data corruption	Digital	Direct attacks on data	Model compromise	Continuous	97
Block Input Manipulation (BIM)	Modify blocks of pixels using gradients to fine-tune the changes	The block size and the number of iterations	Image	Gradient	Data corruption	Digital	Direct attacks on data	Model compromise	Continuous	97
Undetectable Perturbation (UP)	Invisible changes that are not noticeable to humans but affect the model output	Special algorithms for minimizing distortions	Image	Input	Data corruption	Evasion	Direct attacks on data	Evade of detection	Continuous	85
Feature Adversary (FA)	Changing features based on importance analysis to confuse the model without changing the input data	Definition of important features of the model	Image	Features	Data corruption	Evasion	Attacks on parts of the ML model	Evade of detection	Continuous	73
Carlini and Wagner's Attack (C&W)	Optimized changes that minimize the distance between the original and modified images	Weighting factors for different targets	Image	Gradient	Data corruption	Digital	Direct attacks on data	Model compromise	Continuous	97
Iterative Least-Likely Class Method (ILLCM)	Minimizing the probability of the correct class through iterative changes to the input data	Number of iterations and target class	Image	Gradient	Data corruption	Evasion	Direct attacks on data	Evade of detection	Continuous	86
One-Step Target Class Method (OSTCM)	Directed changes that cause the model to misclassify data into a predetermined class	Defining the target class and changing the input data	Image	Output	Data corruption	Evasion	Direct attacks on data	Evade of detection	Continuous	83
Deep Fool (DF)	Iteratively changing the input data to minimize the distance to the class boundary	Number of iterations and change step	Image	Gradient	Data corruption	Digital	Direct attacks on data	Model compromise	Continuous	97
Hot/Cold method (HCM)	Temperature changes in the input data that affect the model's classification process	Temperature and exposure time	Image	Input	Data corruption	Evasion	Direct attacks on data	Evade of detection	Continuous	85
Ground-Truth Attack (GTA)	Incorrect class labels fed to the model to manipulate its output	Type of mislabeled label and its impact on learning	Image	Input	Data corruption	Evasion	Attacks on parts of the ML model	Evade of detection	Discrete	78
Targeted Audio Adversarial Examples (TAAE)	Audio adversarial examples aimed at fooling speech recognition systems	Using psychoacoustic principles	Audio	Input	Data corruption	Evasion	Attacks on parts of the ML model	Evade of detection	Continuous	62

Table 6. Distribution of “grey box” attacks by classification features

Name	Description	Parameters	Data	Level	Feature	Type	Method	Target	Input type	S
Cross-Site Scripting (XSS)	Malicious code is introduced through vulnerabilities in web applications, allowing the attacker to execute scripts on behalf of the user	Exploits vulnerabilities in web applications; parameters may include script type and injection location	Text	Application	Inappropriate data input	Poisoning	Direct attacks on data	Model compromise	Discrete	65
Password Guessing (PG)	Password guessing through systematic testing of possible combinations or using dictionaries	Parameters include password length, hashing algorithm complexity, and dictionaries used	Text	Authentication	Weak authentication	Poisoning	Direct attacks on data	Model compromise	Discrete	60
Cross-Site Request Forgery (CSRF)	Forging requests to a web application on behalf of a user without their knowledge	Exploits vulnerabilities in user authentication; parameters may include request type and target page	Text	Application	Inappropriate data input	Poisoning	Direct attacks on data	Model compromise	Discrete	65
SQL Injection (SQLI)	Inserting malicious SQL code into database queries in order to gain access or manipulate data	Parameters may include database type and table structure; used to gain unauthorized access to data	Text	Database	Inappropriate data input	Poisoning	Direct attacks on data	Model compromise	Discrete	64
Buffer Overflow Attack (BOA)	Buffer overflow by writing more information than allocated, which can lead to arbitrary code execution	Exploits memory management vulnerabilities; parameters may include buffer size and vulnerability type	All	Memory	Overflow	Physical	Attacks on parts of the ML model	Evade of detection	Continuous	40
Weak Authentication Attack (WAA)	Exploiting weak passwords or authentication vulnerabilities to gain unauthorized access to systems	Parameters may include password complexity and authentication methods; used to bypass security systems	All	Authentication	Weak authentication	Physical	Direct attacks on data	Model compromise	Discrete	52

As a result, we obtain the following distribution of attacks by popularity groups:

- Group 1 “High Popularity” ($S > 80$) includes 14 attacks: OPA, PWWS, GSA, IRW, FGSM, IGSM, JSMA, BIM, UP, C&W, ILLCM, OSTCM, DF, HCM;
- Group 2 “Medium Popularity” ($70 < S \leq 80$) includes 9 attacks: BA, ZQA, GAN, ZOO, NES, RWN, FA, GTA, TAAE;
- Group 3 “Low Popularity” ($S \leq 70$) includes 8 attacks: GA, IGA, XSS, PG, CSRF, SQLI, BOA, WAA.

Thus, we have obtained 3 groups into which the above-mentioned adversarial attacks are divided according to the set of features. Group 1 represents attacks that can be described by features that are often used to describe attacks. This allows using a universal approach to detect them. Attacks from Group 2 contain rarer features. Probably, approaches from group 1 will be partially suitable for their detection. However, these attacks have their own specifics. But to detect attacks from Group 3, separate individual approaches are needed.

The proposed approach to classifying adversarial attacks can become the basis for further research and practical experiments.

Let us consider in more detail the most popular adversarial attacks.

FGSM is a white-box attack against neural networks that is used to fool models trained to recognize images. FGSM involves slightly altering an image so that the trained model will mistakenly identify it as a different class. This is done using the gradient descent method, which allows finding the most sensitive pixels in the image. When using FGSM, the initial image is considered as a point on the path from the original to the altered image that provides the maximum change in the rate value of the target loss function. This can be written as follows [112]:

$$Z^* = Z + \epsilon \cdot \text{sign}(\nabla_Z J(\theta, Z, W)), \quad (2)$$

where Z^* – adversarial image, Z – original image, ∇_Z – gradient of the image Z (it shows a directional change in the intensity or color in an image), W – label associated with input image, ϵ – applied noise, θ – model's parameters, and J – loss function.

The gradient of the loss is then calculated for each pixel in the image, after which all pixels with the smallest gradient magnitude are set to zero, and the rest are increased or decreased by an amount that is the sign of the gradient. The FGSM method allows one to create fake images that look almost identical to the originals, but carry altered information that can fool an ML model.

There are several variants of FGSM, which differ in how the gradient descent step size is determined. For example, FGSM can be used with a

fixed step, or it can determine the step at each point using a backtracking line search. FGSM has limitations. One of them is that the method can fool the model only to a certain extent, after which the results are no longer reliable, and the model begins to identify the altered image correctly. In addition, FGSM can only be applied to models that use gradient descent for training. It should also be noted that attack methods such as FGSM can be used not only by attackers, but also for various research tasks related to assessing the level of defense of neural networks and their behavior in various scenarios. In particular, the FGSM method can be used to develop new algorithms for protecting neural networks, allowing to increase the level of defense against such attacks.

IGSM is a variation of the white-box attack method against neural networks that extends the capabilities of a similar FGSM algorithm. It is based on repeated applications of the FGSM method with several modifications [113]:

$$Z'_0 = Z; Z'_{n+1} = \text{Clip}_{Z, \epsilon} \{Z'_n + \alpha \cdot \text{sign}(\nabla_z J(\theta, Z'_n, W))\}, \quad (3)$$

where α – step length, $\text{Clip}_{Z, \epsilon} \{A\}$ – the element-wise clipping of Z .

IGSM is an optimization algorithm that starts with an original image and continues updating it through a series of iterations using FGSM. In each iteration, the pixel values are changed in the direction of increasing the loss of the target function. Unlike FGSM, which uses only one iteration to generate fake images, IGSM repeats the attack procedure in each iteration, which gives a better effect but requires more computational resources. IGSM can be used for both targeted and non-targeted attacks.

JSMA is a white-box attack algorithm for image counterfeiting detection systems based on deep learning methods. This algorithm computes the sensitivity within the gradient of the loss function (L) for the input image (x), leading to favorable outcomes in white-box target-specific attacks [114]. Their execution flow is as follows: $\Delta L_x = \partial L / \partial x_i$. Then, the adversarial samples are calculated based on the algorithm's forward derivatives by the Jacobi matrix, that is computed by

$$\varphi(x, t)[i] = \begin{cases} 0, & \text{if } \frac{\partial L_t(x)}{\partial x_i} < 0 \text{ or } \sum_{i \neq t} \frac{\partial L_t(x)}{\partial x_i} > 0, \\ \frac{\partial L_t(x)}{\partial x_i} \left| \frac{\partial L_t(x)}{\partial x_i} \right|, & \text{otherwise,} \end{cases} \quad (4)$$

where i represent the input feature. The matrix will compute the change in positions of pixels to obtain the target t in the input image.

As a result, JSMA can identify the most “important” features (parts) of an image that influence the model's classification. JSMA starts by selecting a target model to attack. It then calculates a gradient vector for each part of the image, allowing it to identify those parts that influence the classification of the image as a fake. It then increases the influence of these parts of the image while decreasing the influence of other parts, leading to a result where the neural network classifies the image as a fake.

BIM is based on introducing changes into the input data that is fed to the trained model [115]:

$$f_{\theta} \rightarrow f_{\theta'} \Rightarrow \begin{cases} \forall x, g(f_{\theta}, x) \neq g(f_{\theta'}, x'), \\ \forall x, f(x) \approx f(x'), \end{cases} \quad (5)$$

where f_{θ} is a model with the model's parameters θ ; $g(\cdot, \cdot)$ denotes an explanation function, for which both model and data are the input, where the output domain varies between different explanation methods.

The results of an attack can lead to erroneous conclusions or incorrect system actions, which can threaten security and create risks for the business.

Examples of BIM attacks include changing the values of input parameters when training ML models. For example, if a system is trained to identify different patterns based on color, size, and shape, an attacker can trick the system by providing it with input data that contains altered values, such as changed colors or object sizes.

Another example of a BIM attack can be aimed at automated quality control systems, when an attacker sends corrupted data, creating false error signals. Such attacks can cause system failures, equipment malfunctions, or dangerous disruptions in the manufacturing process.

To prevent BIM attacks, it is necessary to include security measures in the design and configuration of AI systems, such as input data validation, data integrity control, and training models on a large amount of data. It is also worth using additional control methods, such as the use of one-time PIN codes or two-factor authentication, in case the system interacts with important data.

UP is a type of white-box attack that involves introducing minor changes to the data or parameters of a model that lead to erroneous conclusions and discredit the results [116]:

$$x \rightarrow x' \Rightarrow \begin{cases} g(f_\theta, x) \neq g(f_\theta, x'), \\ f(x) \approx f(x'). \end{cases} \quad (6)$$

The main goal of such an attack is to bypass the security system and create distorted data so that it is accepted as correct. UP attack can be used in various fields, for example, to manipulate voting results, change weather forecasts, autonomous vehicles, medical diagnostic systems, etc. It is very difficult to detect, as it creates minor changes in the data.

FA is aimed at allowing attackers to penetrate a system that uses ML and change its behavior in order to gain unauthorized access to data or perform other actions to suit their interests. FA attacks can be implemented using various methods, including introducing false data into the system to change its behavior, using black-box attacks on ML models, analyzing the behavior of the system, etc. These attacks can be quite complex. They require specialized skills and knowledge in the field of ML from the attackers [117].

One way to protect against FA attacks is to install software that can detect abnormal behavior of the ML system and block any attempts by the attacker to interfere. It is also necessary to ensure the defense of personal data, as well as control over the correct operation of the ML system. In addition, it is possible to use secure algorithms and technologies to combat built-in vulnerabilities.

BA is a typical black-box attack method based on the classification boundary. Starting from the original adversarial image, it uses binary search to find a sample point that is near the classification boundary. It performs a random walk along the boundary between two opposite regions, thereby decreasing the distance from the target image. According to this iterative method, the distance from the original image is gradually reduced. In other words, the classification result obtained by the input query of the classifier is always the category to be misclassified [118].

The reason this type of algorithm is called a "boundary attack" is that it generates adversarial examples by searching along the boundary until they converge to produce an optimal or rational solution. The results obtained by this method can satisfy the misclassification requirements of a "black box" model. The total perturbation that is increased compared to the original image depends on the performance of the algorithm.

ZQA is a black-box attack that is designed to transfer knowledge between models without access to the input data [119]. They rely on transferring knowledge between models using the outputs of the models rather than the input data. Traditionally, transferring knowledge between models requires access to the original model's input data, which can lead to

the leakage of sensitive information. However, ZQA attacks can transfer knowledge from one model to another using only the model's output data, significantly reducing the risk of information leakage.

Attackers can use ZQA attacks to perform various tasks, such as creating fake images and videos that fool AI-driven computer vision systems or attacking secure facial recognition systems using data obtained from other recognition systems. In addition, ZQA attacks can be used to identify sensitive information. For example, attackers can use them to detect keywords and phrases in documents that should not be publicly available. They can use the knowledge gained from one model to train another model that can identify this sensitive data.

At the same time, ZQA attacks can also be used for constructive purposes. For example, they can be used to transfer experience between models in the field of medicine or scientific research, allowing the model to speed up the training process and create more accurate results.

Defenses against ZQA can be implemented based on anomaly detection and supervised learning methods. These methods allow identifying unusual output data that can be associated with ZQA. Another possible direction of defense is to develop methods for detecting and preventing experience transfer between models using only output data. Additionally, defense can be strengthened by training models, which helps to cope with attacks and slow down the process of transferring experience between models.

NES is a family of numerical optimization algorithms for black-box problems. Like all other evolution strategies, they iteratively update the parameters of the search distribution, following the natural gradient towards higher expected fitness [120]. The general procedure for implementing this method is as follows. A parameterized distribution is used to generate a set of search points. At each point, the fitness function is estimated. The parameters of the distribution allow the algorithm to adaptively fix the values of the fitness function. For example, in the case of a Gaussian distribution, they include the mean and the covariance matrix. Based on the samples, the search gradient towards higher expected fitness is estimated. Then, an ascent step along the natural gradient is performed. This step is crucial, as it prevents oscillations, premature convergence, and undesirable effects arising from the given parameterization. The entire process is repeated until the stopping criterion is met.

GAN is a black-box attack method that uses neural networks to perform various malicious attacks on ML models. The principle of GAN is to train two neural networks, a generator and a discriminator, which are fed data sequentially to each other and learn from each other. In the first stage,

the generator creates fake data examples that are fed to the discriminator along with real examples from the training set [121]. The discriminator learns to distinguish real data from fake, and the generator learns to create data that is difficult to distinguish from real. In the second stage, the generator uses the acquired knowledge about the structure of the data to create malicious attacks on the ML model. These attacks can be different depending on the type of model and the problem it solves. Once the attack is generated, it can be used by an attacker to attack the ML model. Thus, neural GAN allows generating different types of attacks against ML models, which makes them more vulnerable to attacks. This can be used to test the robustness of models and find vulnerabilities in their defenses.

OPA is a black box attack and is based on ML algorithms. It exploits vulnerabilities in neural networks that identify images based on the color values of each pixel [122]. The basic principle of this attack is to change the value of just one pixel in the image so that the neural network incorrectly classifies this image. For example, when editing a photo of a cat, the OPA attack can change the value of the pixel in the place of the cat's nose so that the neural network will believe that it is actually a dog. The OPA attack uses evolutionary algorithms that allow one to determine the optimal pixels and change their values in such a way as to deceive the neural network. Using such algorithms allows one to achieve maximum attack efficiency with a minimum number of changes to the image.

4. Methods of defense against attacks on ML systems. Possible classification features for methods of defense against attacks on ML systems, indicated in review papers [19 – 48], are:

- direction of the defense method;
- implementation complexity;
- applicability level;
- protected data type;
- impact character;
- operating principle.

By direction, defense methods are divided into defensive and counterattacking. Defensive methods are aimed at protecting against specific types of adversarial attacks. Counterattacking methods are aimed at preventing adversarial attacks in general.

By implementation complexity, defense methods are divided into simple and complex. Simple methods are easy to implement and require minimal computing resources. Complex methods require additional calculations, a complex training process, or specialized implementation.

By applicability level, defense methods are divided into universal and specific. Universal methods can be applied to various types of models

and tasks. Specific methods require a specialized approach or are applied only to certain architectures of machine learning systems.

By the protected data type, there are methods that defend input data (images) and methods that protect neural network weights, i.e., defend the parameters of the machine learning model.

By the impact character, defense methods are divided into active and passive. Active methods change the data or model to increase stability. Passive methods analyze data and detect anomalies without changing the data or the model.

By operating principles, methods are divided into those that perform data (image) transformations, model training, and attack detection and filtering. Transformation methods change the input data. For example, they add noise or perform compression. Methods that affect training change the model training process. For example, these include adversarial learning and distillation. Attack detection and filtering methods analyze and filter the input data to detect attacks.

Table 7 displays the classification feature values for the most common adversarial attack protection methods.

The last column of Table 7, by analogy with Tables 4–6, presents the S popularity index of the protection method. It is calculated using a formula similar to the popularity of adversarial attacks. The S index for protection methods ranges from 45 (for the MagNet, DeepHunter, LID and PCA methods) to 66 (for the Mixup method). Therefore, we propose dividing all the protection methods into the following two groups by popularity:

- Group 1 “High Popularity” ($S > 55$) includes 8 attacks: Randomized Smoothing, JPEG Compression, Bit Depth Reduction, Total Variance Minimization, Feature Squeezing, Mixup, L2 Regularization, Median Filtering;

- Group 2 “Low Popularity” ($S \leq 55$) includes 9 attacks: Adversarial Training, Defensive Distillation, MagNet, DeepHunter, LID, PCA, Weight Clipping, Dropout, Spectral Normalization.

The methods of the first group have a wider scope of application. They allow detecting and counteracting a larger number of adversarial attacks. At the same time, the implementation of these methods is usually simple.

The methods of the second group are aimed at protecting against a smaller number of attacks. They are distinguished by increased complexity and specificity.

Let us consider in more detail the most popular adversarial attacks.

Table 7. Classification feature values for the most common adversarial attack defense methods

Defense method	Direction	Complexity	Applicability level	Protected data type	Impact character	Operating principle	S
Randomized Smoothing	Defense	Complex	Universal	Input	Active	Transformation	65
JPEG Compression	Defense	Simple	Universal	Input	Active	Transformation	62
Bit Depth Reduction	Defense	Simple	Universal	Input	Active	Transformation	62
Total Variance Minimization	Defense	Complex	Universal	Input	Active	Transformation	65
Adversarial Training	Defense	Complex	Specific	Weights	Active	Training	54
Feature Squeezing	Defense	Simple	Universal	Input	Active	Transformation	62
Defensive Distillation	Defense	Complex	Specific	Weights	Active	Training	54
Mixup	Defense	Complex	Universal	Input	Active	Training	66
MagNet	Counterattack	Complex	Universal	Input	Passive	Detection and filtering	45
DeepHunter	Counterattack	Complex	Universal	Input	Passive	Detection and filtering	45
Local Intrinsic Dimensionality (LID)	Counterattack	Complex	Universal	Input	Passive	Detection and filtering	45
Principal Component Analysis (PCA)	Counterattack	Complex	Universal	Input	Passive	Detection and filtering	45
Weight Clipping	Defense	Simple	Specific	Weights	Active	Training	51
Dropout	Defense	Simple	Specific	Weights	Active	Training	51
L2 Regularization	Defense	Simple	Universal	Weights	Active	Training	58
Spectral Normalization	Defense	Complex	Specific	Weights	Active	Training	54
Median Filtering	Defense	Simple	Universal	Input	Active	Transformation	62

The idea behind the **Randomized Smoothing** method can be expressed by the following formula [123]:

$$\hat{y} = \frac{1}{N} \sum_{i=1}^N f(x_i + \epsilon_i), \quad (7)$$

where $\epsilon_i \sim \mathcal{N}(0, \sigma^2)$ is some noise introduced into the initial data vector $\{x_i\}$, which has a normal distribution law $\mathcal{N}$ with a zero first moment and a second moment equal to σ^2 ; $f(\cdot)$ is the model training function; $\hat{y}$ is the result of logical inference on the model.

In other words, Randomized Smoothing is a method for protecting against adversarial attacks that relies on adding random noise to the input data and then averaging the model's predictions.

In the formula, $\hat{y}$ is the average prediction of the model after applying noise. Each image x_i is modified by adding random noise ϵ_i taken from a normal distribution $\mathcal{N}(0, \sigma^2)$. In this way, the model receives several noisy versions of the same image and makes predictions for each of them. Averaging these predictions helps reduce the impact of small changes that can be used to create adversarial examples. This method is effective due to its ability to reduce the sensitivity of the model to small changes, making the model more resilient to attacks.

Adding noise from a normal distribution allows the model to account for the uncertainty in the input data and make more stable predictions. The choice of the parameter σ plays a key role. Too small a value of σ may not provide sufficient protection, while too large a value of σ may result in a decrease in the accuracy of the model on normal data. Randomized Smoothing also requires additional computation, since it must make several predictions for each input image [124]. However, its versatility and ease of implementation make it quite popular for improving the resilience of models to adversarial attacks. This method is especially useful in situations where protection is required against small changes that may be difficult to detect with the human eye.

JPEG Compression is a defense method that is based on applying the JPEG compression algorithm to images before feeding them to the model input [125]. The idea of the method can be represented by the following formula:

$$I_{compressed} = \text{JPEG}(I, q), \quad (8)$$

where I is the original image; q is a parameter describing the image quality after compression; $I_{compressed}$ is the compressed image; $\text{JPEG}(\cdot)$ is the image compression algorithm.

The formula describes the process of transforming the original image I into a compressed image $I_{compressed}$ with a given compression quality q . JPEG compression removes high-frequency components of the image, making it less sensitive to small changes that can be used to create adversarial examples. This is because such changes are often in the high-frequency region of the spectrum, which is removed by compression. Thus, the model becomes less susceptible to such attacks, since it processes already “averaged” data.

The choice of the compression quality parameter q is an important aspect of this method. Too low a compression quality can lead to significant information loss, which will negatively affect the performance of the model on normal data [126].

At the same time, too high a compression quality may not provide sufficient protection against adversarial attacks. Therefore, it is important to find a balance between compression quality and model robustness. Despite its simplicity, JPEG Compression can be a very effective defense, especially if adversarial attacks exploit small changes that are easily removed by compression.

Bit Depth Reduction is a method that reduces the bit depth of an image's pixels, thereby reducing its sensitivity to small changes [127]. The idea behind the method is represented by the following formula:

$$I_{reduced} = \frac{\lfloor I \cdot 2^b \rfloor}{2^b}, \quad (9)$$

where I is the original image; $I_{reduced}$ is the transformed image; b is the bit depth.

The original image I is transformed into an image with a reduced bit depth b . The rounding operation $\lfloor I \cdot 2^b \rfloor$ replaces pixel values with the nearest integers in the new bit depth, which significantly reduces the number of possible pixel values. This makes the model less susceptible to small changes that can be used to generate adversarial examples, since they are simply lost in the limited range of values [128].

Total Variance Minimization (TVM) is a method aimed at removing small variations from an image by minimizing pixel variation [129]. The method is described by the following formula:

$$TVM = \min_x ||\nabla x||_1 + \lambda ||x - y||_2^2. \quad (10)$$

The formula describes the optimization process, where $||\nabla x||_1$ is the L1 norm of the image gradient that minimizes pixel variation, and $\lambda ||x - y||_2^2$ is a regularization term that controls the balance between minimizing variation and preserving the original image y .

TVM uses regularization to remove small changes that can be used to generate adversarial examples. This method is especially useful for images containing noise or other unwanted changes.

The regularization coefficient λ plays an important role in this method. It controls how much we strive to preserve the original image y

while minimizing variation. If λ is too large, the model will tend to preserve the original image with minimal changes, which may not provide sufficient protection against adversarial attacks [130]. On the other hand, too small a value of λ can lead to significant blurring of the image, which can also negatively affect the performance of the model. TVM requires solving an optimization problem, which can increase computational costs, but its results can significantly improve the model's resistance to various types of attacks.

Adversarial Training is a method for training a model on a combination of original and adversarial examples [131]. The method uses two neural networks: a generator that generates adversarial examples, and a discriminator that evaluates them [132]. The idea of the method is expressed by the following formula:

$$L(\theta) = E_{(x,y) \sim D}[l(f_{\theta}(x), y)] + E_{(x',y) \sim D'}[l(f_{\theta}(x'), y)]. \quad (11)$$

The formula describes a loss function that consists of two parts. The first part corresponds to the loss on the original data D , and the second part corresponds to the loss on the adversarial data D' . Training a model on adversarial examples makes it more resilient to such attacks, since it has already "seen" and processed such examples during training.

This method requires the creation of adversarial examples, which may require additional computation and time. Adversarial examples are created by making small changes to the original data so that they cause the model to make erroneous predictions. The model is then trained on these examples, which makes it more resilient to such attacks. One drawback of the method is that it can lead to some decrease in the model's performance on ordinary data. However, Adversarial Training remains one of the most effective methods for protecting against specific types of adversarial attacks.

The process of adversarial training can be broken down into several steps:

1. Creating two neural networks: a generator and a discriminator.
2. Training the generator to create fake data, such as images, sounds, texts, etc. At the same time, the discriminator learns to distinguish fake data from real data.
3. The generator and discriminator compete with each other within the framework of the task that was defined for training the model (for example, face recognition).

4. The generator creates new fake data, and the discriminator evaluates how similar it is to real data. The discriminator's assessment is passed back to the generator so that it can improve its skills.

5. The discriminator learns to become more and more accurate in recognizing fakes. The generator, in turn, learns to create better and better fakes. This process is repeated many times until the fakes become almost indistinguishable from real data.

6. The model obtained after training can be used to analyze new data. For example, it can be used to recognize faces in photographs attached to loan applications, to detect lies during a conversation with a client, etc.

Depending on the applications and areas where this method is applied, the steps may vary slightly to achieve better results. However, the general approach is to adversarially train two networks, a generator and a discriminator, to create better fake data and protect ML systems from attacks and hacks.

Feature Squeezing is a method that reduces feature diversity by applying filters or bit compression [133]. The idea of the method is reflected in the following formula:

$$z' = \lfloor z \cdot 2^b \rfloor / 2^b. \quad (12)$$

The formula describes the process of reducing the bit depth of features z . Features are image channel values or other data characteristics that are used by the model for classification. Bit depth compression reduces the number of possible feature values, making the model less sensitive to small changes that can be used to generate adversarial examples. This method is especially useful for models that are sensitive to small changes in features.

Despite its simplicity, Feature Squeezing can be a very effective defense. Reducing the bit depth of features limits the model's ability to exploit small changes to manipulate classification results. However, too aggressive compression can lead to significant information loss, which can negatively affect the model's performance on normal data. Choosing the right bit depth b is a key factor in the success of this method [134]. It is important to find a balance between reducing the bit depth and preserving the information needed for accurate classification.

Defensive Distillation is a method of training a model using distillation to improve its robustness to adversarial attacks [135]. The method is described by the following formula:

$$p_i = \frac{\exp(z_i/T)}{\sum_j \exp(z_j/T)}. \quad (13)$$

The formula describes the process of obtaining soft labels p_i from the model outputs. Here, z_i are the logits (outputs before the activation function) of the model, and T is the temperature, which controls the "softness" of the output probabilities. Increasing the temperature makes the probabilities softer, which helps the model become less sensitive to small changes in the input. This method is especially useful for models that are prone to overfitting and sensitive to small changes in the data.

Distillation involves training a model on soft labels obtained from another model that was trained on the original data. Soft labels represent probabilities of class membership. This approach helps the model learn from more general representations of the data, making it more resilient to adversarial attacks. However, choosing the right value of the temperature T can be challenging, as too high a temperature can lead to a decrease in the accuracy of the model, while too low a temperature will not provide sufficient protection [136]. Defensive Distillation requires a specialized training process and can be more difficult to implement compared to other methods. However, its results can significantly improve the resilience of the model.

Mixup is a method that creates new training data by linearly mixing two samples and their labels [137]. The idea of the method is shown in the following formula:

$$\tilde{x} = \lambda x_i + (1 - \lambda)x_j, \quad \tilde{y} = \lambda y_i + (1 - \lambda)y_j. \quad (14)$$

The formula involves mixing the input data x_i and x_j and their labels y_i and y_j with weights λ and $(1-\lambda)$. This method helps the model become more robust to various types of attacks, as it has already "seen" and processed such examples during training. Linear data mixing creates new points in the feature space, which makes the model more robust and less sensitive to small changes.

The Mixup method requires generating new data during training, which can increase the overall training time. However, using it can significantly improve the model's robustness to adversarial attacks, as it learns from more diverse data. The parameter λ plays a key role in this method: it determines the degree to which the two samples are mixed. Choosing the right distribution for λ (e.g., Beta distribution) can help achieve a better result [138]. Despite the additional computational costs,

Mixup remains a popular method for improving the robustness of models to various types of attacks.

MagNet is a method that uses autoencoders to reconstruct original images and detect adversarial examples [139]. The method is described by the following formula:

$$\min_{\theta} E_{x \sim P_{data}} [|x - g_{\theta}(f_{\theta}(x))|^2]. \quad (15)$$

The formula describes the training process of an autoencoder, where θ are the model parameters, $f_{\theta}(x)$ is the encoder that transforms the input image into a latent representation, and $g_{\theta}(f_{\theta}(x))$ is the decoder that transforms the latent representation back into an image. The goal of training is to minimize the mean squared error between the original image x and the reconstructed image $g_{\theta}(f_{\theta}(x))$. This allows the autoencoder to efficiently reconstruct the original images and detect adversarial examples.

Since adversarial examples often have noticeable differences from normal data, they can be easily detected through large reconstruction errors. The autoencoder is trained on normal data and then used to check the input images for anomalies. This method requires significant computational resources to train the autoencoder, but its effectiveness in detecting adversarial examples makes it a valuable tool for defending artificial intelligence systems [140]. In addition, MagNet can be adapted to different types of data and models, increasing its applicability.

DeepHunter is a tool for automatic detection of adversarial examples using ensembles of models and statistical methods [141]. The idea is represented by the following formula:

$$d(x) = \frac{1}{N} \sum_{i=1}^N (f_i(x) - \bar{f}(x))^2. \quad (16)$$

The formula describes the process of calculating the standard deviation of the models' predictions. Here, $f_i(x)$ is the prediction of the i -th model, and $\bar{f}(x)$ is the average prediction of all models. If the models' predictions diverge greatly, this may indicate the presence of an adversarial example. DeepHunter exploits the idea that different models can interpret the same data differently, and if most models agree on a classification, this indicates a normal image.

Using ensembles of models allows DeepHunter to achieve high accuracy in detecting adversarial examples. This method requires

significant computational resources to train multiple models and analyze their predictions, but its results can significantly improve the system's resilience to attacks. DeepHunter is especially useful in cases where high detection reliability is required, as it uses the combined efforts of multiple models to detect anomalies [142]. This method is passive, as it analyzes the data and detects anomalies without changing the data or the model.

Local Intrinsic Dimensionality (LID) is a method for estimating the local intrinsic dimensionality of data to detect anomalies [143]. The method is described by the following formula:

$$\text{LID}(x) = \lim_{r \rightarrow 0} \frac{\log N(x, r)}{\log r}. \quad (17)$$

The formula describes the process of estimating LID for a data point x . Here $N(x, r)$ is the number of points within a radius r of the point x . Local intrinsic dimensionality is estimated by analyzing the distances between neighboring points in the feature space. Adversarial examples often have higher local dimensionality than normal data, which allows this method to effectively detect them.

The LID method requires calculating the distances between all data points, which can be computationally expensive, especially for large data sets. However, its results can significantly improve the resilience of the system to adversarial attacks. LID is a passive method, as it analyzes the data and detects anomalies without changing the data or the model [144]. This method is especially useful in cases where high accuracy and reliability of adversarial example detection is required, as it uses data properties to detect anomalies.

Principal Component Analysis (PCA) is a method of principal component analysis for anomaly detection [145]. The method is described by the following formula:

$$X_{\text{PCA}} = U\Sigma V^T. \quad (18)$$

In the formula, X_{PCA} is the original data, U is the matrix of eigenvectors, Σ is the diagonal matrix of singular values, and V^T is the transpose of the matrix of eigenvectors. PCA transforms the data into a principal component space, where each component represents the direction of maximum variance. This allows for anomaly detection, since adversarial examples often reside in areas of high variance.

PCA is a powerful tool for anomaly detection, since it can detect data that deviates significantly from a normal distribution. However, it requires performing singular value decomposition on the data, which can be computationally expensive for large datasets. PCA is a passive method, since it analyzes the data and detects anomalies without changing the data or the model. This method is especially useful in cases where high accuracy and robustness in detecting adversarial examples are required.

Weight Clipping is a method that limits the values of a model's weights within a certain range [146]. The method is described by the following formula:

$$w' = \max(-c, \min(c, w)). \quad (19)$$

The formula describes the process of trimming the weights w to the limits $[-c, c]$, where c is the weight limit. This method is based on the idea that too large or too small weight values can make the model more sensitive to small changes in the data. Weight clipping helps control this sensitivity, making the model more resilient to adversarial attacks. This method is especially useful for models that are prone to overfitting and sensitive to large weight values.

Choosing the right value of the limit c is the key to the success of this method. If the limit is too large, it may not provide sufficient protection, while too small a limit can lead to a significant decrease in the model's performance on normal data. Weight clipping is a simple method that is easy to implement and apply to any model. Its advantage is that it does not require significant changes to the model architecture and can be easily integrated into an existing data processing pipeline. This method is especially useful for problems where control over the scale of the model's weights is required.

Dropout is a regularization method based on randomly switching off neurons during training [147]. The method is described by the following formula:

$$h_i = \begin{cases} 0 & \text{with probability } p, \\ \frac{x_i}{1-p} & \text{otherwise.} \end{cases} \quad (20)$$

In the formula, h_i is the output of the neuron after applying the Dropout method, p is the probability of the neuron being disconnected, and x_i is the input of the neuron. This method helps prevent overfitting and

makes the model more resilient to adversarial attacks, as it forces the model to use different paths to classify data.

The Dropout method is a simple and effective regularization method that is widely used in practice. Its implementation requires minimal changes to the model architecture and can be easily integrated into most neural networks. However, choosing the right probability value p is an important aspect of this method. If the probability is too large, it can lead to a significant decrease in the model's performance on normal data. On the other hand, too small a probability may not provide sufficient protection against overfitting and adversarial attacks. Dropout is an active method, as it modifies the model during training, making it more resilient to various types of attacks.

L2 Regularization is a regularization method based on adding a penalty for large values of model weights [148]. The idea of the method is shown in the following formula:

$$L(\theta) = \text{Loss}(\theta) + \lambda \sum_i \theta_i^2. \quad (21)$$

The formula describes a loss function that includes the main loss function $\text{Loss}(\theta)$ and a regularization term $\lambda \sum_i \theta_i^2$. The regularization term adds a penalty for large values of the weights θ , which helps control the complexity of the model and makes it more robust to adversarial attacks. This method is especially useful for models that are prone to overfitting and use too large weights to classify data.

The regularization coefficient λ plays a key role in this method. If λ is too large, it can lead to a significant decrease in the performance of the model on normal data. On the other hand, too small a value of λ may not provide sufficient protection against overfitting and adversarial attacks. L2 Regularization is an active method, as it modifies the model training process by adding a regularization term to the loss function.

Spectral Normalization is a method of normalizing the spectral radius of scales to control their scale [149]. The method is described by the following formula:

$$W' = \frac{W}{\sigma(W)}. \quad (22)$$

The formula describes the process of normalizing the W layer weights, where $\sigma(W)$ is the spectral radius, which is the maximum singular

value of the W matrix. This method normalizes the layer weights, which helps control their scale and improves the robustness of the model. Spectral Normalization is especially useful for models that are sensitive to the scale of the weights and may be vulnerable to adversarial attacks.

Calculating the spectral radius requires performing a singular value decomposition of the weight matrix, which can be computationally expensive. However, the results of this method can significantly improve the robustness of the model to various types of attacks. Spectral Normalization is an active method, as it modifies the model weights during training. This method requires a specialized approach and can be more difficult to implement compared to other methods.

Median Filtering is a method that replaces the value of each pixel with the median of the values in its neighborhood [150]. The method is described by the following formula:

$$I_{\text{filtered}} = \text{median}(I_{\text{window}}). \quad (23)$$

The formula describes the process of filtering an image. Here I_{window} is the filtering window of pixels around the current pixel, and $\text{median}(I_{\text{window}})$ is the median value of the pixels in the window.

The method helps remove small changes, such as noise or adversarial changes, making the model less susceptible to small changes. Median Filtering is a simple and effective method for removing small changes that can be used to generate adversarial examples. It does not require significant changes to the model architecture and can be easily integrated into an existing data processing pipeline. However, too large a filtering window can result in significant blurring of the image, which can also negatively affect the performance of the model. Median Filtering is an active method, as it modifies the input data to improve the robustness of the model.

5. Discussion. Summarizing the results of the conducted research, the following conclusions can be drawn.

1. Currently, there are a considerable number of papers that study adversarial attacks and protection methods against them. Moreover, the number of such articles increases over time. This suggests that the topic of protection against attacks on machine intelligence systems is highly relevant. Moreover, many papers are of a review nature. They systematize well-known and newly emerged attacks and protection methods, propose classification features and classification systems for attacks and protection methods. Therefore, numerous review articles on this topic, published between 2020 and 2025 in journals ranked in the Scopus Q1 quartile, were

selected as the initial data for the conducted research. A total of 30 such review articles were selected [19–48]. It was noted that, in general, the material devoted to the analysis of adversarial attacks and protection methods tends to be repeated across review papers from year to year. Therefore, such a seemingly small number of review papers analyzed in the conducted study seems to be quite sufficient.

2. In addition, papers in the field of adversarial attacks were selected and studied, which are not review-based but research-based and are devoted to the field of energy infrastructure. The analysis of these studies showed that, firstly, adversarial attacks can affect any elements of smart grids or energy systems in which ML models are used for monitoring and control. Secondly, new adversarial attacks on energy infrastructure are constantly appearing. Therefore, the search for defense methods remains crucial. Thirdly, most of the protection methods against adversarial attacks used in energy infrastructures are based on various ML models.

3. Analyzing the classification systems of adversarial attacks proposed in [19–48], it was noted that all of them were unique. They differed both in classification features and in the values of these features. However, in the course of the conducted study, we sought to identify the general trends inherent in the classification of adversarial attacks and to propose a generalized classification of such attacks that would cover other known classifications. As a result, in this paper, a set of eight most relevant classification features of adversarial attacks is proposed, which was tested on a set consisting of 31 of the most well-known attacks. The dominant classification feature is “knowledge of the attacked system”, which takes the values “white box”, “black box”, and “gray box”.

4. Analyzing the distribution of the selected attacks by the values of classification features, it was found that all attacks can be divided into groups with high, medium, and low popularity. To protect against attacks with high popularity, a universal approach can be used to detect them, since they are described by frequently used feature values. Attacks of medium popularity contain rarer features. Therefore, on the one hand, approaches from the previous group are partially suitable for their detection, and on the other hand, these attacks have their own specifics. Low-popularity attacks require separate individual approaches for their detection.

5. Analyzing the classification systems of protection methods against adversarial attacks proposed in [19–48], it was also noted that all of them are unique. Therefore, one of the goals of the study was to form a generalized classification of protection methods against such attacks. A set of six classification features was proposed, which was tested on a set consisting of 17 of the most well-known protection methods. As in the case

of the classification of attacks, an assessment of protection methods by their popularity was carried out, and it was proposed to distribute protection methods into two groups: high and low popularity. High-popularity methods have a wider scope of application and are effective against a wider range of adversarial attacks. At the same time, the implementation of these methods is usually simple. Low-popularity methods provide protection against a smaller number of attacks and are characterized by increased complexity.

6. The conducted study provided a brief description of adversarial attacks and protection methods against them, using, where possible, a formal description of these attacks and methods. Thus, a formal theoretical basis was laid for further research on assessing the capabilities of adversarial attacks and improving protection methods against them.

7. In conclusion, it is important to highlight some features inherent in the implementation of adversarial attacks and protection methods against them, identified in the course of the conducted study.

Thus, the One Pixel Attack becomes ineffective if two or more pooling layers are used in the basic design of the ML model. The specific type of pooling used is irrelevant, since the combination of layers of this type neutralizes the effect of pixel differences at the level of high-level features [151].

The JSMA attack has the following very important feature: it cannot function simultaneously with the ML model. Therefore, the correct organization of data flows in the pipeline for constructing an applied AI model will help to completely eliminate the negative effect of introducing JSMA as a malicious component [152]. In this case, no additional add-ons will be required to control the functioning of the main machine learning model.

It is also worth noting that defense against FA attacks is quite non-trivial: the ML model may malfunction when anomalous values appear in the data set [153]. Anomalous indicators include not only hacking attempts but also, most often, simply the rarest values. For example, an electrical voltage of 5,000 V may seem abnormal to the generalization ability of the algorithm if the model is most often operated with data of the average statistical voltage. For this reason, the implementation of defense components against FA attacks should be done with special caution, since non-standard data can be interpreted by the defense system as an attempt of unauthorized access to data.

Finally, it should be noted that the Defensive Distillation technique has its own characteristics when protecting ensemble models. Thus, if an algorithm with privileged information is used in the decision model, then

the separate functioning of the student model and the teacher model can lead to collisions during the normal operation of the base model.

It should be noted that the analysis of defense methods for ML systems in the energy domain allowed the authors to identify three most effective approaches: noise reduction, compression, and neural cleansing of input data. Preliminary experiments demonstrated the high efficiency of the approach based on a combination of these methods [154]. Optimization of the parameters of combining these methods allows us to almost completely eliminate the negative consequences of such popular attacks as FGSM, ZOO, and OPA. Further research in this direction is ongoing.

6. Conclusion. The article presents the results of a comprehensive analysis of attacks against ML systems and defense methods against them. Energy systems and networks, being an important element of the infrastructure, are increasingly dependent on digital technologies. Modern energy systems are no longer purely physical structures, such as power plants and transmission lines. They are becoming more complex and interconnected due to the introduction of digital technologies: automation, the Internet of Things, and big data. Their integration with artificial intelligence requires the introduction of more stringent security measures.

The presented classification of attack methods and techniques will allow developers to identify the most vulnerable components of such intelligent systems, as well as find a suitable defense method from those proposed in this article. The compilation of the material discussed in this article was carried out on the basis of the inclusion of the latest achievements in the field of information security of applied ML models. Therefore, the presented defense methods will be a relevant guide for developers of AI models. Separately, it is worth noting that the integration of applied AI models into applications may imply other risks to information security that were not described in this article, due to the potential vulnerabilities in the application's underlying software and hardware complex.

The study identified several trends in the direction of attacks against AI-based systems. The first is attacks using deep learning neural networks. With the increasing computing power and availability of equipment, there are more and more opportunities to carry out such attacks. They make models vulnerable to attacks with the distortion of input data. The second trend is subtle, difficult to distinguish attacks that involve making minimal changes to normal data, making their detection even more difficult. The third is compound attacks that occur in several stages. Attacks are becoming more sophisticated, using multi-layered strategies that may include preliminary training on real data before the attack.

Studying the vulnerabilities of AI systems and developing threat models is an important task for ensuring security in the context of rapid technological development and the connection of AI to energy systems, where reliable transmission and processing of real-time data plays a key role. The creation of active monitoring and anomaly detection systems can prevent malicious actions and reduce the risks associated with the use of AI technologies in energy infrastructure. It is important to focus on the need to improve knowledge in the field of artificial intelligence and cybersecurity, namely the training of IT specialists servicing energy facilities. In the context of the growing need for security measures, the development of methods and algorithms for identifying yet-unknown attacks using AI also remains important.

On November 7, 2024, Rafael Midkhatovich Yusupov, a prominent scientist in the fields of computer science, information technology, and control theory, founder and leader of scientific schools on the theoretical foundations of the informatization of society and on the sensitivity theory of complex information and control systems, Doctor of Engineering Sciences, Professor, Corresponding Member of the Russian Academy of Sciences, Honored Scientist and Engineer of the Russian Federation, Head of the Research Department at SPIIRAS, St. Petersburg Federal Research Center of the Russian Academy of Sciences, Director of SPIIRAS (1991-2018), and Editor-in-Chief of the journal "Informatics and Automation," passed away at the age of 90. Rafael Yusupov actively promoted the field of cybersecurity and supported the work of the authors of this article.

References

1. Nascimento E.d.S., Ahmed I., Oliveira E., Palheta M.P., Steinmacher I., Conte T. Understanding Development Process of Machine Learning Systems: Challenges and Solutions. Proceedings of the ACM/IEEE International Symposium on Empirical Software Engineering and Measurement (ESEM). 2019. pp. 1–6.
2. Farhat R., Mourali Y., Jemni M., Ezzedine H. An overview of Machine Learning Technologies and their use in E-learning. Proceedings of the 2020 International Multi-Conference on Organization of Knowledge and Advanced Technologies (OCTA). 2020. pp. 1–4.
3. Liu Q., Li P., Zhao W., Cai W., Yu S., Leung V.C.M. A Survey on Security Threats and Defensive Techniques of Machine Learning: A Data Driven View. IEEE Access. 2018. vol. 6. pp. 12103–12117.
4. Pitropakis N., Panaousis E., Giannetsos T., Anastasiadis E., Loukas G. A taxonomy and survey of attacks against machine learning. Computer Science Review. 2019. vol. 34.
5. Tcukanova O.A., Yarskaya A.A., Torosyan A.A. Artificial Intelligence as a New Stage in the Development of Business Intelligence Systems. Proceedings of the International Conference on Quality Management, Transport and Information Security, Information Technologies (IT&QM&IS). 2022. pp. 315–318.

6. Bland J.A., Petty M.D., Whitaker T.S., Maxwell K.P., Cantrell W.A. Machine Learning Cyberattack and Defense Strategies. *Computers & Security*. 2020. vol. 92.
7. Onyeji I., Bazilian M., Bronk C. Cyber security and critical energy infrastructure. *The Electricity Journal*. 2014. vol. 27. pp. 52–60.
8. Santis E.D., Rizzi A., Sadeghian A. A cluster-based dissimilarity learning approach for localized fault classification in smart grids. *Swarm Evol. Comput.* 2018. vol. 39. pp. 267–278.
9. Umaphathy K., Dinesh Kumar T., Poojitha G., Khyathi Sri D., Pavaneeswar C., Amannah C. Machine Learning Applications for the Smart Grid. *Data Analytics for Smart Grids Applications – A Key to Smart City Development*. Intelligent Systems Reference Library. 2023. vol. 247. pp. 251–270.
10. Zhao Z., Chen G. An Overview of Cyber Security for Smart Grid. *Proceedings of the 2018 IEEE 27th International Symposium on Industrial Electronics (ISIE)*. 2018. pp. 1127–1131.
11. Biggio B., Corona I., Maiorca D., Nelson B., Smidic N., Laskov P., Giacinto G., Roli F. Evasion Attacks against Machine Learning at Test Time. *Advanced Information Systems Engineering. Lecture Notes in Computer Science*. 2013. vol. 7908. pp. 387–402.
12. Goodfellow I.J., Shlens J., Szegedy C. Explaining and Harnessing Adversarial Examples. *Proceedings of the International Conference on Learning Representations (ICLR'15)*. 2015. arXiv preprint arXiv:1412.6572.
13. Nguyen T.N., Liu B.-H., Nguyen N.P., Chou J.-T. Cyber Security of Smart Grid: Attacks and Defenses. *Proceedings of the IEEE International Conference on Communications (ICC)*. 2020. pp. 1–6.
14. Kawoosa A.I., Prashar D. A Review of Cyber Securities in Smart Grid Technology. *Proceedings of the 2nd International Conference on Computation, Automation and Knowledge Management (ICCAKM)*. 2021. pp. 151–156.
15. Madhavarapu V.P.K., Bhattacharjee S., Dasy S.K. A Generative Model for Evasion Attacks in Smart Grid. *Proceedings of the IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*. 2022. pp. 1–6.
16. Cui L., Qu Y., Gao L., Xie G., Yu S. Detecting false data attacks using machine learning techniques in smart grid: A survey. *J. Netw. Comput. Appl.* 2020. vol. 170.
17. Kotenko I., Saenko I., Lauta O., Kribel K., Vasiliev N. Attacks Against Artificial Intelligence Systems: Classification, The Threat Model and the Approach to Protection. *Proceedings of the Sixth International Scientific Conference “Intelligent Information Technologies for Industry” (IITI'22)*. 2023. pp. 293–302.
18. Kotenko I., Saenko I., Lauta O., Vasiliev N., Iatsenko D. Attacks Against Machine Learning Systems: Analysis and GAN-based Approach to Protection. *Proceedings of the Seventh International Scientific Conference «Intelligent Information Technologies for Industry» (IITI'23)*. 2023. pp. 49–59.
19. Huang X., Kroening D., Ruan W., Sharp J., Sun Y., Thamo E., Wu M., Yi X. A survey of safety and trustworthiness of deep neural networks: Verification, testing, adversarial attack and defence, and interpretability. *Computer Science Review*. 2020. vol. 37.
20. Martins N., Cruz J.M., Cruz T., Henriques Abreu P. Adversarial Machine Learning Applied to Intrusion and Malware Scenarios: A Systematic Review. *IEEE Access*. 2020. vol. 8. pp. 35403–35419.
21. Oseni A., Moustafa N., Janicke H., Liu P., Tari Z., Vasilakos A. Security and Privacy for Artificial Intelligence: Opportunities and Challenges. 2020. arXiv preprint arXiv:2102.04661.

22. Xu H., Ma Y., Liu H.C., Deb D., Liu H., Tang J.-L., Jain A.K. Adversarial Attacks and Defenses in Images, Graphs and Text: A Review. *Int. J. Autom. Comput.* 2020. vol. 17. pp. 151–178.
23. Ren K., Zheng T., Qin Zh., Liu X. Adversarial Attacks and Defenses in Deep Learning. *Engineering*. 2020. vol. 6. pp. 346–360.
24. Zhou X., Canady R., Li Y., Koutsoukos X., Gokhale A. Overcoming Stealthy Adversarial Attacks on Power Grid Load Predictions Through Dynamic Data Repair. *Dynamic Data Driven Applications Systems (DDDAS 2020). Lecture Notes in Computer Science*. 2020. vol. 12312. pp. 102–109.
25. Akhtar N., Mian A., Kardan N., Shah M. Advances in Adversarial Attacks and Defenses in Computer Vision: A Survey. *IEEE Access*. 2021. vol. 9. pp. 155161–155196.
26. Zhang H., Liu B., Wu H. Smart Grid Cyber-Physical Attack and Defense: A Review. *IEEE Access*. 2021. vol. 9. pp. 29641–29659.
27. Chakraborty A., Alam M., Dey V., Chattopadhyay A., Mukhopadhyay D. A survey on adversarial attacks and defences. *CAAI Trans. Intell. Technol.* 2021. vol. 6. pp. 25–45.
28. Rosenberg I., Shabtai A., Elovici Y., Rokach L. Adversarial Machine Learning Attacks and Defense Methods in the Cyber Security Domain. *ACM Comput. Surv.* 2021. vol. 54(5). pp. 1–36.
29. Tian J., Wang B., Li J., Konstantinou C. Adversarial attack and defense methods for neural network-based state estimation in smart grid. *IET Renewable Power Generation*. 2022. vol. 16. pp. 3507–3518.
30. Kong Z., Xue J., Wang Y., Huang L., Niu Z., Li F., Meng W. A Survey on Adversarial Attack in the Age of Artificial Intelligence. *Wireless Communications and Mobile Computing*. 2021. vol. 2021(1).
31. Zhou Sh., Liu Ch., Ye D., Zhu T., Zhou W., Yu Ph.S. Adversarial Attacks and Defenses in Deep Learning: From a Perspective of Cybersecurity. *ACM Comput. Surv.* 2022. vol. 55(8). pp. 1–39.
32. Khamaisch S.Y., Bagagem D., Al-Alaj A., Mancino M., Alomari H.W. Adversarial Deep Learning: A Survey on Adversarial Attacks and Defense Mechanisms on Image Classification. *IEEE Access*. 2022. vol. 10. pp. 102266–102291.
33. Liang H., He E., Zhao Y., Jia Z., Li H. Adversarial Attack and Defense: A Survey. *Electronics*. 2022. vol. 11.
34. Tian Q., Zhang S., Mao Sh., Lin Y. Adversarial attacks and defenses for digital communication signals identification. *Digital Communications and Networks*. 2024. vol. 10. no. 3. pp. 756–764.
35. Anastasiou Th., Karagiorgou S., Petrou P., Papamartzivanos D., Giannetsos Th., Tsigiotaki G., Keizer J. Towards Robustifying Image Classifiers against the Perils of Adversarial Attacks on Artificial Intelligence Systems. *Sensors*. 2022. vol. 22.
36. Li Y., Cheng M., Hsieh Ch.-J., Lee Th.C.M. A Review of Adversarial Attack and Defense for Classification Methods. *The American Statistician*. 2022. vol. 76(4). pp. 329–345.
37. Tian J., Wang B., Li J., Wang Z. Adversarial Attacks and Defense for CNN Based Power Quality Recognition in Smart Grid. *IEEE Transactions on Network Science and Engineering*. 2021. vol. 9(2). pp. 807–819.
38. Li H., Namiot D. A Survey of Adversarial Attacks and Defenses for Image Data on Deep Learning. *International Journal of Open Information Technologies*. 2022. vol. 10. pp. 9–16.
39. Girdhar M., Hong J., Moore J. Cybersecurity of Autonomous Vehicles: A Systematic Literature Review of Adversarial Attacks and Defense Models. *IEEE Open Journal of Vehicular Technology*. 2023. vol. 4. pp. 417–437.

40. Goyal Sh., Doddapaneni S., Khapra M.M., Ravindran B. A Survey of Adversarial Defenses and Robustness in NLP. *ACM Comput. Surv.* 2023. vol. 55.
41. Al-Khassawneh Y.A. A Review of Artificial Intelligence in Security and Privacy: Research Advances, Applications, Opportunities, and Challenges. *Indonesian Journal of Science and Technology.* 2023. vol. 8. pp. 79–96.
42. He K., Kim D.D., Asghar M.R. Adversarial Machine Learning for Network Intrusion Detection Systems: A Comprehensive Survey. *IEEE Communications Surveys & Tutorials.* 2023. vol. 25(1). pp. 538–566.
43. Sun L., Dou Y., Yang C., Zhang K., Wang J., Yu Ph.S., He L., Li B. Adversarial Attack and Defense on Graph Data: A Survey. *IEEE Transactions on Knowledge and Data Engineering.* 2023. vol. 35(8). pp. 7693–7711.
44. Qureshi A.U.H., Larijani H., Yousefi M., Adeel A., Mtetwa N. An Adversarial Approach for Intrusion Detection Systems Using Jacobian Saliency Map Attacks (JSMA) Algorithm. *Computers.* 2020. vol. 9(3).
45. Hoang V.-T., Ergu Y.A., Nguyen V.-L., Chang R.-G. Security risks and countermeasures of adversarial attacks on AI-driven applications in 6G networks: A survey. *Journal of Network and Computer Applications.* 2024. vol. 232.
46. Baniecki H., Biecek P. Adversarial attacks and defenses in explainable artificial intelligence: A survey. *Information Fusion.* 2024. vol. 107.
47. Andrade E. de O., Guérin J., Viterbo J., Sampaio I.G.B. Adversarial attacks and defenses in person search: A systematic mapping study and taxonomy. *Image and Vision Computing.* 2024. vol. 148.
48. Zhang Ch., Hu M., Li W., Wang L. Adversarial attacks and defenses on text-to-image diffusion models: A survey. *Information Fusion.* 2025. vol. 114.
49. Hao J., Tao Y. Adversarial attacks on deep learning models in smart grids. *Energy Reports.* 2022. vol. 8. pp. 123–129.
50. Sande-Rios J., Canal-Sánchez J., Manzano-Hernández C., Pastrana S. Threat Analysis and Adversarial Model for Smart Grids. *Proceeding of the IEEE European Symposium on Security and Privacy Workshops (EuroS&PW).* 2024. pp. 130–145.
51. Elsisi M., Su C.-L., Ali M.N. Design of Reliable IoT Systems With Deep Learning to Support Resilient Demand Side Management in Smart Grids Against Adversarial Attacks. *Transactions on Industry Applications.* 2024. vol. 60. pp. 2095–2106.
52. Ahmadian S., Malki H., Han Z. Cyber Attacks on Smart Energy Grids Using Generative Adversarial Networks. *Proceedings of the IEEE Global Conference on Signal and Information Processing (GlobalSIP).* 2018. pp. 942–946.
53. Takiddin A., Ismail M., Zafar U., Serpedin E. Robust Electricity Theft Detection Against Data Poisoning Attacks in Smart Grids. *IEEE Transactions on Smart Grid.* 2021. vol. 12. pp. 2675–2684.
54. Takiddin A., Ismail M., Serpedin E. Robust Data-Driven Detection of Electricity Theft Adversarial Evasion Attacks in Smart Grids. *IEEE Transactions on Smart Grid.* 2023. vol. 14. pp. 663–676.
55. Bondok A.H., Mahmoud M., Badr M.M., Fouda M.M., Abdallah M., Alsabaan M. Novel Evasion Attacks against Adversarial Training Defense for Smart Grid Federated Learning. *IEEE Access.* 2023. vol. 11. pp. 112953–112972.
56. Sampedro G.A., Ojo S., Krichen M., Alamro M.A., Mihoub A., Karovic V. Defending AI Models Against Adversarial Attacks in Smart Grids Using Deep Learning. *IEEE Access.* 2024. vol. 12. pp. 157408–157417. DOI: 10.1109/ACCESS.2024.3473531.
57. Khan S.U., Mynuddin M., Nabil M. AdaptEdge: Targeted Universal Adversarial Attacks on Time Series Data in Smart Grids. *IEEE Transactions on Smart Grid.* 2024. vol. 15. pp. 5072–5086.

58. Berghout T., Benbouzid M., Amirat Y. Towards Resilient and Secure Smart Grids against PMU Adversarial Attacks: A Deep Learning-Based Robust Data Engineering Approach. *Electronics*. 2023. vol. 12.
59. Teryak H., Albaseer A., Abdallah M., Al-Kuwari S., Qaraqe M. Double-Edged Defense: Thwarting Cyber Attacks and Adversarial Machine Learning in IEC 60870-5-104 Smart Grids. *IEEE Open Journal of the Industrial Electronics Society*. 2023. vol. 4. pp. 629–642.
60. Ness S. Adversarial Attack Detection in Smart Grids Using Deep Learning Architectures. *IEEE Access*. 2025. vol. 13. pp. 16314–16323. DOI: 10.1109/ACCESS.2024.3523409.
61. Zhang G., Sikdar B. Ensemble and Transfer Adversarial Attack on Smart Grid Demand-Response Mechanisms. *Proceedings of the IEEE International Conference on Communications, Control, and Computing Technologies for Smart Grids (SmartGridComm)*. IEEE, 2022. pp. 53–58.
62. Zhang G., Sikdar B. A Novel Adversarial FDI Attack and Defense Mechanism for Smart Grid Demand-Response Mechanisms. *IEEE Transactions on Industrial Cyber-Physical Systems*. 2024. vol. 2. pp. 380–390. DOI: 10.1109/TICPS.2024.3448380.
63. Reda H.T., Anwar A., Mahmood A. Comprehensive survey and taxonomies of false data injection attacks in smart grids: attack models, targets, and impacts. *Renewable and Sustainable Energy Reviews*. 2022. vol. 163.
64. Ren Y., Zhang H., Yang W., Li M., Zhang J., Li H. Transferable Adversarial Attack Against Deep Reinforcement Learning-Based Smart Grid Dynamic Pricing System. *IEEE Transactions on Industrial Informatics*. 2024. vol. 20. pp. 9015–9025.
65. He J., Xiang T., Wu T., Chen Z., Wang N., Guo S. Maintaining Privacy in Smart Grid: Utilizing the Adversarial Attack Paradigm to Counter Non-Intrusive Load Monitoring Models. *IEEE Internet of Things Journal*. 2024.
66. Guihai Z., Sikdar B. Adversarial Machine Learning Against False Data Injection Attack Detection for Smart Grid Demand Response. *Proceedings of the IEEE International Conference on Communications, Control, and Computing Technologies for Smart Grids (SmartGridComm)*. 2021. pp. 352–357.
67. Aurangzeb M., Wang Y., Iqbal S., Naveed A., Ahmed Z., Alenezi M., Shouran M. Enhancing cybersecurity in smart grids: Deep black box adversarial attacks and quantum voting ensemble models for blockchain privacy-preserving storage. *Energy Reports*. 2024. vol. 11. pp. 2493–2515.
68. Melendez K.A., Matamala Y. Adversarial attacks in demand-side electricity markets. *Applied Energy*. 2025. vol. 377.
69. Nguyen T., Wang S., Alhazmi M., Nazemi M., Estebarsari A., Dehghanian P. Electric Power Grid Resilience to Cyber Adversaries: State of the Art. *IEEE Access*. 2020. vol. 8. pp. 87592–87608.
70. Efatinasab E., Brighente A., Rampazzo M., Azadi N., Conti M. GAN-GRID: A Novel Generative Attack on Smart Grid Stability Prediction. *Computer Security – ESORICS 2024. Lecture Notes in Computer Science*. 2024. vol. 14982. pp. 374–393.
71. Takiddin A., Ismail M., Atat R., Serpedin E. Spatio-temporal Graph-Based Generation and Detection of Adversarial False Data Injection Evasion Attacks in Smart Grids. *IEEE Transactions on Artificial Intelligence*. 2024. vol. 5(12). pp. 6601–6616. DOI: 10.1109/TAI.2024.3464511.
72. Omara A., Kantarci B. An AI-driven solution to prevent adversarial attacks on mobile Vehicle-to-Microgrid services. *Simulation Modelling Practice and Theory*. 2024. vol. 137.
73. Tian J., Shen C., Wang B., Ren C., Xia X., Dong R., Cheng T. EVADE: Targeted Adversarial False Data Injection Attacks for State Estimation in Smart Grid. *IEEE Transactions on Sustainable Computing*. 2024.

74. Efatinasab E., Sinigaglia A., Azadi N., Susto G.A., Rampazzo M. Adversarially Robust Fault Zone Prediction in Smart Grids With Bayesian Neural Networks. *IEEE Access*. 2024. vol. 12. pp. 121169–121184.
75. Zhang Z., Liu M., Sun M., Deng R., Cheng P., Niyato D., Chow M.-Y., Chen J. Vulnerability of Machine Learning Approaches Applied in IoT-Based Smart Grid: A Review. *IEEE Internet of Things Journal*. 2024. vol. 11(11). pp. 18951–18975.
76. Sánchez G., Elbez G., Hagenmeyer V. Attacking Learning-based Models in Smart Grids: Current Challenges and New Frontiers. *Proceedings of the 15th ACM International Conference on Future and Sustainable Energy Systems (e-Energy '24)*. 2024. pp. 589–595.
77. Huang R., Li Y. Adversarial Attack Mitigation Strategy for Machine Learning-Based Network Attack Detection Model in Power System. *IEEE Transactions on Smart Grid*. 2022. vol. 14(3). pp. 2367–2376.
78. Ruan J., Wang Q., Chen S., Lyu H., Liang G., Zhao J., Dong Z.Y. On Vulnerability of Renewable Energy Forecasting: Adversarial Learning Attacks. *IEEE Transactions on Industrial Informatics*. 2023. vol. 20(3). pp. 3650–3663.
79. Heinrich R., Scholz C., Vogt S., Lehna M. Targeted adversarial attacks on wind power forecasts. *Mach Learn*. 2024. vol. 113(2). pp. 863–889.
80. Tang N., Mao S., Nelms R.M. Adversarial Attacks to Solar Power Forecast. *Proceedings of the 2021 IEEE Global Communications Conference (GLOBECOM)*. 2021. pp. 1–6.
81. Khan S.U., Mynuddin M., Adom I., Mahmoud M.N. Mitigating Targeted Universal Adversarial Attacks on Time Series Power Quality Disturbances Models. *Proceedings of the 5th IEEE International Conference on Trust, Privacy and Security in Intelligent Systems and Applications (TPS-ISA)*. 2023. pp. 91–100.
82. Gunn S., Jang D., Paradise O., Spangher L., Spanos C.J. Adversarial poisoning attacks on reinforcement learning-driven energy pricing. *Proceedings of the 9th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation (BuildSys '22)*. 2024. pp. 262–265.
83. Zhang L., Jiang C., Pang A., He Y. Super-efficient detector and defense method for adversarial attacks in power quality classification. *Applied Energy*. 2024. vol. 361.
84. Sayghe A., Zhao J., Konstantinou C. Evasion Attacks with Adversarial Deep Learning Against Power System State Estimation. *Proceedings of the IEEE Power & Energy Society General Meeting (PESGM)*. 2020. pp. 1–5.
85. Akter K., Rahman M.A., Islam R.M., Sheikh R.I., Hossain M.J. Attack-resilient framework for wind power forecasting against civil and adversarial attacks. *Electric Power Systems Research*. 2025. vol. 238.
86. Li J., Wang J., Chen L., Yu Y. Defending Against Adversarial Attacks by Energy Storage Facility. *Proceedings of the IEEE Power & Energy Society General Meeting (PESGM)*. 2022. pp. 1–5.
87. Farajzadeh-Zanjani M., Hallaji E., Razavi-Far R., Saif M., Parvania M. Adversarial Semi-Supervised Learning for Diagnosing Faults and Attacks in Power Grids. *IEEE Transactions on Smart Grid*. 2021. vol. 12(4). pp. 3468–3478.
88. Zhou Y., Ding Z., Wen Q., Wang Y. Robust Load Forecasting Towards Adversarial Attacks via Bayesian Learning. *IEEE Transactions on Power Systems*. 2023. vol. 38. pp. 1445–1459.
89. Zhang L., Jiang C., Chai Z., He Y. Adversarial attack and training for deep neural network based power quality disturbance classification. *Engineering Applications of Artificial Intelligence*. 2024. vol. 127.
90. Li W., Deka D., Wang R., Paternina M.R.A. Physics-Constrained Adversarial Training for Neural Networks in Stochastic Power Grids. *IEEE Transactions on*

- Artificial Intelligence. 2024. vol. 5(3). pp. 1121–1131. DOI: 10.1109/TAI.2023.3236377.
91. Afrin A., Ardakanian O. Adversarial Attacks on Machine Learning-Based State Estimation in Power Distribution Systems. *Proceedings of the 14th ACM International Conference on Future Energy Systems (e-Energy '23)*. 2023. pp. 446–458.
 92. Tian J., Wang B., Wang Z., Cao K., Li J., Ozay M. Joint Adversarial Example and False Data Injection Attacks for State Estimation in Power Systems. *IEEE Transactions on Cybernetics*. 2022. vol. 52(12). pp. 13699–13713.
 93. Ali M.N., Amer M., Elsisy M. Reliable IoT Paradigm with Ensemble Machine Learning for Faults Diagnosis of Power Transformers Considering Adversarial Attacks. *IEEE Transactions on Instrumentation and Measurement*. 2023. vol. 72. pp. 1–13.
 94. Kim B., Shi Y., Sagduyu Y.E., Erpek T., Ulukus S. Adversarial Attacks against Deep Learning Based Power Control in Wireless Communications. *Proceedings of the IEEE Globecom Workshops (GC Wkshps)*. 2021. pp. 1–6.
 95. Manoj B.R., Sadeghi M., Larsson E.G. Adversarial Attacks on Deep Learning Based Power Allocation in a Massive MIMO Network. *Proceedings of the ICC – IEEE International Conference on Communications*. 2021. pp. 1–6.
 96. Manoj B.R., Sadeghi M., Larsson E.G. Downlink Power Allocation in Massive MIMO via Deep Learning: Adversarial Attacks and Training. *IEEE Transactions on Cognitive Communications and Networking*. 2022. vol. 8(2). pp. 707–719.
 97. Santos P.M., Manoj B.R., Sadeghi M., Larsson E.G. Universal Adversarial Attacks on Neural Networks for Power Allocation in a Massive MIMO System. *IEEE Wireless Communications Letters*. 2022. vol. 11(1). pp. 67–71.
 98. Shabbir A., Manzoor H.U., Ahmed R.A., Halim Z. Resilience of Federated Learning Against False Data Injection Attacks in Energy Forecasting. *Proceedings of the 2024 International Conference on Green Energy, Computing and Sustainable Technology (GECOST)*. 2024. pp. 245–249.
 99. Xu S., Yu L., Lin X. Robust Low-Overhead Control of DER Reactive Power under Adversarial Attacks and Uncertainty. *Proceedings of the ICC 2024 – IEEE International Conference on Communications*. 2024. pp. 3097–3103.
 100. Kuzlu M., Sarp S., Catak F.O., Cali U., Zhao Y., Elma O., Guler O. Analysis of deceptive data attacks with adversarial machine learning for solar photovoltaic power generation forecasting. *Electr. Eng.* 2024. vol. 106(2). pp. 1815–1823.
 101. Zideh M.J., Khalghani M.R., Solanki S.K. An unsupervised adversarial autoencoder for cyber attack detection in power distribution grids. *Electric Power Systems Research*. 2024. vol. 232.
 102. Zeng L., Qiu D., Sun M. Resilience enhancement of multi-agent reinforcement learning-based demand response against adversarial attacks. *Applied Energy*. 2022. vol. 324.
 103. Zografopoulos I., Hatziaegyriou N.D., Konstantinou C. Distributed Energy Resources Cybersecurity Outlook: Vulnerabilities, Attacks, Impacts, and Mitigations. *IEEE Systems Journal*. 2023. vol. 17. pp. 6695–6709.
 104. Santana E.J., Silva R.P., Zarpelão B.B., Barbon Junior S. Detecting and Mitigating Adversarial Examples in Regression Tasks: A Photovoltaic Power Generation Forecasting Case Study. *Information*. 2021. vol. 12(10).
 105. Youssef E.-N.S., Labeau F., Kassouf M. Adversarial Dynamic Load-Altering Cyberattacks Against Peak Shaving Using Residential Electric Water Heaters. *IEEE Transactions on Smart Grid*. 2024. vol. 15(2). pp. 2073–2088.
 106. Bhattacharjee A., Bai G., Tushar W., Verma A., Mishra S., Saha T.K. DeeBBAA: A Benchmark Deep Black-Box Adversarial Attack Against Cyber–Physical Power

- Systems. *IEEE Internet of Things Journal*. 2024. vol. 11(24). pp. 40670–40688. DOI: 10.1109/JIOT.2024.3454257.
107. Genç D., Özüysal M., Tomur E. A Taxonomic Survey of Model Extraction Attacks. *Proceedings of the 2023 IEEE International Conference on Cyber Security and Resilience (CSR)*. 2023. pp. 200–205.
108. Fan J., Yan Q., Li M., Qu G., Xiao Y. A Survey on Data Poisoning Attacks and Defenses. *Proceedings of the 2022 7th IEEE International Conference on Data Science in Cyberspace (DSC)*. 2022. pp. 48–55.
109. Wang S., Ko R.K.L., Bai G., Dong N., Choi T., Zhang Y. Evasion Attack and Defense on Machine Learning Models in Cyber-Physical Systems: A Survey. *IEEE Communications Surveys & Tutorials*. 2024. vol. 26. pp. 930–966.
110. El-Toukhy A.T., Mahmoud M., Bondok A.H., Fouda M.M., Alsabaan M. Evasion Attacks in Smart Power Grids: A Deep Reinforcement Learning Approach. *Proceedings of the 2024 IEEE 21st Consumer Communications & Networking Conference (CCNC)*. 2024. pp. 708–713.
111. Ali Alatwi H., Morisset C. Threat Modeling for Machine Learning-Based Network Intrusion Detection Systems. *Proceedings of the IEEE International Conference on Big Data (Big Data)*. 2022. pp. 4226–4235.
112. Qiu S., Liu Q., Zhou S., Wu C. Review of Artificial Intelligence Adversarial Attack and Defense Technologies. *Applied sciences*. 2019. vol. 9(5).
113. Tuna O.F., Catak F.O., Eskil M.T. Exploiting epistemic uncertainty of the deep learning models to generate adversarial samples. *Multimed Tools Appl*. 2022. vol. 81. pp. 11479–11500.
114. Dubrovkin J. Evaluation of undetectable perturbations of the peak parameters estimated by the least square curve fitting of analytical signal consisting of overlapping peaks. *Chemom. and Intel. Lab. Syst*. 2016. vol. 153. pp. 9–21.
115. Lecun Y., Bottou L., Bengio Y., Haffner P. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*. 2002. vol. 86(11). pp. 2278–2324.
116. Liu X., Cheng M., Zhang H., Hsieh C.J. Towards Robust Neural Networks via Random Self-ensemble. *Computer Vision – ECCV 2018. Lecture Notes in Computer Science*. 2018. vol. 11211. pp. 381–397.
117. McCarthy A., Ghadafi E., Andriotis P., Legg P. Functionality-Preserving Adversarial Machine Learning for Robust Classification in Cybersecurity and Intrusion Detection Domains: A Survey. *Journal of Cybersecurity and Privacy*. 2022. vol. 2(1). pp. 154–190.
118. Goodfellow I., McDaniel P., Papernot N. Making Machine Learning Robust Against Adversarial Inputs. *Communications of the ACM*. 2018. vol. 61(7). pp. 56–66.
119. Li J., Yang Y., Sun J., Tomsovic K.L., Qi H. Towards Adversarial-Resilient Deep Neural Networks for False Data Injection Attack Detection in Power Grids. *Proceedings of the 32nd International Conference on Computer Communications and Networks (ICCCN)*. 2023. pp. 1–10.
120. Li Y., Wang Y. Defense Against Adversarial Attacks in Deep Learning. *Applied Sciences*. 2018. vol. 9(1).
121. Zantedeschi V., Nicolae M.-I., Rawat A. Efficient Defenses Against Adversarial Attacks. *Proceedings of the 10th ACM Workshop on Artificial Intelligence and Security (AISec '17)*. 2017. pp. 39–49.
122. Sarker A., Shen H., Sen T., Mendelson Q. Efficient Black-Box Adversarial Attacks for Deep Driving Maneuver Classification Models. *Proceedings of the 18th International Conference on Mobile Ad Hoc and Smart Systems*. 2021. pp. 536–544.
123. Gibert D., Zizzo G., Le Q., Planes J. Adversarial Robustness of Deep Learning-Based Malware Detectors via (De)Randomized Smoothing. *IEEE Access*. 2024. vol. 12. pp. 61152–61162.

124. Yuan X., He P., Zhu Q., Li X. Adversarial Examples: Attacks and Defenses for Deep Learning. *IEEE Transactions on Neural Networks and Learning Systems*. 2019. vol. 30(9). pp. 2805–2824.
125. Ozturk E., Mesut A. Comparison of Learned Image Compression Methods and JPEG. *Proceedings of the 2024 Innovations in Intelligent Systems and Applications Conference (ASYU)*. 2024. pp. 1–6.
126. Zhou Z., Firestone C. Humans can decipher adversarial images. *Nat. Commun*. 2019. vol. 10(1).
127. Mohammed A., Ali Z., Ahmad I. Enhancing adversarial robustness with randomized interlayer processing. *Expert Systems with Applications*. 2024. vol. 245.
128. Shi Y., Zeng H., Nguyen T.T. Adversarial Machine Learning for Network Security. *Proceedings of the International Symposium on Technologies for Homeland Security*. 2019. pp. 1–7.
129. Eleftheriadis C., Symeonidis A., Katsaros P. Adversarial robustness improvement for deep neural networks. *Machine Vision and Applications*. 2024. vol. 35(3).
130. Han S., Lin Ch., Shen Ch., Wang Q., Guan X. Interpreting Adversarial Examples in Deep Learning: A Review. *ACM Comput. Surv.* 2023. vol. 55. no. 14s. pp. 1–38.
131. Li B., Liu W. WAT: Improve the Worst-Class Robustness in Adversarial Training. In *Proceedings of the AAAI Conference on Artificial Intelligence*. 2023. vol. 37(12). pp. 14982–14990.
132. Liu N., Du M., Guo R., Liu H., Hu X. Adversarial Attacks and Defenses: An Interpretation Perspective. *SIGKDD Explor. Newsl*. 2021. vol. 23(1). pp. 86–99.
133. Peng Y., Fu G., Luo Y., Yu Q., Wang L. CNN-based Steganalysis Detects Adversarial Steganography via Adversarial Training and Feature Squeezing. *Proceedings of the 4th International Conference on Information Science, Parallel and Distributed Systems (ISPDS)*. 2023. pp. 165–169.
134. Wang W., Park Y., Lee T., Molloy I., Tang P., Xiong L. Utilizing Multimodal Feature Consistency to Detect Adversarial Examples on Clinical Summaries. *Proceedings of the 3rd Clinical Natural Language Processing Workshop, Association for Computational Linguistics*. 2020. pp. 259–268.
135. Shao J., Geng S., Fu Z., Xu W., Liu T., Hong S. CardioDefense: Defending against adversarial attack in ECG classification with adversarial distillation training. *Biomedical Signal Processing and Control*. 2024. vol. 91. DOI: 10.1016/j.bspc.2023.105922.
136. Li H., Xu X., Zhang X., Yang Sh., Li B. Qeba: Query-efficient boundary-based blackbox attack. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 2020. pp. 1221–1230.
137. RyouD Ha I., Yoo H., Kim D., Han B. Robust Image Denoising Through Adversarial Frequency Mixup. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 2024. pp. 2723–2732.
138. Tyukin I.Y., Higham D.J., Gorban A.N. On Adversarial Examples and Stealth Attacks in Artificial Intelligence Systems. *Proceedings of the International Joint Conference on Neural Networks (IJCNN)*. 2020. pp. 1–6.
139. Shimizu Y. Automatic Design System with Generative Adversarial Network and Vision Transformer for Efficiency Optimization of Interior Permanent Magnet Synchronous Motor. *IEEE Transactions on Industrial Electronics*. 2024. vol. 71(11). pp. 14600–14609.
140. Barth A., Rubinstein B.I.P., Sundararajan M., Mitchell J.C., Song D., Bartlett P.L. A Learning-Based Approach to Reactive Security. *IEEE Transactions on Dependable and Secure Computing*. 2012. vol. 9(4). pp. 482–493.

141. Aly A., Iqbal S., Youssef A., Mansour E. MEGR-APT: A Memory-Efficient APT Hunting System Based on Attack Representation Learning. *IEEE Transactions on Information Forensics and Security*. 2024. vol. 19. pp. 5257–5271.
142. Deng Y., Zheng X., Zhang T., Chen C., Lou G., Kim M. An Analysis of Adversarial Attacks and Defenses on Autonomous Driving Models. *Proceedings of the IEEE International Conference on Pervasive Computing and Communications (PerCom)*. 2020. pp. 1–10.
143. He X., Yao J., Wang Y., Tang Z., Cheung K.C., See S., Han B., Chu X. NAS-LID: Efficient Neural Architecture Search with Local Intrinsic Dimension. *Proceedings of the AAAI Conference on Artificial Intelligence*. 2023. vol. 37(6). pp. 7839–7847.
144. Biggio B., Roli F. Wild patterns: Ten years after the rise of adversarial machine learning. *Proceedings of the 2018 ACM SIGSAC Conference on Computer and Communications Security (CCS '18)*. 2018. pp. 2154–2156.
145. Kwatra S., Torra V. Data Reconstruction Attack against Principal Component Analysis. *Security and Privacy in Social Networks and Big Data (SocialSec 2023)*. *Lecture Notes in Computer Sciences*. 2023. vol. 14097. pp. 79–92.
146. Pachika S., Reddy A.B., Pachika B., Karnam A. Generative Adversarial Networks: Overview. *Proceedings of the Fifth International Conference on Computer and Communication Technologies (IC3T 2023)*. *Lecture Notes in Computer Sciences*. 2024. vol. 897. pp. 319–328.
147. Li H., Yu W., Huang H. Strengthening transferability of adversarial examples by adaptive inertia and amplitude spectrum dropout. *Neural Networks*. 2023. vol. 165. pp. 925–937.
148. Chen Z. Robust Sparse Online Learning through Adversarial Sparsity Constraints. *Proceedings of the 2024 9th IEEE International Conference on Smart Cloud (SmartCloud)*. 2024. pp. 42–47.
149. Liu X., Chen X., Cheng J., Zhou L., Chen L., Li C., Zu S. Simulation of Complex Geological Architectures Based on Multistage Generative Adversarial Networks Integrating with Attention Mechanism and Spectral Normalization. *IEEE Transactions on Geoscience and Remote Sensing*. 2023. vol. 61. pp. 1–15.
150. Ahmed S., Islam S. Methods in detection of median filtering in digital images: a survey. *Multimed. Tools Appl*. 2023. vol. 82. pp. 43945–43965.
151. Elderman R., Pater L.J.J., Thie A.S., Drugan M.M., Wiering M.A. Adversarial Reinforcement Learning in a Cyber Security Simulation. *Proceedings of the 9th International Conference on Agents and Artificial Intelligence*. 2017. pp. 559–566.
152. Ling Y., Yong Z., Pengfei W. Euclidean and Rapid Jacobian-based Saliency Maps Attacks. *Proceedings of the 16th International Conference on Intelligent Systems and Knowledge Engineering (ISKE)*. 2021. pp. 355–361.
153. Vorobeychik Y., Kantarcioglu M. Introduction. *Adversarial Machine Learning. Synthesis Lectures on Artificial Intelligence and Machine Learning*. 2018. pp. 1–4.
154. Kotenko I., Saenko I., Lauta O., Vasiliev N., Sadovnikov V. A Noise-Based Approach Augmented with Neural Cleanse and JPEG Compression to Counter Adversarial Attacks against Image Classification Systems. *Proceedings of the 33rd Euromicro International Conference on Parallel, Distributed, and Network-Based Processing (PDP)*. 2025. pp. 576–583.

Kotenko Igor — Ph.D., Dr.Sci., Professor, Honored scientist of the Russian Federation, Chief researcher, head of the laboratory, Laboratory of computer security problems, St. Petersburg Federal Research Center of the Russian Academy of Sciences (SPC RAS). Research interests: computer network security, including security policy management, access control, authentication, security analysis, detection of computer attacks, firewalls, protection against viruses and network worms, analysis and verification of security protocols and information

security systems, protection of software from hacking and digital rights management, modeling and visualization technologies to counter cyberterrorism. The number of publications — 1000. ivkote@comsec.spb.ru; 39, 14-th Line V.O., 199178, St. Petersburg, Russia; office phone: +7(812)328-7181; fax: +7(812)328-4450.

Saenko Igor — Ph.D., Dr.Sci., Professor, Chief researcher, Laboratory of computer security problems, St. Petersburg Federal Research Center of the Russian Academy of Sciences (SPC RAS). Research interests: automated information systems, information security, data processing and transmission via communication channels, modeling theory and mathematical statistics, information theory. The number of publications — 500. ibsaen@comsec.spb.ru; 39, 14-th Line V.O., 199178, St. Petersburg, Russia; office phone: +7(812)328-7181; fax: +7(812)328-4450.

Lauta Oleg — Ph.D., Dr.Sci., Associate Professor, Professor of the department, Department of comprehensive information security, Admiral Makarov State University of Maritime and Inland Shipping. Research interests: protection against computer attacks. The number of publications — 200. laos-82@yandex.ru; 5/7, Dvinskaya St., 198035, St. Petersburg, Russia; office phone: +7(911)842-0228.

Sadovnikov Vladimir — Postgraduate student, Laboratory of computer security problems, St. Petersburg Federal Research Center of the Russian Academy of Sciences (SPC RAS). Research interests: protection against computer attacks. The number of publications — 25. bladimir1998@mail.ru; 39, 14-th Line V.O., 199178, St. Petersburg, Russia; office phone: +7(812)328-7181; fax: +7(812)328-4450.

Ichetovkin Egor — postgraduate student, Laboratory of computer security problems, St. Petersburg Federal Research Center of the Russian Academy of Sciences (SPC RAS). Research interests: protection against computer attacks. The number of publications — 7. egor.email@list.ru; 39, 14-th Line V.O., 199178, St. Petersburg, Russia; office phone: +7(812)328-7181; fax: +7(812)328-4450.

Li Wei — Professor, College of computer science and technology, Harbin Engineering University. Research interests: protection against computer attacks, machine learning, artificial intelligence. The number of publications — 200. wei.li@hrbeu.edu.cn; 145, улица Наньтун, район Нанган, 150001, Harbin, China; office phone: 0086(0451)8251921.

Acknowledgements. The research is partially supported by the budget theme FFZF-2025-0016.

И.В. Котенко, И.Б. Саенко, О.С. Лаута, В.Е. Садовников,
Е.А. Ичетовкин, В. Ли

АНАЛИЗ СОВРЕМЕННЫХ ИССЛЕДОВАНИЙ ПО ЗАЩИТЕ ОТ СОСТЯЗАТЕЛЬНЫХ АТАК В ЭНЕРГЕТИЧЕСКИХ СИСТЕМАХ

Котенко И.В., Саенко И.Б., Лаута О.С., Садовников В.Е., Ичетовкин Е.А., Ли В. Анализ современных исследований по защите от состязательных атак в энергетических системах.

Аннотация. Системы на основе машинного обучения в настоящее время являются привлекательными мишенями для злоумышленников, поскольку нарушение работы таких систем может иметь серьезные последствия для объектов критической инфраструктуры, в частности, энергетических систем. В связи с этим количество различных типов кибератак на системы машинного обучения, которые называются состязательными атаками, постоянно растёт, и эти атаки являются предметом изучения многих исследователей. Соответственно, ежегодно появляется множество публикаций, посвящённых обзорам состязательных атак и методов защиты от них. Многие виды состязательных атак и методы защиты в этих обзорных статьях пересекаются. Однако в более поздних исследованиях содержится информация о новых типах атак и методах защиты. Цель данной статьи – проанализировать исследования, проведённые за последние шесть лет и опубликованные в высокорейтинговых журналах, с акцентом на обзорные работы. Результатом исследования является уточнённая классификация состязательных атак, характеристика наиболее распространённых атак, а также уточнённая классификация и характеристика методов защиты от этих атак. Основное внимание в анализе уделяется состязательным атакам, нацеленным на энергетические системы. В заключительной части статьи рассматриваются преимущества и недостатки различных методов противодействия состязательным атакам.

Ключевые слова: кибератаки, искусственный интеллект, машинное обучение, состязательные атаки, модель угроз, методы защиты, обзор, энергетические системы, классификация.

Литература

1. Nascimento E.d.S., Ahmed I., Oliveira E., Palheta M.P., Steinmacher I., Conte T. Understanding Development Process of Machine Learning Systems: Challenges and Solutions. Proceedings of the ACM/IEEE International Symposium on Empirical Software Engineering and Measurement (ESEM). 2019. pp. 1–6.
2. Farhat R., Mourali Y., Jemni M., Ezzedine H. An overview of Machine Learning Technologies and their use in E-learning. Proceedings of the 2020 International Multi-Conference on Organization of Knowledge and Advanced Technologies (OCTA). 2020. pp. 1–4.
3. Liu Q., Li P., Zhao W., Cai W., Yu S., Leung V.C.M. A Survey on Security Threats and Defensive Techniques of Machine Learning: A Data Driven View. IEEE Access. 2018. vol. 6. pp. 12103–12117.
4. Pitropakis N., Panaousis E., Giannetsos T., Anastasiadis E., Loukas G. A taxonomy and survey of attacks against machine learning. Computer Science Review. 2019. vol. 34.
5. Tcukanova O.A., Yarskaya A.A., Torosyan A.A. Artificial Intelligence as a New Stage in the Development of Business Intelligence Systems. Proceedings of the

- International Conference on Quality Management, Transport and Information Security, Information Technologies (IT&QM&IS). 2022. pp. 315–318.
6. Bland J.A., Petty M.D., Whitaker T.S., Maxwell K.P., Cantrell W.A. Machine Learning Cyberattack and Defense Strategies. *Computers & Security*. 2020. vol. 92.
 7. Onyeji I., Bazilian M., Bronk C. Cyber security and critical energy infrastructure. *The Electricity Journal*. 2014. vol. 27. pp. 52–60.
 8. Santis E.D., Rizzi A., Sadeghian A. A cluster-based dissimilarity learning approach for localized fault classification in smart grids. *Swarm Evol. Comput.* 2018. vol. 39. pp. 267–278.
 9. Umapathy K., Dinesh Kumar T., Poojitha G., Khyathi Sri D., Pavaneeswar C., Amannah C. Machine Learning Applications for the Smart Grid. *Data Analytics for Smart Grids Applications – A Key to Smart City Development*. Intelligent Systems Reference Library. 2023. vol. 247. pp. 251–270.
 10. Zhao Z., Chen G. An Overview of Cyber Security for Smart Grid. *Proceedings of the 2018 IEEE 27th International Symposium on Industrial Electronics (ISIE)*. 2018. pp. 1127–1131.
 11. Biggio B., Corona I., Maiorca D., Nelson B., Srndic N., Laskov P., Giacinto G., Roli F. Evasion Attacks against Machine Learning at Test Time. *Advanced Information Systems Engineering. Lecture Notes in Computer Science*. 2013. vol. 7908. pp. 387–402.
 12. Goodfellow I.J., Shlens J., Szegedy C. Explaining and Harnessing Adversarial Examples. *Proceedings of the International Conference on Learning Representations (ICLR'15)*. 2015. arXiv preprint arXiv:1412.6572.
 13. Nguyen T.N., Liu B.-H., Nguyen N.P., Chou J.-T. Cyber Security of Smart Grid: Attacks and Defenses. *Proceedings of the IEEE International Conference on Communications (ICC)*. 2020. pp. 1–6.
 14. Kawoosa A.I., Prashar D. A Review of Cyber Securities in Smart Grid Technology. *Proceedings of the 2nd International Conference on Computation, Automation and Knowledge Management (ICCAKM)*. 2021. pp. 151–156.
 15. Madhavarapu V.P.K., Bhattacharjee S., Dasy S.K. A Generative Model for Evasion Attacks in Smart Grid. *Proceedings of the IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*. 2022. pp. 1–6.
 16. Cui L., Qu Y., Gao L., Xie G., Yu S. Detecting false data attacks using machine learning techniques in smart grid: A survey. *J. Netw. Comput. Appl.* 2020. vol. 170.
 17. Kotenko I., Saenko I., Lauta O., Kribel K., Vasiliev N. Attacks Against Artificial Intelligence Systems: Classification, The Threat Model and the Approach to Protection. *Proceedings of the Sixth International Scientific Conference “Intelligent Information Technologies for Industry” (IITI'22)*. 2023. pp. 293–302.
 18. Kotenko I., Saenko I., Lauta O., Vasiliev N., Iatsenko D. Attacks Against Machine Learning Systems: Analysis and GAN-based Approach to Protection. *Proceedings of the Seventh International Scientific Conference «Intelligent Information Technologies for Industry» (IITI'23)*. 2023. pp. 49–59.
 19. Huang X., Kroening D., Ruan W., Sharp J., Sun Y., Thamo E., Wu M., Yi X. A survey of safety and trustworthiness of deep neural networks: Verification, testing, adversarial attack and defence, and interpretability. *Computer Science Review*. 2020. vol. 37.
 20. Martins N., Cruz J.M., Cruz T., Henriques Abreu P. Adversarial Machine Learning Applied to Intrusion and Malware Scenarios: A Systematic Review. *IEEE Access*. 2020. vol. 8. pp. 35403–35419.
 21. Oseni A., Moustafa N., Janicke H., Liu P., Tari Z., Vasilakos A. Security and Privacy for Artificial Intelligence: Opportunities and Challenges. 2020. arXiv preprint arXiv:2102.04661.

22. Xu H., Ma Y., Liu H.C., Deb D., Liu H., Tang J.-L., Jain A.K. Adversarial Attacks and Defenses in Images, Graphs and Text: A Review. *Int. J. Autom. Comput.* 2020. vol. 17. pp. 151–178.
23. Ren K., Zheng T., Qin Zh., Liu X. Adversarial Attacks and Defenses in Deep Learning. *Engineering.* 2020. vol. 6. pp. 346–360.
24. Zhou X., Canady R., Li Y., Koutsoukos X., Gokhale A. Overcoming Stealthy Adversarial Attacks on Power Grid Load Predictions Through Dynamic Data Repair. *Dynamic Data Driven Applications Systems (DDDAS 2020). Lecture Notes in Computer Science.* 2020. vol. 12312. pp. 102–109.
25. Akhtar N., Mian A., Kardan N., Shah M. Advances in Adversarial Attacks and Defenses in Computer Vision: A Survey. *IEEE Access.* 2021. vol. 9. pp. 155161–155196.
26. Zhang H., Liu B., Wu H. Smart Grid Cyber-Physical Attack and Defense: A Review. *IEEE Access.* 2021. vol. 9. pp. 29641–29659.
27. Chakraborty A., Alam M., Dey V., Chattopadhyay A., Mukhopadhyay D. A survey on adversarial attacks and defences. *CAAI Trans. Intell. Technol.* 2021. vol. 6. pp. 25–45.
28. Rosenberg I., Shabtai A., Elovici Y., Rokach L. Adversarial Machine Learning Attacks and Defense Methods in the Cyber Security Domain. *ACM Comput. Surv.* 2021. vol. 54(5). pp. 1–36.
29. Tian J., Wang B., Li J., Konstantinou C. Adversarial attack and defense methods for neural network-based state estimation in smart grid. *IET Renewable Power Generation.* 2022. vol. 16. pp. 3507–3518.
30. Kong Z., Xue J., Wang Y., Huang L., Niu Z., Li F., Meng W. A Survey on Adversarial Attack in the Age of Artificial Intelligence. *Wireless Communications and Mobile Computing.* 2021. vol. 2021(1).
31. Zhou Sh., Liu Ch., Ye D., Zhu T., Zhou W., Yu Ph.S. Adversarial Attacks and Defenses in Deep Learning: From a Perspective of Cybersecurity. *ACM Comput. Surv.* 2022. vol. 55(8). pp. 1–39.
32. Khamaisch S.Y., Bagagem D., Al-Alaj A., Mancino M., Alomari H.W. Adversarial Deep Learning: A Survey on Adversarial Attacks and Defense Mechanisms on Image Classification. *IEEE Access.* 2022. vol. 10. pp. 102266–102291.
33. Liang H., He E., Zhao Y., Jia Z., Li H. Adversarial Attack and Defense: A Survey. *Electronics.* 2022. vol. 11.
34. Tian Q., Zhang S., Mao Sh., Lin Y. Adversarial attacks and defenses for digital communication signals identification. *Digital Communications and Networks.* 2024. vol. 10. no. 3. pp. 756–764.
35. Anastasiou Th., Karagiorgou S., Petrou P., Papamartzivanos D., Giannetsos Th., Tsirigotaki G., Keizer J. Towards Robustifying Image Classifiers against the Perils of Adversarial Attacks on Artificial Intelligence Systems. *Sensors.* 2022. vol. 22.
36. Li Y., Cheng M., Hsieh Ch.-J., Lee Th.C.M. A Review of Adversarial Attack and Defense for Classification Methods. *The American Statistician.* 2022. vol. 76(4). pp. 329–345.
37. Tian J., Wang B., Li J., Wang Z. Adversarial Attacks and Defense for CNN Based Power Quality Recognition in Smart Grid. *IEEE Transactions on Network Science and Engineering.* 2021. vol. 9(2). pp. 807–819.
38. Li H., Namiot D. A Survey of Adversarial Attacks and Defenses for Image Data on Deep Learning. *International Journal of Open Information Technologies.* 2022. vol. 10. pp. 9–16.
39. Girdhar M., Hong J., Moore J. Cybersecurity of Autonomous Vehicles: A Systematic Literature Review of Adversarial Attacks and Defense Models. *IEEE Open Journal of Vehicular Technology.* 2023. vol. 4. pp. 417–437.

40. Goyal Sh., Doddapaneni S., Khapra M.M., Ravindran B. A Survey of Adversarial Defenses and Robustness in NLP. *ACM Comput. Surv.* 2023. vol. 55.
41. Al-Khassawneh Y.A. A Review of Artificial Intelligence in Security and Privacy: Research Advances, Applications, Opportunities, and Challenges. *Indonesian Journal of Science and Technology.* 2023. vol. 8. pp. 79–96.
42. He K., Kim D.D., Asghar M.R. Adversarial Machine Learning for Network Intrusion Detection Systems: A Comprehensive Survey. *IEEE Communications Surveys & Tutorials.* 2023. vol. 25(1). pp. 538–566.
43. Sun L., Dou Y., Yang C., Zhang K., Wang J., Yu Ph.S., He L., Li B. Adversarial Attack and Defense on Graph Data: A Survey. *IEEE Transactions on Knowledge and Data Engineering.* 2023. vol. 35(8). pp. 7693–7711.
44. Qureshi A.U.H., Larijani H., Yousefi M., Adeel A., Mtetwa N. An Adversarial Approach for Intrusion Detection Systems Using Jacobian Saliency Map Attacks (JSMA) Algorithm. *Computers.* 2020. vol. 9(3).
45. Hoang V.-T., Ergu Y.A., Nguyen V.-L., Chang R.-G. Security risks and countermeasures of adversarial attacks on AI-driven applications in 6G networks: A survey. *Journal of Network and Computer Applications.* 2024. vol. 232.
46. Baniecki H., Biecek P. Adversarial attacks and defenses in explainable artificial intelligence: A survey. *Information Fusion.* 2024. vol. 107.
47. Andrade E. de O., Guérin J., Viterbo J., Sampaio I.G.B. Adversarial attacks and defenses in person search: A systematic mapping study and taxonomy. *Image and Vision Computing.* 2024. vol. 148.
48. Zhang Ch., Hu M., Li W., Wang L. Adversarial attacks and defenses on text-to-image diffusion models: A survey. *Information Fusion.* 2025. vol. 114.
49. Hao J., Tao Y. Adversarial attacks on deep learning models in smart grids. *Energy Reports.* 2022. vol. 8. pp. 123–129.
50. Sande-Rios J., Canal-Sánchez J., Manzano-Hernández C., Pastrana S. Threat Analysis and Adversarial Model for Smart Grids. *Proceeding of the IEEE European Symposium on Security and Privacy Workshops (EuroS&PW).* 2024. pp. 130–145.
51. Elsisi M., Su C.-L., Ali M.N. Design of Reliable IoT Systems With Deep Learning to Support Resilient Demand Side Management in Smart Grids Against Adversarial Attacks. *Transactions on Industry Applications.* 2024. vol. 60. pp. 2095–2106.
52. Ahmadian S., Malki H., Han Z. Cyber Attacks on Smart Energy Grids Using Generative Adversarial Networks. *Proceedings of the IEEE Global Conference on Signal and Information Processing (GlobalSIP).* 2018. pp. 942–946.
53. Takiddin A., Ismail M., Zafar U., Serpedin E. Robust Electricity Theft Detection Against Data Poisoning Attacks in Smart Grids. *IEEE Transactions on Smart Grid.* 2021. vol. 12. pp. 2675–2684.
54. Takiddin A., Ismail M., Serpedin E. Robust Data-Driven Detection of Electricity Theft Adversarial Evasion Attacks in Smart Grids. *IEEE Transactions on Smart Grid.* 2023. vol. 14. pp. 663–676.
55. Bondok A.H., Mahmoud M., Badr M.M., Fouda M.M., Abdallah M., Alsabaan M. Novel Evasion Attacks against Adversarial Training Defense for Smart Grid Federated Learning. *IEEE Access.* 2023. vol. 11. pp. 112953–112972.
56. Sampedro G.A., Ojo S., Krichen M., Alamro M.A., Mihoub A., Karovic V. Defending AI Models Against Adversarial Attacks in Smart Grids Using Deep Learning. *IEEE Access.* 2024. vol. 12. pp. 157408–157417. DOI: 10.1109/ACCESS.2024.3473531.
57. Khan S.U., Mynuddin M., Nabil M. AdaptEdge: Targeted Universal Adversarial Attacks on Time Series Data in Smart Grids. *IEEE Transactions on Smart Grid.* 2024. vol. 15. pp. 5072–5086.

58. Berghout T., Benbouzid M., Amirat Y. Towards Resilient and Secure Smart Grids against PMU Adversarial Attacks: A Deep Learning-Based Robust Data Engineering Approach. *Electronics*. 2023. vol. 12.
59. Teryak H., Albaseer A., Abdallah M., Al-Kuwari S., Qaraqe M. Double-Edged Defense: Thwarting Cyber Attacks and Adversarial Machine Learning in IEC 60870-5-104 Smart Grids. *IEEE Open Journal of the Industrial Electronics Society*. 2023. vol. 4. pp. 629–642.
60. Ness S. Adversarial Attack Detection in Smart Grids Using Deep Learning Architectures. *IEEE Access*. 2025. vol. 13. pp. 16314–16323. DOI: 10.1109/ACCESS.2024.3523409.
61. Zhang G., Sikdar B. Ensemble and Transfer Adversarial Attack on Smart Grid Demand-Response Mechanisms. *Proceedings of the IEEE International Conference on Communications, Control, and Computing Technologies for Smart Grids (SmartGridComm)*. IEEE, 2022. pp. 53–58.
62. Zhang G., Sikdar B. A Novel Adversarial FDI Attack and Defense Mechanism for Smart Grid Demand-Response Mechanisms. *IEEE Transactions on Industrial Cyber-Physical Systems*. 2024. vol. 2. pp. 380–390. DOI: 10.1109/TICPS.2024.3448380.
63. Reda H.T., Anwar A., Mahmood A. Comprehensive survey and taxonomies of false data injection attacks in smart grids: attack models, targets, and impacts. *Renewable and Sustainable Energy Reviews*. 2022. vol. 163.
64. Ren Y., Zhang H., Yang W., Li M., Zhang J., Li H. Transferable Adversarial Attack Against Deep Reinforcement Learning-Based Smart Grid Dynamic Pricing System. *IEEE Transactions on Industrial Informatics*. 2024. vol. 20. pp. 9015–9025.
65. He J., Xiang T., Wu T., Chen Z., Wang N., Guo S. Maintaining Privacy in Smart Grid: Utilizing the Adversarial Attack Paradigm to Counter Non-Intrusive Load Monitoring Models. *IEEE Internet of Things Journal*. 2024.
66. Guihai Z., Sikdar B. Adversarial Machine Learning Against False Data Injection Attack Detection for Smart Grid Demand Response. *Proceedings of the IEEE International Conference on Communications, Control, and Computing Technologies for Smart Grids (SmartGridComm)*. 2021. pp. 352–357.
67. Aurangzeb M., Wang Y., Iqbal S., Naveed A., Ahmed Z., Alenezi M., Shouran M. Enhancing cybersecurity in smart grids: Deep black box adversarial attacks and quantum voting ensemble models for blockchain privacy-preserving storage. *Energy Reports*. 2024. vol. 11. pp. 2493–2515.
68. Melendez K.A., Matamala Y. Adversarial attacks in demand-side electricity markets. *Applied Energy*. 2025. vol. 377.
69. Nguyen T., Wang S., Alhazmi M., Nazemi M., Estebarsari A., Dehghanian P. Electric Power Grid Resilience to Cyber Adversaries: State of the Art. *IEEE Access*. 2020. vol. 8. pp. 87592–87608.
70. Efatinasab E., Brighente A., Rampazzo M., Azadi N., Conti M. GAN-GRID: A Novel Generative Attack on Smart Grid Stability Prediction. *Computer Security – ESORICS 2024. Lecture Notes in Computer Science*. 2024. vol. 14982. pp. 374–393.
71. Takiddin A., Ismail M., Atat R., Serpedin E. Spatio-temporal Graph-Based Generation and Detection of Adversarial False Data Injection Evasion Attacks in Smart Grids. *IEEE Transactions on Artificial Intelligence*. 2024. vol. 5(12). pp. 6601–6616. DOI: 10.1109/TAI.2024.3464511.
72. Omara A., Kantarci B. An AI-driven solution to prevent adversarial attacks on mobile Vehicle-to-Microgrid services. *Simulation Modelling Practice and Theory*. 2024. vol. 137.
73. Tian J., Shen C., Wang B., Ren C., Xia X., Dong R., Cheng T. EVADE: Targeted Adversarial False Data Injection Attacks for State Estimation in Smart Grid. *IEEE Transactions on Sustainable Computing*. 2024.

74. Efatinasab E., Sinigaglia A., Azadi N., Susto G.A., Rampazzo M. Adversarially Robust Fault Zone Prediction in Smart Grids With Bayesian Neural Networks. *IEEE Access*. 2024. vol. 12. pp. 121169–121184.
75. Zhang Z., Liu M., Sun M., Deng R., Cheng P., Niyato D., Chow M.-Y., Chen J. Vulnerability of Machine Learning Approaches Applied in IoT-Based Smart Grid: A Review. *IEEE Internet of Things Journal*. 2024. vol. 11(11). pp. 18951–18975.
76. Sánchez G., Elbez G., Hagenmeyer V. Attacking Learning-based Models in Smart Grids: Current Challenges and New Frontiers. *Proceedings of the 15th ACM International Conference on Future and Sustainable Energy Systems (e-Energy '24)*. 2024. pp. 589–595.
77. Huang R., Li Y. Adversarial Attack Mitigation Strategy for Machine Learning-Based Network Attack Detection Model in Power System. *IEEE Transactions on Smart Grid*. 2022. vol. 14(3). pp. 2367–2376.
78. Ruan J., Wang Q., Chen S., Lyu H., Liang G., Zhao J., Dong Z.Y. On Vulnerability of Renewable Energy Forecasting: Adversarial Learning Attacks. *IEEE Transactions on Industrial Informatics*. 2023. vol. 20(3). pp. 3650–3663.
79. Heinrich R., Scholz C., Vogt S., Lehna M. Targeted adversarial attacks on wind power forecasts. *Mach Learn*. 2024. vol. 113(2). pp. 863–889.
80. Tang N., Mao S., Nelms R.M. Adversarial Attacks to Solar Power Forecast. *Proceedings of the 2021 IEEE Global Communications Conference (GLOBECOM)*. 2021. pp. 1–6.
81. Khan S.U., Mynuddin M., Adom I., Mahmoud M.N. Mitigating Targeted Universal Adversarial Attacks on Time Series Power Quality Disturbances Models. *Proceedings of the 5th IEEE International Conference on Trust, Privacy and Security in Intelligent Systems and Applications (TPS-ISA)*. 2023. pp. 91–100.
82. Gunn S., Jang D., Paradise O., Spangher L., Spanos C.J. Adversarial poisoning attacks on reinforcement learning-driven energy pricing. *Proceedings of the 9th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation (BuildSys '22)*. 2024. pp. 262–265.
83. Zhang L., Jiang C., Pang A., He Y. Super-efficient detector and defense method for adversarial attacks in power quality classification. *Applied Energy*. 2024. vol. 361.
84. Sayghe A., Zhao J., Konstantinou C. Evasion Attacks with Adversarial Deep Learning Against Power System State Estimation. *Proceedings of the IEEE Power & Energy Society General Meeting (PESGM)*. 2020. pp. 1–5.
85. Akter K., Rahman M.A., Islam R.M., Sheikh R.I., Hossain M.J. Attack-resilient framework for wind power forecasting against civil and adversarial attacks. *Electric Power Systems Research*. 2025. vol. 238.
86. Li J., Wang J., Chen L., Yu Y. Defending Against Adversarial Attacks by Energy Storage Facility. *Proceedings of the IEEE Power & Energy Society General Meeting (PESGM)*. 2022. pp. 1–5.
87. Farajzadeh-Zanjani M., Hallaji E., Razavi-Far R., Saif M., Parvania M. Adversarial Semi-Supervised Learning for Diagnosing Faults and Attacks in Power Grids. *IEEE Transactions on Smart Grid*. 2021. vol. 12(4). pp. 3468–3478.
88. Zhou Y., Ding Z., Wen Q., Wang Y. Robust Load Forecasting Towards Adversarial Attacks via Bayesian Learning. *IEEE Transactions on Power Systems*. 2023. vol. 38. pp. 1445–1459.
89. Zhang L., Jiang C., Chai Z., He Y. Adversarial attack and training for deep neural network based power quality disturbance classification. *Engineering Applications of Artificial Intelligence*. 2024. vol. 127.
90. Li W., Deka D., Wang R., Paternina M.R.A. Physics-Constrained Adversarial Training for Neural Networks in Stochastic Power Grids. *IEEE Transactions on*

- Artificial Intelligence. 2024. vol. 5(3). pp. 1121–1131. DOI: 10.1109/TAI.2023.3236377.
91. Afrin A., Ardakanian O. Adversarial Attacks on Machine Learning-Based State Estimation in Power Distribution Systems. *Proceedings of the 14th ACM International Conference on Future Energy Systems (e-Energy '23)*. 2023. pp. 446–458.
92. Tian J., Wang B., Wang Z., Cao K., Li J., Ozay M. Joint Adversarial Example and False Data Injection Attacks for State Estimation in Power Systems. *IEEE Transactions on Cybernetics*. 2022. vol. 52(12). pp. 13699–13713.
93. Ali M.N., Amer M., Elsisy M. Reliable IoT Paradigm with Ensemble Machine Learning for Faults Diagnosis of Power Transformers Considering Adversarial Attacks. *IEEE Transactions on Instrumentation and Measurement*. 2023. vol. 72. pp. 1–13.
94. Kim B., Shi Y., Sagduyu Y.E., Erpek T., Ulukus S. Adversarial Attacks against Deep Learning Based Power Control in Wireless Communications. *Proceedings of the IEEE Globecom Workshops (GC Wkshps)*. 2021. pp. 1–6.
95. Manoj B.R., Sadeghi M., Larsson E.G. Adversarial Attacks on Deep Learning Based Power Allocation in a Massive MIMO Network. *Proceedings of the ICC – IEEE International Conference on Communications*. 2021. pp. 1–6.
96. Manoj B.R., Sadeghi M., Larsson E.G. Downlink Power Allocation in Massive MIMO via Deep Learning: Adversarial Attacks and Training. *IEEE Transactions on Cognitive Communications and Networking*. 2022. vol. 8(2). pp. 707–719.
97. Santos P.M., Manoj B.R., Sadeghi M., Larsson E.G. Universal Adversarial Attacks on Neural Networks for Power Allocation in a Massive MIMO System. *IEEE Wireless Communications Letters*. 2022. vol. 11(1). pp. 67–71.
98. Shabbir A., Manzoor H.U., Ahmed R.A., Halim Z. Resilience of Federated Learning Against False Data Injection Attacks in Energy Forecasting. *Proceedings of the 2024 International Conference on Green Energy, Computing and Sustainable Technology (GECOST)*. 2024. pp. 245–249.
99. Xu S., Yu L., Lin X. Robust Low-Overhead Control of DER Reactive Power under Adversarial Attacks and Uncertainty. *Proceedings of the ICC 2024 – IEEE International Conference on Communications*. 2024. pp. 3097–3103.
100. Kuzlu M., Sarp S., Catak F.O., Cali U., Zhao Y., Elma O., Guler O. Analysis of deceptive data attacks with adversarial machine learning for solar photovoltaic power generation forecasting. *Electr. Eng.* 2024. vol. 106(2). pp. 1815–1823.
101. Zideh M.J., Khalghani M.R., Solanki S.K. An unsupervised adversarial autoencoder for cyber attack detection in power distribution grids. *Electric Power Systems Research*. 2024. vol. 232.
102. Zeng L., Qiu D., Sun M. Resilience enhancement of multi-agent reinforcement learning-based demand response against adversarial attacks. *Applied Energy*. 2022. vol. 324.
103. Zografopoulos I., Hatziaargyriou N.D., Konstantinou C. Distributed Energy Resources Cybersecurity Outlook: Vulnerabilities, Attacks, Impacts, and Mitigations. *IEEE Systems Journal*. 2023. vol. 17. pp. 6695–6709.
104. Santana E.J., Silva R.P., Zarpelão B.B., Barbon Junior S. Detecting and Mitigating Adversarial Examples in Regression Tasks: A Photovoltaic Power Generation Forecasting Case Study. *Information*. 2021. vol. 12(10).
105. Youssef E.-N.S., Labeau F., Kassouf M. Adversarial Dynamic Load-Altering Cyberattacks Against Peak Shaving Using Residential Electric Water Heaters. *IEEE Transactions on Smart Grid*. 2024. vol. 15(2). pp. 2073–2088.
106. Bhattacharjee A., Bai G., Tushar W., Verma A., Mishra S., Saha T.K. DeeBBAA: A Benchmark Deep Black-Box Adversarial Attack Against Cyber–Physical Power

- Systems. IEEE Internet of Things Journal. 2024. vol. 11(24). pp. 40670–40688. DOI: 10.1109/JIOT.2024.3454257.
107. Genç D., Özüysal M., Tomur E. A Taxonomic Survey of Model Extraction Attacks. Proceedings of the 2023 IEEE International Conference on Cyber Security and Resilience (CSR). 2023. pp. 200–205.
 108. Fan J., Yan Q., Li M., Qu G., Xiao Y. A Survey on Data Poisoning Attacks and Defenses. Proceedings of the 2022 7th IEEE International Conference on Data Science in Cyberspace (DSC). 2022. pp. 48–55.
 109. Wang S., Ko R.K.L., Bai G., Dong N., Choi T., Zhang Y. Evasion Attack and Defense on Machine Learning Models in Cyber-Physical Systems: A Survey. IEEE Communications Surveys & Tutorials. 2024. vol. 26. pp. 930–966.
 110. El-Toukhy A.T., Mahmoud M., Bondok A.H., Fouda M.M., Alsabaan M. Evasion Attacks in Smart Power Grids: A Deep Reinforcement Learning Approach. Proceedings of the 2024 IEEE 21st Consumer Communications & Networking Conference (CCNC). 2024. pp. 708–713.
 111. Ali Alatwi H., Morisset C. Threat Modeling for Machine Learning-Based Network Intrusion Detection Systems. Proceedings of the IEEE International Conference on Big Data (Big Data). 2022. pp. 4226–4235.
 112. Qiu S., Liu Q., Zhou S., Wu C. Review of Artificial Intelligence Adversarial Attack and Defense Technologies. Applied sciences. 2019. vol. 9(5).
 113. Tuna O.F., Catak F.O., Eskil M.T. Exploiting epistemic uncertainty of the deep learning models to generate adversarial samples. Multimed Tools Appl. 2022. vol. 81. pp. 11479–11500.
 114. Dubrovkin J. Evaluation of undetectable perturbations of the peak parameters estimated by the least square curve fitting of analytical signal consisting of overlapping peaks. Chemom. and Intel. Lab. Syst. 2016. vol. 153. pp. 9–21.
 115. Lecun Y., Bottou L., Bengio Y., Haffner P. Gradient-based learning applied to document recognition. Proceedings of the IEEE. 2002. vol. 86(11). pp. 2278–2324.
 116. Liu X., Cheng M., Zhang H., Hsieh C.J. Towards Robust Neural Networks via Random Self-ensemble. Computer Vision – ECCV 2018. Lecture Notes in Computer Science. 2018. vol. 11211. pp. 381–397.
 117. McCarthy A., Ghadafi E., Andriotis P., Legg P. Functionality-Preserving Adversarial Machine Learning for Robust Classification in Cybersecurity and Intrusion Detection Domains: A Survey. Journal of Cybersecurity and Privacy. 2022. vol. 2(1). pp. 154–190.
 118. Goodfellow I., McDaniel P., Papernot N. Making Machine Learning Robust Against Adversarial Inputs. Communications of the ACM. 2018. vol. 61(7). pp. 56–66.
 119. Li J., Yang Y., Sun J., Tomsovic K.L., Qi H. Towards Adversarial-Resilient Deep Neural Networks for False Data Injection Attack Detection in Power Grids. Proceedings of the 32nd International Conference on Computer Communications and Networks (ICCCN). 2023. pp. 1–10.
 120. Li Y., Wang Y. Defense Against Adversarial Attacks in Deep Learning. Applied Sciences. 2018. vol. 9(1).
 121. Zantedeschi V., Nicolae M.-I., Rawat A. Efficient Defenses Against Adversarial Attacks. Proceedings of the 10th ACM Workshop on Artificial Intelligence and Security (AISec '17). 2017. pp. 39–49.
 122. Sarker A., Shen H., Sen T., Mendelson Q. Efficient Black-Box Adversarial Attacks for Deep Driving Maneuver Classification Models. Proceedings of the 18th International Conference on Mobile Ad Hoc and Smart Systems. 2021. pp. 536–544.
 123. Gibert D., Zizzo G., Le Q., Planes J. Adversarial Robustness of Deep Learning-Based Malware Detectors via (De)Randomized Smoothing. IEEE Access. 2024. vol. 12. pp. 61152–61162.

124. Yuan X., He P., Zhu Q., Li X. Adversarial Examples: Attacks and Defenses for Deep Learning. *IEEE Transactions on Neural Networks and Learning Systems*. 2019. vol. 30(9). pp. 2805–2824.
125. Ozturk E., Mesut A. Comparison of Learned Image Compression Methods and JPEG. *Proceedings of the 2024 Innovations in Intelligent Systems and Applications Conference (ASYU)*. 2024. pp. 1–6.
126. Zhou Z., Firestone C. Humans can decipher adversarial images. *Nat. Commun*. 2019. vol. 10(1).
127. Mohammed A., Ali Z., Ahmad I. Enhancing adversarial robustness with randomized interlayer processing. *Expert Systems with Applications*. 2024. vol. 245.
128. Shi Y., Zeng H., Nguyen T.T. Adversarial Machine Learning for Network Security. *Proceedings of the International Symposium on Technologies for Homeland Security*. 2019. pp. 1–7.
129. Eleftheriadis C., Symeonidis A., Katsaros P. Adversarial robustness improvement for deep neural networks. *Machine Vision and Applications*. 2024. vol. 35(3).
130. Han S., Lin Ch., Shen Ch., Wang Q., Guan X. Interpreting Adversarial Examples in Deep Learning: A Review. *ACM Comput. Surv.* 2023. vol. 55. no. 14s. pp. 1–38.
131. Li B., Liu W. WAT: Improve the Worst-Class Robustness in Adversarial Training. In *Proceedings of the AAAI Conference on Artificial Intelligence*. 2023. vol. 37(12). pp. 14982–14990.
132. Liu N., Du M., Guo R., Liu H., Hu X. Adversarial Attacks and Defenses: An Interpretation Perspective. *SIGKDD Explor. Newsl*. 2021. vol. 23(1). pp. 86–99.
133. Peng Y., Fu G., Luo Y., Yu Q., Wang L. CNN-based Steganalysis Detects Adversarial Steganography via Adversarial Training and Feature Squeezing. *Proceedings of the 4th International Conference on Information Science, Parallel and Distributed Systems (ISPDS)*. 2023. pp. 165–169.
134. Wang W., Park Y., Lee T., Molloy I., Tang P., Xiong L. Utilizing Multimodal Feature Consistency to Detect Adversarial Examples on Clinical Summaries. *Proceedings of the 3rd Clinical Natural Language Processing Workshop, Association for Computational Linguistics*. 2020. pp. 259–268.
135. Shao J., Geng S., Fu Z., Xu W., Liu T., Hong S. CardioDefense: Defending against adversarial attack in ECG classification with adversarial distillation training. *Biomedical Signal Processing and Control*. 2024. vol. 91. DOI: 10.1016/j.bspc.2023.105922.
136. Li H., Xu X., Zhang X., Yang Sh., Li B. Qeba: Query-efficient boundary-based blackbox attack. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 2020. pp. 1221–1230.
137. RyouD Ha I., Yoo H., Kim D., Han B. Robust Image Denoising Through Adversarial Frequency Mixup. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 2024. pp. 2723–2732.
138. Tyukin I.Y., Higham D.J., Gorban A.N. On Adversarial Examples and Stealth Attacks in Artificial Intelligence Systems. *Proceedings of the International Joint Conference on Neural Networks (IJCNN)*. 2020. pp. 1–6.
139. Shimizu Y. Automatic Design System with Generative Adversarial Network and Vision Transformer for Efficiency Optimization of Interior Permanent Magnet Synchronous Motor. *IEEE Transactions on Industrial Electronics*. 2024. vol. 71(11). pp. 14600–14609.
140. Barth A., Rubinstein B.I.P., Sundararajan M., Mitchell J.C., Song D., Bartlett P.L. A Learning-Based Approach to Reactive Security. *IEEE Transactions on Dependable and Secure Computing*. 2012. vol. 9(4). pp. 482–493.

141. Aly A., Iqbal S., Youssef A., Mansour E. MEGR-APT: A Memory-Efficient APT Hunting System Based on Attack Representation Learning. *IEEE Transactions on Information Forensics and Security*. 2024. vol. 19. pp. 5257–5271.
142. Deng Y., Zheng X., Zhang T., Chen C., Lou G., Kim M. An Analysis of Adversarial Attacks and Defenses on Autonomous Driving Models. *Proceedings of the IEEE International Conference on Pervasive Computing and Communications (PerCom)*. 2020. pp. 1–10.
143. He X., Yao J., Wang Y., Tang Z., Cheung K.C., See S., Han B., Chu X. NAS-LID: Efficient Neural Architecture Search with Local Intrinsic Dimension. *Proceedings of the AAAI Conference on Artificial Intelligence*. 2023. vol. 37(6). pp. 7839–7847.
144. Biggio B., Roli F. Wild patterns: Ten years after the rise of adversarial machine learning. *Proceedings of the 2018 ACM SIGSAC Conference on Computer and Communications Security (CCS '18)*. 2018. pp. 2154–2156.
145. Kwatra S., Torra V. Data Reconstruction Attack against Principal Component Analysis. *Security and Privacy in Social Networks and Big Data (SocialSec 2023)*. *Lecture Notes in Computer Sciences*. 2023. vol. 14097. pp. 79–92.
146. Pachika S., Reddy A.B., Pachika B., Karnam A. Generative Adversarial Networks: Overview. *Proceedings of the Fifth International Conference on Computer and Communication Technologies (IC3T 2023)*. *Lecture Notes in Computer Sciences*. 2024. vol. 897. pp. 319–328.
147. Li H., Yu W., Huang H. Strengthening transferability of adversarial examples by adaptive inertia and amplitude spectrum dropout. *Neural Networks*. 2023. vol. 165. pp. 925–937.
148. Chen Z. Robust Sparse Online Learning through Adversarial Sparsity Constraints. *Proceedings of the 2024 9th IEEE International Conference on Smart Cloud (SmartCloud)*. 2024. pp. 42–47.
149. Liu X., Chen X., Cheng J., Zhou L., Chen L., Li C., Zu S. Simulation of Complex Geological Architectures Based on Multistage Generative Adversarial Networks Integrating with Attention Mechanism and Spectral Normalization. *IEEE Transactions on Geoscience and Remote Sensing*. 2023. vol. 61. pp. 1–15.
150. Ahmed S., Islam S. Methods in detection of median filtering in digital images: a survey. *Multimed. Tools Appl*. 2023. vol. 82. pp. 43945–43965.
151. Elderman R., Pater L.J.J., Thie A.S., Drugan M.M., Wiering M.A. Adversarial Reinforcement Learning in a Cyber Security Simulation. *Proceedings of the 9th International Conference on Agents and Artificial Intelligence*. 2017. pp. 559–566.
152. Ling Y., Yong Z., Pengfei W. Euclidean and Rapid Jacobian-based Saliency Maps Attacks. *Proceedings of the 16th International Conference on Intelligent Systems and Knowledge Engineering (ISKE)*. 2021. pp. 355–361.
153. Vorobeychik Y., Kantarcioglu M. Introduction. *Adversarial Machine Learning. Synthesis Lectures on Artificial Intelligence and Machine Learning*. 2018. pp. 1–4.
154. Kotenko I., Saenko I., Lauta O., Vasiliev N., Sadovnikov V. A Noise-Based Approach Augmented with Neural Cleanse and JPEG Compression to Counter Adversarial Attacks against Image Classification Systems. *Proceedings of the 33rd Euromicro International Conference on Parallel, Distributed, and Network-Based Processing (PDP)*. 2025. pp. 576–583.

Котенко Игорь Витальевич — д-р техн. наук, профессор, заслуженный деятель науки российской федерации, главный научный сотрудник, руководитель лаборатории, лаборатория проблем компьютерной безопасности, Федеральное государственное бюджетное учреждение науки «Санкт-Петербургский Федеральный исследовательский центр Российской академии наук» (СПб ФИЦ РАН). Область научных интересов: безопасность компьютерных сетей, включая управление политиками безопасности,

контроль доступа, аутентификацию, анализ безопасности, обнаружение компьютерных атак, брандмауэры, защиту от вирусов и сетевых червей, анализ и проверку протоколов безопасности и систем информационной безопасности, защиту программного обеспечения от взлома и управление цифровыми правами, технологии моделирования и визуализации для борьбы с кибертерроризмом. Число научных публикаций — 1000. ivkote@comsec.spb.ru; 14-я линия В.О., 39, 199178, Санкт-Петербург, Россия; р.т.: +7(812)328-7181; факс: +7(812)328-4450.

Саенко Игорь Борисович — д-р техн. наук, профессор, главный научный сотрудник, лаборатория проблем компьютерной безопасности, Санкт-Петербургский Федеральный исследовательский центр Российской академии наук (СПб ФИЦ РАН). Область научных интересов: автоматизированные информационные системы, информационная безопасность, обработка и передача данных по каналам связи, теория моделирования и математическая статистика, теория информации. Число научных публикаций — 500. ibsaen@comsec.spb.ru; 14-я линия В.О., 39, 199178, Санкт-Петербург, Россия; р.т.: +7(812)328-7181; факс: +7(812)328-4450.

Лаута Олег Сергеевич — д-р техн. наук, доцент, профессор кафедры, кафедра комплексного обеспечения информационной безопасности, Государственный университет морского и речного флота имени адмирала С.О. Макарова. Область научных интересов: защита от компьютерных атак. Число научных публикаций — 200. laos-82@yandex.ru; улица Двинская, 5/7, 198035, Санкт-Петербург, Россия; р.т.: +7(911)842-0228.

Садовников Владимир Евгеньевич — аспирант лаборатории, лаборатория проблем компьютерной безопасности, Санкт-Петербургский Федеральный исследовательский центр Российской академии наук (СПб ФИЦ РАН). Область научных интересов: защита от компьютерных атак. Число научных публикаций — 25. bladimir1998@mail.ru; 14-я линия В.О., 39, 199178, Санкт-Петербург, Россия; р.т.: +7(812)328-7181; факс: +7(812)328-4450.

Ичетовкин Егор Андреевич — аспирант, лаборатория проблем компьютерной безопасности, Санкт-Петербургский Федеральный исследовательский центр Российской академии наук (СПб ФИЦ РАН). Область научных интересов: защита от компьютерных атак. Число научных публикаций — 7. egor.email@list.ru; 14-я линия В.О., 39, 199178, Санкт-Петербург, Россия; р.т.: +7(812)328-7181; факс: +7(812)328-4450.

Ли Вэй — профессор, колледж компьютерных наук и технологий, Харбинский инженерный университет. Область научных интересов: защита от компьютерных атак, машинное обучение, искусственный интеллект. Число научных публикаций — 200. wei.li@hrbeu.edu.cn; Nantong St., Nangang District, 145, 150001, Харбин, Китай; р.т.: 0086(0451)8251921.

Поддержка исследований. Исследование частично поддержано бюджетной темой FFZF-2025-0016.

А.А. Молдовян, Д.Н. Молдовян, А.А. Костина
**РАНДОМИЗАЦИЯ В ПОСТКВАНТОВЫХ АЛГОРИТМАХ ЭЦП
С СЕКРЕТНОЙ ГРУППОЙ**

Молдовян А.А., Молдовян Д.Н., Костина А.А. Рандомизация в постквантовых алгоритмах ЭЦП с секретной группой.

Аннотация. В области постквантовой двухключевой криптографии значительный интерес представляет разработка практических алгебраических схем электронной цифровой подписи (ЭЦП) с секретной группой, стойкость которых базируется на вычислительной сложности решения систем степенных уравнений с многими неизвестными. В качестве алгебраического носителя в таких криптосхемах используются ассоциативные некоммутативные конечные алгебры (АНКА). Специфическим моментом, связанным с обеспечением стойкости к атакам на основе известных подписей при разработке схем ЭЦП данного типа, является наличие проблемы совершенствования механизма рандомизации подгоночного элемента цифровой подписи, представляющего собой вектор S , многократно входящий в уравнение верификации в качестве множителя. Известное решение этой проблемы на основе вычисления S в зависимости от двух векторов, выбираемых случайным образом из двух коммутативных секретных групп, таких, что элементы одной из них некоммутативны с элементами другой, при использовании в проверочном уравнении значения хеш-функции от S приводит к увеличению размера ЭЦП, обусловленного необходимостью задания двух вспомогательных подгоночных элементов подписи. В статье предлагается новый механизм рандомизации подписи, отличающийся вычислением значения S в зависимости от двух векторов, выбираемых случайным образом из одной коммутативной секретной группы. Предложенный механизм обеспечивает приемлемый уровень стойкости к атакам на основе известных подписей и может быть положен в основу разработки схем ЭЦП с одним уравнением верификации и одним вспомогательным подгоночным элементом подписи. Разработан новый алгебраический алгоритм ЭЦП, представляющий интерес как прототип практического постквантового стандарта ЭЦП. Даны оценки его параметров при реализации на четырехмерных АНКА и на алгебре матриц 3×3 , заданных над простым конечным полем.

Ключевые слова: постквантовая криптография, многомерная криптография, алгоритм ЭЦП, рандомизация ЭЦП, криптоалгоритм на конечных алгебрах, криптосхема на некоммутативных алгебрах, секретная группа.

1. Введение. Исследования в области разработки двухключевых криптографических алгоритмов, стойких к криптоанализу с использованием квантовых компьютеров, имеют высокую степень актуальности [1, 2]. Это обусловлено ожидаемым появлением в ближайшем будущем практически доступных квантовых вычислителей, для которых известны полиномиальные по времени алгоритмы решения как задачи факторизации (ЗФ), так и задачи дискретного логарифмирования (ЗДЛ) [3, 4]. Проблема разработки практических постквантовых криптографических алгоритмов является текущим вызовом в области прикладной криптографии, поскольку на

практике наиболее широко применяемые алгоритмы открытого согласования секретного ключа, открытого шифрования и электронной цифровой подписи (ЭЦП) основаны на ЗДЛ и/или ЗФ. Действительно, появление практически доступного многокубитного квантового вычислителя будет означать необходимость обновления арсенала используемых на практике криптографических двухключевых алгоритмов. Для обеспечения постквантовой стойкости разрабатываемые криптографические алгоритмы и протоколы должны базироваться на использовании вычислительной трудности задач, для которых неизвестны и предположительно не будут найдены эффективные (полиномиальные) алгоритмы их решения на квантовом компьютере. В этом направлении разработано большое число криптоалгоритмов, относящихся к различным типам, например, к криптосхемам на группах [5], на кодах, исправляющих ошибки [6, 7], на алгебраических решетках [8, 9], на трудно обратимых функциях [10] и на булевых функциях [11].

Значительное внимание со стороны криптографического сообщества уделяется разработке и криптоанализу постквантовых алгоритмов открытого шифрования, открытого согласования ключа и ЭЦП, базирующихся на трудно обратимых нелинейных отображениях с секретной лазейкой [12 – 15]. Стойкость алгоритмов данного типа определяется вычислительной сложностью решения больших систем степенных уравнений (БССУ), обычно квадратных и в отдельных случаях уравнений третьей и более высокой степени, с многими неизвестными. Это направление известно с 1988 г. как “multivariate cryptography” (далее будем использовать термин многомерная криптография) и в его рамках разработано большое число алгоритмов, которые достаточно детально проанализированы, а многие из них прошли несколько этапов совершенствования. Однако в известных разработках, относящихся к многомерной криптографии так и не был устранен существенный прикладной недостаток, связанный с непрактично большим размером открытого ключа (от десятков Кбайт до нескольких Мбайт для различных уровней стойкости от 2^{80} до 2^{256}). В частности, непрактичность алгоритмов ЭЦП с таким размером открытого ключа связана с его распечаткой на бумажном носителе, который подписывается создателем (владельцем) этого ключа и хранится в удостоверяющем центре.

Недавно предложенная в работах [16, 17] парадигма построения алгоритмов многомерной криптографии, использующая задание трудно обратимых отображений с секретной лазейкой как операции экспоненцирования в конечных векторных полях с секретной

модификацией, потенциально позволяет уменьшить размер открытого ключа в 10 раз и более. Однако, и в этом случае длина открытого ключа остается непрактично большой.

В статьях [18, 19] предложена концепция построения алгоритмов многомерной криптографии, относящихся к новому типу – алгебраическим алгоритмам ЭЦП с секретной группой, стойкость которых основана на вычислительной сложности решения БССУ. В криптоалгоритмах этого типа в качестве алгебраического носителя используются ассоциативные некоммутативные конечные алгебры (АНКА), включая как частный случай алгебры матриц. Достоинством постквантовых алгоритмов ЭЦП данного типа является сочетание сравнительно малых размеров открытого ключа и цифровой подписи по сравнению с известными. Их особенностью является использование подгоночного элемента подписи в виде вектора S , используемого в качестве множителя, входящего в проверочное уравнение два и более раз. При этом вектор S вычисляется в зависимости от предварительно генерируемого рандомизирующего элемента подписи e .

В отличие от рандомизированных алгоритмов ЭЦП, стойкость которых основана на вычислительной трудности решения ЗДЛ, в алгоритмах ЭЦП с секретной группой при фиксированном секретном ключе подгоночный элемент потенциально принимает значения из существенно ограниченного подмножества элементов алгебраического носителя. Несмотря на уникальность упомянутого подмножества, возникают предпосылки для реализации атак на основе известных подписей, которые состоят в вычислении элементов секретного ключа по некоторому набору известных подписей. Действительно, в работе [20] было показано, что механизм рандомизации элемента S , используемый в алгоритмах [18, 19], является недостаточным для обеспечения стойкости к указанной атаке по причине вычисления значения S в зависимости от случайного вектора, выбираемого из ограниченного подмножества элементов АНКА (а именно, из секретной группы, порядок которой существенно меньше порядка алгебраического носителя).

Предложенный в [20] способ совершенствования механизма рандомизации ЭЦП требует использования удвоенного уравнения верификации подписи. Последнее обуславливает значительное снижение производительности процедуры проверки подлинности ЭЦП, а также появление предпосылок специальных атак типа подделка подписи. Способ совершенствования механизма рандомизации ЭЦП, дающий возможность использовать в процедуре верификации только одно проверочное уравнение с многократным

вхождением вектора $\mathbf{S}$ разработан в статье [21], в которой вычисление подгоночного элемента подписи $\mathbf{S}$ задается в зависимости от двух взаимно некоммутативных векторов, выбираемых случайным образом из двух секретных коммутативных групп. В работах [21, 22] рассмотрены перспективные варианты проверочных уравнений, в которых используется дополнительный прием предотвращения атак на основе решения проверочного уравнения относительно неизвестного вектора $\mathbf{S}$, состоящий в задании в проверочном уравнении операции возведения в степень $\rho = \Phi(\mathbf{S})$, где $\Phi(\cdot)$ – некоторая специфицированная хеш-функция. Этот прием представляется эффективным для обозначенной цели, однако в алгоритмах ЭЦП с двумя секретными группами требуется использования двух дополнительных подгоночных элементов подписи, что приводит к увеличению размера ЭЦП.

В настоящей статье предлагается новый механизм рандомизации в алгебраических алгоритмах ЭЦП с одной секретной коммутативной группой и одним проверочным уравнением, обеспечивающий возможность использовать только один дополнительный подгоночный элемент подписи в схемах ЭЦП с проверочным уравнением, включающем значение хеш-функции $\Phi(\mathbf{S})$ в качестве дополнительного механизма защиты от атак, состоящим в решении проверочного уравнения относительно неизвестного вектора $\mathbf{S}$. Новым в разработанном способе является то, что вычисление вектора $\mathbf{S}$ выполняется в зависимости от случайного выбора двух векторов $\mathbf{G}$ и $\mathbf{H}$ из секретной коммутативной группы, которые служат левым и правым множителями относительно некоторого фиксированного секретного вектора $\mathbf{V}$, не принадлежащего секретной группе и некоммутативного с $\mathbf{G}$ и $\mathbf{H}$. Благодаря использованию одной секретной коммутативной группы, обеспечивается возможность использования только одного дополнительного подгоночного элемента подписи, что приводит к уменьшению размера ЭЦП при заданном уровне стойкости.

2. Алгебраические носители алгоритма. Для постквантовых алгоритмов ЭЦП с векторным проверочным уравнением, включающим подгоночный элемент ЭЦП $\mathbf{S}$ в качестве множителя, принципиальным является обеспечение стойкости к атакам, основанным на решении указанного уравнения относительно неизвестного вектора $\mathbf{S}$. Для реализации этого требования в работах [21, 22] используется многократное вхождение $\mathbf{S}$ в уравнение верификации, причем таким образом, что некоторая группа множителей, включающая $\mathbf{S}$, возводится в достаточно большую степень, благодаря чему сведение

решения проверочного уравнения к решению системы скалярных степенных уравнений с координатами $\mathbf{S}$ в качестве неизвестных вычислительно неосуществимо.

Этот прием обуславливает наличие в уравнении верификации ЭЦП операций возведения в степень достаточно большого размера, что приводит к необходимости использования алгоритма быстрого возведения в степень, корректность работы которого базируется на свойстве ассоциативности операции умножения. Таким образом, в качестве носителя алгоритмов ЭЦП, основанных на вычислительной трудности решения БССУ, следует рассматривать конечные алгебраические структуры с ассоциативным умножением, в частности АНКА, в том числе, конечные алгебры матриц.

Для формирования АНКА интерес представляет способ задания в конечном векторном пространстве дополнительной операции – операции векторного умножения, обладающей свойствами замкнутости, ассоциативности и дистрибутивности слева и справа относительно операции сложения векторов. Например, операция умножения может быть определена следующим образом. Пусть конечное m -мерное векторное пространство задано над полем $GF(p)$, где p – простое число. Векторы $\mathbf{V}$ этого пространства записываются 1) в виде упорядоченного набора значений из $GF(p)$: $\mathbf{V} = (v_0, v_1, \dots, v_{m-1})$, где $v_i \in GF(p)$ – координаты вектора; или 2) в виде суммы его компонент: $\mathbf{V} = \sum_{i=0}^{m-1} v_i \mathbf{e}_i$, где $\mathbf{e}_i$ – базисные векторы; $v_i \mathbf{e}_i$ – компоненты вектора. Естественным представляется задание результата операции умножения векторов $\mathbf{V}$ и $\mathbf{U} = \sum_{j=0}^{m-1} u_j \mathbf{e}_j$ по правилу перемножения каждой компоненты вектора $\mathbf{V}$ с каждой компонентой вектора $\mathbf{U}$, а именно, по следующей формуле:

$$\mathbf{V}\mathbf{U} = \sum_{i,j=0}^{m-1} v_i u_j (\mathbf{e}_i \mathbf{e}_j), \quad (1)$$

в которой требуется выполнить замену каждого произведения вида $\mathbf{e}_i \mathbf{e}_j$ некоторым базисным вектором или однокомпонентным вектором $\lambda \mathbf{e}_k$ (значение $\lambda \neq 1$ называется структурной константой). Аналогичным путем может быть задана конечная алгебра над различными расширениями простого поля $GF(p)$, в том числе над полями $GF(2^s)$, элементами которых являются двоичные многочлены степени не выше $s-1$, при различных степенях расширения s . Для разработки

алгебраических алгоритмов ЭЦП с секретной группой наиболее интересными для использования в качестве алгебраических носителей представляются АНКА, заданные над простыми полями $GF(p)$ и над полями характеристики два, т. е. расширениями двоичного поля $GF(2)$.

Для выполнения m^2 замен в формуле (1) составляется некоторая таблица умножения базисных векторов (ТУБВ). Обычно полагается, что левый множитель в произведении $e_i e_j$ указывает строку, а правый – столбец, на пересечении которых имеем ячейку, содержащую нужный для выполнения замены базисный вектор e_k или однокомпонентный вектор λe_i . На самом деле в ячейках ТУБВ можно записывать произвольные многокомпонентные векторы. Формула (1) обеспечивает свойства замкнутости и дистрибутивности умножения относительно операции сложения векторов и в этом случае. Однако задача составления ТУБВ, задающей ассоциативное некоммутативное умножение векторов, значительно упрощается, если в ячейках ТУБВ указывать только базисные и/или однокомпонентные векторы.

Известны унифицированные способы задания АНКА произвольных четных размерностей $m \geq 6$ [23] и $m \geq 2$ [24]. Алгебра квадратных матриц размерности $n \times n$, заданных над конечными полями (например, над простым полем $GF(p)$ или над расширением двоичного поля $GF(2^n)$), может быть рассмотрена как АНКА размерности $m = n^2$ с умножением, определенным по прореженным ТУБВ специального вида, составленным в соответствии с правилами матричного умножения. Для случаев алгебр матриц 2×2 и 3×3 такие ТУБВ показаны как таблицы 1 и 2 соответственно. Для разработки алгебраических алгоритмов ЭЦП с секретной группой, стойкость которых основана на вычислительной сложности решения БССУ, существенный интерес имеет также использование конечных алгебр матриц размеров 5×5 и 7×7 в качестве алгебраического носителя. Потенциально последние два варианта могут обеспечить более высокий уровень стойкости при приемлемом снижении производительности процедур генерации и верификации ЭЦП. Однако применение последних двух вариантов алгебраического носителя в данной работе не рассматривается.

Задание АНКА по прореженным (с большой долей ячеек с нулевой структурной константой) ТУБВ представляет интерес, поскольку при фиксированном значении размерности вычислительная сложность операции векторного умножения существенно меньше по сравнению со случаем использования ТУБВ без нулевых структурных констант.

Таблица 1. Трактовка умножения матриц $2 \times 2 \parallel a_{ij} \parallel$ как умножения четырехмерных векторов $A = (a_0, a_1, a_2, a_3)$, где $a_0 = a_{11}; a_1 = a_{12}; a_2 = a_{21}; a_3 = a_{22}$

$\times$	e_0	e_1	e_2	e_3
e_0	e_0	e_1	0	0
e_1	0	0	e_0	e_1
e_2	e_2	e_3	0	0
e_3	0	0	e_2	e_3

Таблица 2. Трактовка умножения матриц $3 \times 3 \parallel a_{ij} \parallel$ как умножения девятимерных векторов $A = (a_0, a_1, \dots, a_8)$, где $a_0 = a_{11}; a_1 = a_{12}; a_2 = a_{13}; a_3 = a_{21}; a_4 = a_{22}; a_5 = a_{23}; a_6 = a_{31}; a_7 = a_{32}; a_8 = a_{33}$

$\times$	e_0	e_1	e_2	e_3	e_4	e_5	e_6	e_7	e_8
e_0	e_0	e_1	e_2	0	0	0	0	0	0
e_1	0	0	0	e_0	e_1	e_2	0	0	0
e_2	0	0	0	0	0	0	e_0	e_1	e_2
e_3	e_3	e_4	e_5	0	0	0	0	0	0
e_4	0	0	0	e_3	e_4	e_5	0	0	0
e_5	0	0	0	0	0	0	e_3	e_4	e_5
e_6	e_6	e_7	e_8	0	0	0	0	0	0
e_7	0	0	0	e_6	e_7	e_8	0	0	0
e_8	0	0	0	0	0	0	e_6	e_7	e_8

В алгоритмах ЭЦП со скрытой (секретной) группой секретный и открытый ключи генерируются в виде наборов обратимых векторов, в которых каждый элемент открытого ключа вычисляется как произведение нескольких секретных векторов (элементов секретного ключа). Часть таких произведений включают в качестве множителей обратные значения последних. Это обстоятельство обуславливает использование (в качестве алгебраического носителя алгоритма ЭЦП) АНКА с глобальной двухсторонней единицей, обозначаемой как вектор E , вид которого определяется ТУБВ. Способ [23] позволяет задать АНКА с глобальной двухсторонней единицей для произвольных четных размерностей $m \geq 6$. Унифицированный способ [24] позволяет задать АНКА с множеством глобальных односторонних единиц для различных четных значений размерности и только для случая $m = 4$ найдена возможность включения в ТУБВ структурных констант, выбором значений которых обеспечивается наличие глобальной двухсторонней единицы (таблица 3).

Вычисление вектора V^{-1} , обратного по отношению к вектору V , осуществляется путем решения системы линейных скалярных уравнений, к которой сводится векторное уравнение $XV = E$ с неизвестным вектором X . Обычно в рассматриваемых алгебраических алгоритмах ЭЦП вычислительная сложность нахождения обратного вектора этим путем многократно меньше сложности операции возведения векторов в степень, выполняемой в АНКА.

С учетом определенной операции умножения векторов, предлагаемый способ рандомизации можно выразить следующей формулой:

$$S = DGVHF, \quad (2)$$

где G и H – векторы, выбираемые случайным образом из секретной коммутативной группы; D , V и F – фиксированные секретные векторы (некоммутативные с векторами G и H).

Таблица 3. Задание операции умножения в четырехмерных АНКА с глобальной двухсторонней единицей (структурные константы λ и ε удовлетворяют неравенству $\lambda\varepsilon \neq 1$)

$\times$	e_0	e_1	e_2	e_3
e_0	e_0	εe_3	εe_0	e_3
e_1	λe_2	e_1	e_2	λe_1
e_2	e_2	εe_1	εe_2	e_1
e_3	λe_0	e_3	e_0	λe_3

3. Строение четырехмерных АНКА заданных над полями четной характеристики. При разработке алгебраических алгоритмов ЭЦП с секретной коммутативной группой, а также при выполнении их криптоанализа, важным является знание декомпозиции АНКА на множество коммутативных подалгебр. Этот вопрос достаточно подробно изучен в статьях [25, 26] для различных типов четырехмерных АНКА, заданных над простыми полями $GF(p)$ нечетной характеристики, и показана общность строения таких алгебр. В связи с предполагаемым использованием четырехмерных АНКА, заданных по таблице 1 и таблице 3 над полями четной характеристики $GF(2^s)$, авторами было изучено их строение с точки зрения декомпозиции на коммутативные подалгебры. При этом была использована методика изучения, апробированная в [25, 26]. В результате выполненного исследования была установлена общность

строения для изученных двух случаев, которая характеризуется следующими основными моментами:

1. Четырехмерная АНКА, определенная над полем $GF(2^s)$, включает $\eta = 2^{2s} + 2^s + 1$ коммутативных подалгебр, каждая из которых имеет порядок, равный 2^{2s} , и включает все скалярные векторы (т.е. векторы вида $\mathbf{L} = \alpha \mathbf{E}$, где $\alpha \in GF(2^s)$).

2. Существуют ровно три типа коммутативных подалгебр порядка 2^{2s} :

2.1. Подалгебры первого типа, содержащие мультипликативную группу, имеющую циклическое строение и порядок, равный $\Omega_1 = 2^{2s} - 1$. Число подалгебр первого типа равно $\eta_1 = 2^{s-1}(2^s - 1)$.

2.2. Подалгебры второго типа, включающие мультипликативную группу порядка $\Omega_2 = (2^s - 1)^2$, которая порождается минимальной системой образующих, включающей два вектора порядка, равного $2^s - 1$. Число подалгебр второго типа равно $\eta_2 = 2^{s-1}(2^s + 1)$.

2.3. Подалгебры третьего типа, содержащие мультипликативную группу, имеющую циклическое строение и порядок, равный $\Omega_3 = 2^s(2^s - 1)$. Число подалгебр третьего типа равно $\eta_3 = 2^s + 1$.

3. Координаты каждого из векторов $\mathbf{V} = (v_0, v_1, v_2, v_3)$ заданной подалгебры могут быть выражены через координаты любого нескаларного вектора $\mathbf{C} = (c_0, c_1, c_2, c_3)$, принадлежащего заданной подалгебре (ее представителя), и уникальную пару скалярных переменных $d, h \in GF(p)$. Вид выражения, описывающего координаты v_0, v_1, v_2 и v_3 зависит от ТУБВ, по которой задается АНКА, и от типа заданной подалгебры. В случае АНКА, заданной по таблице 1, для подалгебр первого и второго типов получена следующая формула:

$$\mathbf{V} = (v_0, v_1, v_2, v_3) = (d, h, c_2 c_1^{-1} h, d + (c_3 + c_0) c_1^{-1} h). \quad (3)$$

Возможность описания коммутативной подалгебры по координатам ее представителя и две скалярные переменные имеет существенное значение при выполнении криптоанализа алгебраических алгоритмов ЭЦП, основанных на вычислительной сложности решения БССУ.

С учетом применения АНКА, заданных над полями $GF(2^s)$, в качестве алгебраического носителя разрабатываемого алгоритма ЭЦП и полученных результатов можно заметить, что степень расширения s следует выбирать с учетом получения минимального числа простых

делителей малого размера для чисел $2^s - 1$ и $2^s + 1$. Действительно последние два числа задают возможные порядки ω элементов рассматриваемых алгебр, в том числе и векторов, содержащихся в секретной коммутативной группе, а при вычислении подписи в типовых алгебраических алгоритмах ЭЦП имеет место ситуация, когда требуется вычислить обратное значение по модулю ω для случайных чисел. Последнее возможно только для чисел взаимно простых с ω . Если условие взаимной простоты не будет выполняться, то ряд шагов процедуры генерации подписи потребуется повторить при других значениях параметров рандомизации. Вероятность таких повторов может быть снижена до приемлемо малого значения, если задавать простое значение ω или значение ω , содержащее только делители достаточно большого размера (более 10 бит). В таблице 4 представлены приемлемые значения степени расширения полей $GF(2^s)$.

С учетом строения четырехмерных АНКА, заданных над полем $GF(2^s)$, и разложений чисел $2^s - 1$ и $2^s + 1$ (для всех приемлемых степеней расширения s число $2^s + 1$ делится на 3) видно, что в качестве секретной коммутативной группы целесообразно использовать мультипликативную группу подалгебры первого или второго типов. Для первого случая генерируется случайный вектор $\mathbf{J}_1$ порядка $\omega_1 = (2^{2s} - 1)/3$, выполняя следующие три шага:

1. Сгенерировать случайный вектор $\mathbf{V}$ и вычислить вектор $\mathbf{W} = \mathbf{V}^3$.
2. Для каждого простого делителя d числа ω_1 вычислить натуральное число $\sigma = \omega_1/d$ и проверить выполнимость неравенства $\mathbf{W}^\sigma \neq \mathbf{E}$. Если хотя бы для одного нетривиального делителя d имеет место равенство $\mathbf{W}^\sigma = \mathbf{E}$, то перейти к шагу 1.
3. Взять вектор $\mathbf{W}$ в качестве вектора $\mathbf{J}_1$, имеющего порядок ω_1 .

С учетом того, что число подалгебр первого типа равно $\eta_1 = 2^{s-1}(2^s - 1)$, легко установить, что цикл, включающий первые два шага, в среднем будет выполняться ≈ 2 раза. Выбор случайных векторов $\mathbf{G}$ из секретной группы осуществляется путем генерации случайного натурального числа $k < \omega_1$ и вычисления вектора $\mathbf{G} = \mathbf{J}_1^k$.

Чтобы задать в качестве секретной коммутативной группы мультипликативную группу подалгебры второго типа, генерируется нескаллярный случайный вектор $\mathbf{J}_2$ порядка $\omega_2 = 2^s - 1$, выполняя следующие три шага:

1. Сгенерировать случайный нескаллярный вектор $\mathbf{V}$.

2. Для каждого простого делителя d числа ω_2 вычислить натуральное число $\sigma = \omega_2/d$ и проверить выполнимость неравенства $V^\sigma \neq E$. Если хотя бы для одного нетривиального делителя d имеет место равенство $V^\sigma = E$, то перейти к шагу 1.

3. Взять вектор W в качестве вектора J_2 , имеющего порядок ω_2 .

С учетом того, что число подалгебр второго типа равно $\eta_2 = 2^{s-1}(2^s + 1)$, легко установить, что цикл, включающий первые два шага, в среднем будет выполняться ≈ 2 раза. Выбор случайных векторов G из секретной группы осуществляется путем генерации двоичного многочлена α , являющегося примитивным элементом поля $GF(2^s)$, и случайных натуральных чисел $t < \omega_2$ и $u < \omega_2$ с последующим вычислением вектора $G = \alpha' J_2^k$ (заметим, что фиксированное значение α может использоваться для случайного выбора многих различных векторов).

Таблица 4. Приемлемые значения степени расширения s для задания четырехмерных АНКА над полями $GF(2^s)$

Степень s	Число простых делителей (и их размер, бит)	
	для значения $2^s - 1$	для значения $(2^s + 1)/3$
61	1 (61)	1 (60)
67	2 (28 и 40)	2 (23 и 43)
71	3 (18, 26 и 28)	2 (26 и 44)
79	3 (12, 28 и 41)	1 (78)
101	2 (43 и 59)	1 (100)
103	2 (32 и 72)	2 (39 и 63)
109	2 (30 и 80)	2 (27 и 81)
127	1 (127)	1 (126)
137	2 (65 и 73)	4 (11, 14, 55 и 57)
139	2 (43 и 97)	2 (23 и 116)
149	2 (67 и 83)	4 (11, 20, 26 и 93)
181	4 (16, 21, 23 и 123)	3 (11, 25 и 144)
193	3 (24, 76 и 94)	4 (13, 26, 71 и 84)
197	2 (13 и 185)	2 (98 и 98)
199	2 (38 и 162)	1 (198)
211	3 (14, 66 и 132)	5 (13, 24, 33, 38 и 105)
223	6 (15, 18, 21, 22, 71 и 79)	2 (38 и 184)
227	2 (55 и 173)	3 (19, 62 и 146)
229	4 (21, 25, 56 и 129)	2 (25 и 204)
241	2 (25 и 217)	3 (12, 88 и 142)
257	3 (49, 80 и 129)	2 (69 и 187)

4. Атака на основе известных подписей. Алгебраические алгоритмы ЭЦП, стойкость которых базируется на вычислительной сложности решения БССУ, относятся к недетерминированным схемам ЭЦП. Используемый в них механизм рандомизации подписи оставляет принципиальную возможность вычисления элементов секретного ключа, входящих в формулу для вычисления подгоночного элемента подписи S . Это означает, что при разработке алгоритмов такого типа требуется выполнить оценку стойкости к атакам на основе известных подписей. Если вычислительная сложность такой атаки достаточно высока, то рандомизацию ЭЦП можно считать достаточно полной [22]. Способ выполнения оценки сложности атаки на основе известных подписей рассматривается в работе [22]. Применяя этот способ к алгоритму ЭЦП, использующему рандомизацию подписи по формуле (2), получаем следующую модель атаки.

При наличии z известных подписей имеем z векторных уравнений, в каждом из которых присутствуют фиксированные неизвестные векторные значения D , V и F (элементы секретного ключа) и уникальные неизвестные G и H , выбираемые случайным образом из коммутативной секретной группы. Тот факт, что выбор уникальных неизвестных осуществляется из существенно ограниченного множества элементов АНКА, используемой в качестве алгебраического носителя, обуславливает недостаточность рандомизации элемента подписи S . При наличии некоторого набора известных подписей вычисление фиксированных секретных векторов D , V и F может быть в принципе выполнено путем сведения решения системы из z векторных степенных уравнений к решению системы из $4z$ скалярных степенных уравнений. При этом выбор каждого из случайных векторов G и H потенциально может быть описан через $\mu < m$ скалярных неизвестных (уникальных для каждой известной подписи), если задан некоторый нескаллярный вектор C из секретной коммутативной группы.

Пусть имеются известные подписи с подгоночными элементами S_i , где $i = 1, 2, \dots, z$, причем S_i вычислено по паре уникальных значений G_i и H_i из секретной группы. Рассматривая вектор G_1 в качестве представителя секретной группы каждый из случайных векторов G_i для $i = 2, 3, \dots, z$ и векторов H_i для $i = 1, 2, \dots, z$ может быть описан через μ скалярных неизвестных. С учетом этого, по z известным подписям можно записать zm скалярных уравнений с $4m$ фиксированными скалярными неизвестными (которыми являются координаты секретных векторов D , V , F и координаты неизвестного представителя секретной группы G_1) и $\mu + 2\mu(z - 1)$ уникальными

скалярными неизвестными. С увеличением значения z число уравнений растет быстрее, чем число неизвестных. Натуральное значение z_0 , удовлетворяющее уравнению

$$z_0 m = 4m + \mu + 2\mu(z_0 - 1), \quad (4)$$

может быть принято за число необходимых для выполнения атаки известных подписей, а сложность решения системы из $z_0 m$ степенных уравнений в поле $GF(p)$ с $z_0 m$ неизвестными – за стойкость алгоритма ЭЦП с механизмом рандомизации по формуле (2).

Значение μ зависит от размерности используемого алгебраического носителя и от декомпозиции последнего на коммутативные подалгебры. Строение ряда четырехмерных АНКА с глобальной двухсторонней единицей (в том числе алгебры матриц 2×2), заданных над $GF(p)$, с точки зрения декомпозиции на коммутативные подалгебры порядка p^2 хорошо изучено, показана общность их строения и получены формулы, включающие две скалярные переменные и описывающие по координатам заданного представителя подалгебры все содержащиеся в ней векторы [25].

С учетом результатов выполненного исследования строения четырехмерных АНКА, заданных по таблице 1 и таблице 3 над полями четной характеристики $GF(2^5)$, легко видеть, что и в этом случае имеется возможность описать выбор вектора из заданной подалгебры по координатам представителя последней и двум скалярным переменным, принимающих значения в $GF(2^5)$.

Таким образом, при использовании в качестве алгебраического носителя конечных алгебр, заданных по таблицам 1 и 3, в рассмотренной модели атаки следует принять значение $\mu = 2$, при котором условие (4) не выполняется ни при каких значениях z_0 . С учетом существенной неоднозначности получаемых решений можно принять вывод, что при применении четырехмерных АНКА с глобальной двухсторонней единицей формула (2) обеспечивает достаточно полную рандомизацию подписи.

Строение девятимерной АНКА, заданной по таблице 2, может быть исследовано по способу, примененному в работе [25]. Это составляет задачу независимого рассмотрения, однако получение значения μ для этого случая представляется уместным. Все элементы коммутативной подалгебры, содержащей не скалярный девятимерный вектор C , могут быть найдены как решения системы из девяти скалярных линейных уравнений, соответствующей векторному уравнению $CX = XC$. Главный определитель системы скалярных

уравнений имеет ранг 6, т.е. пространство решений имеет размерность три и, следовательно, описывается по координатам вектора $\mathbf{C}$ формулой, включающей три скалярные переменные, т.е. имеем $\mu = 3$. При $m = 9$ и $\mu = 3$ уравнение (4) имеет решение $z_0 = 11$, что в рамках атаки на основе известных подписей соответствует решению системы из 99 степенных уравнений в поле $GF(p)$. Для лучших известных алгоритмов решения систем степенных уравнений оценка сложности решения этой задачи при числе уравнений, равном 99, составляет не менее 2^{256} даже в полях сравнительно малого порядка, а именно в полях $GF(256)$ [13]. Таким образом, при использовании конечной алгебры матриц 3×3 в качестве алгебраического носителя алгоритмы ЭЦП с рандомизацией подписи по формуле (2) также обеспечивают высокий уровень стойкости к атаке на основе известных подписей.

5. Разработанный алгоритм ЭЦП. Решение уравнений с многократным вхождением неизвестной, заданных в конечной некоммутативной алгебре может быть выполнено путем сведения к системе скалярных степенных уравнений, если неизвестный вектор (или некоторое выражение, куда он входит) возводится в сравнительно малую степень. В случае степени, имеющей разрядность 32 бит и более, это представляет собой вычислительно трудную задачу. Задание в проверочном уравнении хотя бы одной операции возведения в степень большой разрядности является приемом обеспечения стойкости к атакам, связанным с решением проверочного уравнения относительно подгоночного элемента подписи $\mathbf{S}$ как неизвестной. Другим приемом достижения такой цели является использование в проверочном уравнении операции возведения в степень, зависящую от значения $\mathbf{S}$ (например, вычисляемую как хеш-функция от $\mathbf{S}$), в результате чего проверочное уравнение приобретает вид экспоненциального уравнения.

В разработанном алгебраическом алгоритме ЭЦП используются оба указанных приема, обеспечивая высокую стойкость к указанным атакам. При этом предполагается использование алгебраических носителей различных типов:

1. четырехмерных АНКА, заданных по таблице 1 над простым полем $GF(p)$ при 128-битном простом порядке p ;
2. четырехмерных АНКА, заданных по таблице 1 или таблице 3 над расширением двоичного поля $GF(2^s)$ при степени $s = 127$, которая обеспечивает наличие в разложениях чисел $2^s + 1$ и $2^s - 1$ только достаточно больших простых делителей за исключением числа 3, которое является делителем числа $2^s + 1$ при натуральных

значениях s , соответствующих приемлемому разложению значения $2^s - 1$;

3. девятимерной АНКА (алгебры матриц 3×3), заданной по таблице 2 над простым полем $GF(p)$ при 64-битном простом порядке p .

При использовании четырехмерной АНКА, заданной над полем $GF(p)$, в качестве порядка поля используется простое число вида $p = 2q + 1$, где q – 127-битное простое число. Генерация такого числа p выполняется методом подбора 127-битных простых значений q до тех пор, пока вычисленное по указанной формуле число p не будет простым. В качестве секретной группы берется подгруппа мультипликативной группы коммутативной подалгебры третьего типа. Для задания секретной группы генерируется случайный вектор (элемент секретного ключа) $\mathbf{J}$, порядок которого равен числу $\omega = pq$, по следующему алгоритму:

1. Выбрать случайные элементы h и d поля $GF(p)$, такие, что h имеет порядок q , а d отличен от нуля и единицы.

2. Сгенерировать случайный вектор $\mathbf{V}$ и вычислить вектор $\mathbf{J} = \mathbf{V}\mathbf{J}'\mathbf{V}^{-1}$, где вектор $\mathbf{J}' = (h, d, 0, h)$ имеет порядок $\omega = pq$ (действительно, легко показать, что $\mathbf{J}'^\omega = \mathbf{E}, \mathbf{J}'^q \neq \mathbf{E}$ и $\mathbf{J}'^p \neq \mathbf{E}$).

3. Взять вектор $\mathbf{J}$ в качестве элемента секретного ключа.

При использовании четырехмерной АНКА, заданной над полем $GF(2^s)$, берется значение $s = 127$, при котором имеем:

$2^s - 1 = 170141183460469231731687303715884105727$ – 127-битное простое число;

$(2^s + 1)/3 = 56713727820156410577229101238628035243$ – 126-битное простое число.

В качестве секретной группы задается подгруппа мультипликативной группы коммутативной подалгебры первого типа. Для этого генерируется случайный вектор $\mathbf{J}$ порядка $\omega = (2^{2s} - 1)/3$, равного произведению двух простых чисел длины 126 и 127 бит. Вектор $\mathbf{J}$ используется как генератор секретной циклической группы.

При использовании в качестве алгебраического носителя девятимерной АНКА, заданной по таблице 2 над полем $GF(p)$, используется 64-битное простое вида $p = 2q + 1$, где q – простое число, при этом p таково, что число $r = p^2 + p + 1$ является простым. При этом генератором секретной группы служит вектор $\mathbf{J}$ порядка $\omega = r$. Примеры подходящих простых значений p представлены в таблице 5 (где знак «\» обозначает перенос записи числа в следующую строку). Простое p , соответствующее приведенным двум формулам, формируется путем многократной (в среднем не более 10^3 раз для разрядности p до 96 бит) генерации различных простых 63-битных

чисел q и проверкой на простоту чисел p и r , вычисляемых по двум указанным выше формулам.

Таблица 5. Примеры подходящих значений p (для различных значений разрядности) и соответствующие простые делители порядка ω

Значение p (длина в битах)	Простые делители ω (и их длина, бит)	
	q	r
959171755463 (40)	479585877731 (39)	9200104564789322\\ 42099833 (80)
25271244599\\ 5463 (48)	12635622299\\ 7731 (47)	6386358036101005597\\ 7462579833 (96)
36802238809\\ 418339 (56)	1840111940\\ 4709169 (55)	1354404781385457398\\ 845650318937261 (111)
133147932671\\ 28944783 (64)	1769732813521\\ 900658699 (63)	177283719746382279559\\ 337772146191861873 (128)
5021288074440\\ 707076923 (72)	2510644037220\\ 353538461 (71)	1252781692502456964370\\ 7613109378931741442601 (144)
12045013962879\\ 22323235223 (80)	60225069814396116 1617611 (79)	14508236136595544966\\ 356791484611805221437\\ 11094953 (160)
431629884442508\\ 29318249230143 (96)	21581494222125414 659124615071 (95)	1863043571438530627066\\ 078753338467297005900\\ 409930429030593 (191)
2669778488756826071 8840929163053897581 9 (128)	13348892443784\\ 130359420464581\\ 5269487909 (127)	712771717902868207632017149\\ 7774134208043365636618092703\\ 4615677976504005696581 (256)
8966932799355719\\ 59062298173795513\\ 824251385322239 (160)	4483466399677\\ 8597953114908\\ 68977569121256\\ 92661119 (159)	804058838281614017299\\ 346720056488863119472\\ 5801739743507509616055\\ 25138210914628357933\\ 457253295361 (319)

В разработанном алгоритме ЭЦП в качестве секретного ключа служат натуральные числа u, w, x, y, z , каждое из которых меньше значения ω , и набор попарно некоммутативных обратимых векторов A, B, D, F, J и V , генерируемых случайным образом, причем значение z является взаимно простым с ω и порядок вектора J равен ω . Открытый ключ представляет собой набор из девяти векторов ($U, Y, Z, T_1, T_2, T_3, T_4, T_5, T_6$), которые вычисляются по следующим формулам:

$$U = DJ^z D^{-1}, Y = AJA^{-1}, Z = B^{-1} J^{-1} B, \quad (5)$$

$$T_1 = AJ^u D^{-1}, T_2 = F^{-1} J^w V^{-1} A^{-1}, T_3 = F^{-1} J^x V^{-1} A^{-1}, \quad (6)$$

$$\mathbf{T}_4 = \mathbf{A}\mathbf{V}\mathbf{J}^w\mathbf{V}^{-1}\mathbf{B}, \mathbf{T}_5 = \mathbf{B}^{-1}\mathbf{V}\mathbf{J}^f\mathbf{F}, \mathbf{T}_6 = \mathbf{D}\mathbf{J}^{-u}\mathbf{B}. \quad (7)$$

Длина секретного (открытого) ключа составляет ≈ 540 (580) и ≈ 550 (650) байт в случае реализации алгоритма на четырехмерных и девятимерных АНКА соответственно.

Процедура формирования ЭЦП к документу M осуществляется с использованием секретного ключа и включает следующую последовательность шагов:

1. Сгенерировать случайные натуральные числа k и t ($1 < k < \omega$; $1 < t < \omega$) и вычислить рандомизирующий вектор (вектор-фиксатор)

$$\mathbf{R} = \mathbf{A}\mathbf{J}^k\mathbf{V}\mathbf{J}^t\mathbf{V}^{-1}\mathbf{B}. \quad (8)$$

2. Используя 512-битную коллизийно стойкую хеш-функцию Φ , вычислить рандомизирующий элемент подписи в виде 512-битного хеш-значения $e = e_1 || e_2 = \Phi(M || \mathbf{R})$, представленного в виде конкатенации двух 256-битных строк, представляющих числа e_1 и e_2 . Если наибольший общий делитель чисел $e_1 - e_2 + 1$ и ω не равен единице, то перейти к шагу 1 (вероятность этого пренебрежимо мала (менее 2^{-60}), поэтому эта теоретическая возможность практически не влияет на среднюю производительность процедуры вычисления ЭЦП).

3. Вычислить натуральные значения n и d :

$$n = -e_1 - u \bmod \omega, \quad d = \frac{t - ze_2 - xe_1 - w - y}{e_1 - e_2 + 1} \bmod \omega. \quad (9)$$

4. Вычислить подгоночный элемент подписи в виде вектора $\mathbf{S}$ по формуле

$$\mathbf{S} = \mathbf{D}\mathbf{J}^n\mathbf{V}\mathbf{J}^d\mathbf{F}. \quad (10)$$

5. Вычислить значение хеш-функции $\Phi(\mathbf{S})$ и натуральное число $\rho = \Phi(\mathbf{S}) \bmod \omega$.

6. Вычислить вспомогательный подгоночный элемент подписи в виде натурального числа σ по формуле

$$\sigma = z^{-1}(k - \rho - u - n) \bmod \omega. \quad (11)$$

7. Выдать тройку значений (e, σ, S) в качестве сгенерированной ЭЦП.

Вычислительная сложность процедуры генерации ЭЦП определяется главным образом четырьмя операциями экспоненцирования, выполняемыми в АНКА, используемой в качестве алгебраического носителя. В случае реализации алгоритма на четырехмерной и девятимерной АНКА указанная сложность равна ≈ 12300 и ≈ 15500 операций умножения в конечном поле, над которым задана АНКА, соответственно. Размер подписи равен 160 (168) байт в случае реализации алгоритма на четырехмерных (девятимерных) АНКА.

Процедура верификации ЭЦП (e, σ, S) к документу M выполняется с использованием открытого ключа по следующему алгоритму:

1. Вычислить значение $\rho = \Phi(S) \bmod \omega$ и контрольный вектор R' по формуле

$$R' = Y^{\rho} T_1 U^{\sigma} S T_2 \left(Y^{e_1} T_1 S T_3 \right)^{e_1} T_4 \left(T_5 S^{-1} T_6 Z^{e_1} \right)^{e_2}. \quad (12)$$

2. Вычислить значение хеш-функции $e' = \Phi(M \| R')$.

3. Если $e' = e$, то ЭЦП признается подлинной, в противном случае подпись отклоняется как ложная.

Вычислительная сложность процедуры верификации ЭЦП определяется шестью операциями возведения в степень, выполняемыми в АНКА, используемой в качестве алгебраического носителя. В случае реализации алгоритма на четырехмерной и девятимерной АНКА трудоемкость равна ≈ 18500 и ≈ 23300 операций умножения в конечном поле, над которым задана АНКА, соответственно.

Корректность работы предложенного алгоритма доказывается как факт того, что подпись (e, σ, S) , вычисленная в соответствии с процедурой генерации подписи, верифицируется как подлинная ЭЦП. Действительно, с учетом формул (5)–(11) и хорошо известного равенства $(XJX^{-1})^x = XJ^xX^{-1}$ из проверочного уравнения (12) получаем:

$$\begin{aligned}
 \mathbf{R}' &= (\mathbf{A}\mathbf{J}\mathbf{A}^{-1})^{\rho} \mathbf{A}\mathbf{J}^u \mathbf{D}^{-1} (\mathbf{D}\mathbf{J}^z \mathbf{D}^{-1})^{\sigma} (\mathbf{D}\mathbf{J}^n \mathbf{V}\mathbf{J}^d \mathbf{F}) \mathbf{F}^{-1} \mathbf{J}^y \mathbf{V}^{-1} \mathbf{A}^{-1} \times \\
 &\times \left[(\mathbf{A}\mathbf{J}\mathbf{A}^{-1})^{e_1} \mathbf{A}\mathbf{J}^u \mathbf{D}^{-1} (\mathbf{D}\mathbf{J}^n \mathbf{V}\mathbf{J}^d \mathbf{F}) \mathbf{F}^{-1} \mathbf{J}^x \mathbf{V}^{-1} \mathbf{A}^{-1} \right]^{e_1} \mathbf{A}\mathbf{V}\mathbf{J}^w \mathbf{V}^{-1} \mathbf{B} \times \\
 &\times \left[\mathbf{B}^{-1} \mathbf{V}\mathbf{J}^z \mathbf{F} (\mathbf{F}^{-1} \mathbf{J}^{-d} \mathbf{V}^{-1} \mathbf{J}^{-n} \mathbf{D}^{-1}) \mathbf{D}\mathbf{J}^{-u} \mathbf{B} (\mathbf{B}^{-1} \mathbf{J}^{-1} \mathbf{B})^{e_1} \right]^{e_2} = \\
 &= \mathbf{A}\mathbf{J}^{\rho+u+z\sigma+n} \mathbf{V}\mathbf{J}^{d+y} \mathbf{V}^{-1} \mathbf{A}^{-1} (\mathbf{A}\mathbf{J}^{e_1+u+n} \mathbf{V}\mathbf{J}^{d+x} \mathbf{V}^{-1} \mathbf{A}^{-1})^{e_1} \mathbf{A}\mathbf{V}\mathbf{J}^w \mathbf{V}^{-1} \mathbf{B} \times \\
 &\times (\mathbf{B}^{-1} \mathbf{V}\mathbf{J}^{z-d} \mathbf{V}^{-1} \mathbf{J}^{-n-u-e_1} \mathbf{B})^{e_2} = \mathbf{A}\mathbf{J}^{\rho+u+z\sigma+n} \mathbf{V}\mathbf{J}^{d+y} \mathbf{V}^{-1} \mathbf{A}^{-1} \times \\
 &\times (\mathbf{A}\mathbf{J}^0 \mathbf{V}\mathbf{J}^{d+x} \mathbf{V}^{-1} \mathbf{A}^{-1})^{e_1} \mathbf{A}\mathbf{V}\mathbf{J}^w \mathbf{V}^{-1} \mathbf{B} (\mathbf{B}^{-1} \mathbf{V}\mathbf{J}^{z-d} \mathbf{V}^{-1} \mathbf{J}^0 \mathbf{B})^{e_2} = \\
 &= \mathbf{A}\mathbf{J}^{\rho+u+z\sigma+n} \mathbf{V}\mathbf{J}^{d+y} \mathbf{V}^{-1} \mathbf{A}^{-1} (\mathbf{A}\mathbf{V}\mathbf{J}^{d+x} \mathbf{V}^{-1} \mathbf{A}^{-1})^{e_1} \mathbf{A}\mathbf{V}\mathbf{J}^w \mathbf{V}^{-1} \mathbf{B} \times \\
 &\times (\mathbf{B}^{-1} \mathbf{V}\mathbf{J}^{z-d} \mathbf{V}^{-1} \mathbf{B})^{e_2} = \\
 &= \mathbf{A}\mathbf{J}^{\rho+u+(k-\rho-u-n)+n} \mathbf{V}\mathbf{J}^{d+y+e_1(d+x)+w+e_2(z-d)} \mathbf{V}^{-1} \mathbf{B} = \\
 &= \mathbf{A}\mathbf{J}^k \mathbf{V}\mathbf{J}^{d(1+e_1-e_2)+y+e_1x+w+e_2z} \mathbf{V}^{-1} \mathbf{B} = \mathbf{A}\mathbf{J}^k \mathbf{V}\mathbf{J}^t \mathbf{V}^{-1} \mathbf{B} = \mathbf{R}.
 \end{aligned}$$

С учетом полученного равенства $\mathbf{R} = \mathbf{R}'$ имеем $e' = \Phi(M\|\mathbf{R}') = \Phi(M\|\mathbf{R}) = e$, т. е. корректно вычисленная подпись проходит процедуру верификации как подлинная ЭЦП.

6. Обсуждение. Постквантовая стойкость разработанного алгоритма ЭЦП обеспечивается тем, что она базируется на вычислительной сложности решения БССУ, для чего квантовый компьютер не является эффективным. Атака, состоящая в решении системы уравнений, связывающих координаты векторов, составляющих открытый ключ, с координатами секретных векторов, являющихся элементами секретного ключа, называется прямой атакой. Такая система записывается по формулам (5), (6) и (7), представляя каждое степенное векторное уравнение в виде системы из m скалярных уравнений. При $m = 4$ ($m = 9$) имеем 36 (81) скалярных уравнений второй и более высоких степеней.

С учетом того, что секретные векторы $\mathbf{A}$, $\mathbf{B}$, $\mathbf{D}$, $\mathbf{F}$, $\mathbf{J}$ и $\mathbf{V}$ в случае четырехмерных АНКА задают 24 скалярных неизвестных, а координаты каждого из неизвестных векторов $\mathbf{J}^u, \mathbf{J}^w, \mathbf{J}^x, \mathbf{J}^y$ и $\mathbf{J}^z$ могут быть представлен через координаты представителя секретной группы $\mathbf{J}$ и две уникальных скалярных неизвестных, то в системе из 36 степенных

уравнений имеются 34 скалярные неизвестные (система является переопределенной, но она имеет хотя бы одно решение по построению).

В случае девятимерных векторов координаты каждого из неизвестных векторов $\mathbf{J}^u, \mathbf{J}^w, \mathbf{J}^x, \mathbf{J}^y$ и $\mathbf{J}^z$ могут быть представлен через координаты представителя секретной группы $\mathbf{J}$ и три уникальных скалярных неизвестных, то в системе из 81 степенного уравнения имеются 69 скалярных неизвестных. Используя оценки трудоемкости решения больших систем степенных уравнений [13] при равенстве числа уравнений и неизвестных для случаев $m = 4$ и $m = 9$ получаем оценки уровня стойкости к прямой атаке $>2^{100}$ и $>2^{192}$ соответственно. Стойкость к атакам на основе известных подписей превышает уровень стойкости к прямой атаке, что позволяет сделать вывод о достаточной полноте рандомизации подписи, обеспечиваемой в разработанном алгоритме.

Подделка подписи к документу M' для разработанного алгебраического алгоритма ЭЦП теоретически может быть выполнена путем генерации случайного обратимого вектора $\mathbf{R}'$, выбора случайного значения σ и вычисления значения $e = e_1 || e_2 = \Phi(M || \mathbf{R}')$ с последующим решением проверочного уравнения относительно вектора $\mathbf{S}$ как неизвестной. Действительно, если решение проверочного уравнения будет найдено, то легко видеть, что подпись $(e, \sigma, \mathbf{S})$, сгенерированная таким путем пройдет проверочную процедуру как подлинная подпись. Однако в разработанном алгоритме используются два различных механизма задания вычислительной невозможности решения уравнения верификации относительно неизвестного значения подгоночного элемента подписи $\mathbf{S}$.

Если атаку по подделке подписи попытаться выполнить путем генерации случайного обратимого вектора $\mathbf{R}'$, выбора случайного вектора $\mathbf{S}$ и вычисления значений $\rho = \Phi(\mathbf{S}) \bmod \omega$ и $e = e_1 || e_2 = \Phi(M || \mathbf{R}')$ с последующим решением проверочного уравнения относительно натурального значения σ как неизвестной, то легко показать, что решение возникающего на последнем шаге уравнения существует с пренебрежимо малой вероятностью, не более 2^{-127} . Действительно, относительно неизвестной U^σ рассматриваемое уравнение имеет единственное решение $\mathbf{N}$, которое попадает в коммутативную подалгебру, содержащую элемент открытого ключа $\mathbf{U}$ с вероятностью $1/\eta_1 \approx 2^{-255}$ ($1/\eta_2 \approx 2^{-255}$), если $\mathbf{U}$ принадлежит подалгебре первого (второго) типа, или с вероятностью $1/\eta_3 \approx 2^{-127}$, если $\mathbf{U}$ принадлежит подалгебре третьего типа. При этом вероятность того, что при формировании открытого ключа будет реализовано последнее событие равна $p^{-1} \approx 2^{-128}$.

Предложенный механизм рандомизации ЭЦП, так же как и его аналог [21], обеспечивает уровень стойкости к атакам на основе известных подписей не менее уровня стойкости к прямым атакам. При одинаковой стойкости к прямым атакам, равной 2^{192} , вычислительная сложность процедура генерации (верификации) разработанного алгоритма ЭЦП равна ≈ 15500 (≈ 23300) умножений в поле $GF(p)$, что меньше соответствующих значений алгоритма ЭЦП [21], равных ≈ 36000 (≈ 55000) умножений в поле $GF(p)$. При этом размер ЭЦП в разработанном алгоритме составляет 168 байт, что на 19,2% меньше размера подписи, равного 208 байт, в алгоритме [21].

9. Заключение. Предложенный новый механизм рандомизации в алгебраических схемах ЭЦП с секретной группой, стойкость которых базируется на вычислительной сложности решения БССУ, обеспечивает стойкость к атаке на основе известных подписей, превышающую стойкость к прямой атаке (устранение основной предпосылки для потенциальной уязвимости к атаке на основе известных подписей) и позволяет уменьшить размер подписи для заданного уровня стойкости. На основе указанного способа разработан новый алгоритм ЭЦП, представляющий интерес как практическая постквантовая криптосхема. Элементы техники его построения представляют интерес для разработки прототипа алгебраического постквантового стандарта ЭЦП, относящегося к многомерной криптографии. Для повышения уровня стойкости представляет интерес реализация алгебраических алгоритмов данного типа на АНКА больших размерностей, в том числе на алгебрах матриц 5×5 и 7×7 , заданных над простыми конечными полями $GF(p)$ с 32-битным порядком p , а также над полями $GF(2^s)$ со значениями $s = 16-64$, однако это представляет задачу самостоятельного исследования.

Становление нового направления в области постквантовой криптографии, связанного с разработкой алгебраических алгоритмов цифровой подписи стало возможным благодаря вниманию и поддержке выдающегося ученого в области информатики, информационных технологий и теории управления, д.т.н, члена-корр. РАН, заслуженного деятеля науки и техники РФ Юсупова Рафаэля Мидхатовича многих инициативных начинаний сотрудников СПИИРАН, работающих в области криптографии. В сердцах авторов и его ближайших коллег благодарная память об этом останется навсегда.

Литература

1. Saarinen M.J., Smith-Tone D. Post-Quantum Cryptography. 15-th International Conference, PQCrypto 2024, Oxford, UK, 2024, Proceedings Part I // Lecture Notes in Computer Science. Springer, Cham, 2024. 434 p. DOI: 10.1007/978-3-031-62743-9.

2. Johansson T., Smith-Tone D. Post-Quantum Cryptography. 14-th International Conference, PQCrypto 2023, College Park, MD, USA, Proceedings // Lecture Notes in Computer Science. Springer, Cham, 2023. 714 p. DOI: 10.1007/978-3-031-40003-2.
3. Yan S.Y. Quantum Computational Number Theory. Springer. 2015. 252 p. DOI: 10.1007/978-3-319-25823-2.
4. Yan S.Y. Quantum Attacks on Public-Key Cryptosystems. Springer. 2014. 207 p. DOI: 10.1007/978-1-4419-7722-9.
5. Battarbee C., Kahrobaei D., Perret L., Shahandashti S.F. SPDH-Sign: Towards Efficient, Post-quantum Group-Based Signatures // Post-Quantum Cryptography. PQCrypto 2023. Lecture Notes in Computer Science. Springer, Cham, 2023. vol. 14154. DOI: 10.1007/978-3-031-40003-2_5.
6. Alamelou Q., Blazy O., Cauchie S., Gaborit P. A code-based group signature scheme // Designs, Codes and Cryptography. 2017. vol. 82. pp. 469–493. DOI: 10.1007/s10623-016-0276-6.
7. Kosolapov Y.V., Turchenko O.Y. On the construction of a semantically secure modification of the McEliece cryptosystem // Prikladnaya Diskretnaya Matematika. 2019. no. 45. pp. 33–43. DOI: 10.17223/20710410/45/4.
8. Gärtner J. NTWE: A Natural Combination of NTRU and LWE // Post-Quantum Cryptography. PQCrypto 2023. Lecture Notes in Computer Science, Springer, Cham, 2023. vol. 14154. DOI: 10.1007/978-3-031-40003-2_12.
9. Lysakov I.V. Solving some cryptanalytic problems for lattice-based cryptosystems with quantum annealing method // Matematicheskie Voprosy Kriptografii [Mathematical Aspects of Cryptography]. 2023. vol. 14. no. 2. pp. 111–122. DOI: 10.4213/mvk441.
10. Hamlin B., Song F. Quantum Security of Hash Functions and Property-Preservation of Iterated Hashing. Post-Quantum Cryptography. PQCrypto 2019. Lecture Notes in Computer Science. Springer, Cham, 2019. vol. 11505. pp. 329–349. DOI: 10.1007/978-3-030-25510-7_18.
11. Agibalov G.P. ElGamal cryptosystems on Boolean functions // Prikladnaya Diskretnaya Matematika. 2018. no. 42. pp. 57–65. DOI: 10.17223/20710410/42/4.
12. Shuaiting Q., Wenbao H., Yifa Li, Luyao J. Construction of Extended Multivariate Public Key Cryptosystems // International Journal of Network Security. 2016. vol. 18. no. 1. pp. 60–67.
13. Ding J., Petzoldt A. Current State of Multivariate Cryptography // IEEE Security and Privacy Magazine. 2017. vol. 15. no. 4. pp. 28–36.
14. Ding J., Petzoldt A., Schmidt D.S. Multivariate Cryptography // Multivariate Public Key Cryptosystems. Advances in Information Security. Springer, New York, NY, 2020. vol. 80. pp. 7–23. DOI: 10.1007/978-1-0716-0987-3_2.
15. Cartor R., Cartor M., Lewis M., Smith-Tone D. IPRainbow // Post-Quantum Cryptography. Lecture Notes in Computer Science. Springer, Cham, 2022. vol. 13512. pp. 170–184. DOI: 10.1007/978-3-031-17234-2_9.
16. Moldovyan N.A. Finite algebras in the design of multivariate cryptography algorithms // Bulletin of Academy of Sciences of Moldova. Mathematics. 2023. no. 3(103). pp. 80–89. DOI: 10.56415/basm.y2023.i3.p80.
17. Moldovyan N.A. Parameterized method for specifying vector finite fields of arbitrary dimensions // Quasigroups and related systems. 2024. vol. 32. no. 2. pp. 299–312. DOI: 10.56415/qrs.v32.21.
18. Moldovyan N.A. Algebraic signature algorithms with a hidden group, based on hardness of solving systems of quadratic equations // Quasigroups and Related Systems. 2022. vol. 30. no. 2(48). pp. 287–298. DOI: 10.56415/qrs.v30.24.
19. Moldovyan D.N. A new type of digital signature algorithms with a hidden group // Computer Science Journal of Moldova. 2023. vol. 31. no. 1(91). pp. 111–124. DOI: 10.56415/csjm.v31.06.

20. Молдовян А.А., Молдовян Д.Н., Молдовян Н.А. Постквантовые двухключевые криптосхемы на конечных алгебрах // Информатика и автоматизация. 2024. Т. 23. № 4. С. 1246–1276. DOI: 10.15622/ia.23.4.12.
21. Молдовян Н.А., Петренко А.С. Алгебраический алгоритм ЭЦП с двумя скрытыми группами // Вопросы кибербезопасности. 2024. № 6(64). С. 98–107. DOI: 10.21681/2311-3456-2024-6-98-107.
22. Молдовян Н.А., Петренко А.С. Типовые уравнения верификации в схемах ЭЦП с двумя скрытыми группами // Вопросы кибербезопасности. 2025. № 3(67). С. 45–54. DOI: 10.21681/2311-3456-2025-3-XX-XX. (в печати)
23. Moldovyan N.A. Unified Method for Defining Finite Associative Algebras of Arbitrary Even Dimensions // Quasigroups and Related Systems. 2018. vol. 26. no. 2. pp. 263–270.
24. Moldovyan N.A., Moldovyan A.A. Finite Non-commutative Associative Algebras as carriers of Hidden Discrete Logarithm Problem. Bulletin of the South Ural State University. Ser. Mathematical Modelling, Programming and Computer Software (Bulletin SUSU MMCS). 2019. vol. 12. no. 1. pp. 66–81. DOI: 10.14529/mmp190106.
25. Moldovyan D.N., Moldovyan A.A., Moldovyan N.A. Structure of a finite non-commutative algebra set by a sparse multiplication table // Quasigroups and Related Systems. 2022, vol. 30. no. 1. pp. 133–140. DOI: 10.56415/qrs.v30.11.
26. Duong M.T., Moldovyan A.A., Moldovyan D.N., Nguyen M.H., Do B.T. Structure of quaternion-type algebras and a post-quantum signature algorithm // International Journal of Electrical and Computer Engineering (IJECE). 2025. vol. 15. no. 3. pp. 2965–2976. DOI: 10.11591/ijece.v15i3.pp2965-2976.

Молдовян Александр Андреевич — д-р техн. наук, профессор, главный научный сотрудник, лаборатория проблем компьютерной безопасности, Федеральное государственное бюджетное учреждение науки «Санкт-Петербургский Федеральный исследовательский центр Российской академии наук» (СПб ФИЦ РАН). Область научных интересов: криптография, постквантовые криптоалгоритмы с открытым ключом, электронная цифровая подпись, криптографические протоколы, компьютерная безопасность. Число научных публикаций — 200. maa1305@yandex.ru; 14-я линия В.О., 39, 199178, Санкт-Петербург, Россия; р.т.: +7(812)328-7181.

Молдовян Дмитрий Николаевич — канд. техн. наук, доцент кафедры, кафедра информационных систем, Санкт-Петербургский государственный электротехнический университет «ЛЭТИ». Область научных интересов: криптография, постквантовые двухключевые криптоалгоритмы на алгебрах и нелинейных отображениях, цифровая подпись, информационная безопасность. Число научных публикаций — 100. mdn.spectr@mail.ru; улица Профессора Попова, 5, 197022, Санкт-Петербург, Россия; р.т.: +7(812)234-2772.

Костина Анна Александровна — научный сотрудник, лаборатория проблем компьютерной безопасности, Федеральное государственное бюджетное учреждение науки «Санкт-Петербургский Федеральный исследовательский центр Российской академии наук» (СПб ФИЦ РАН). Область научных интересов: криптосхемы с открытым ключом, электронная цифровая подпись, криптографические протоколы, информационная безопасность. Число научных публикаций — 40. to.ann@inbox.ru; 14-я линия В.О., 39, 199178, Санкт-Петербург, Россия; р.т.: +7(812)328-7181.

Поддержка исследований. Работа выполнена при финансовой поддержке РФН: проект № 24-21-00225.

A. MOLDOVYAN, D. MOLDOVYAN, A. KOSTINA
**RANDOMIZATION IN POST-QUANTUM DIGITAL SIGNATURE
ALGORITHMS WITH A SECRET GROUP**

Moldovyan A., Moldovyan D., Kostina A. Randomization in Post-Quantum Digital Signature Algorithms with a Secret Group.

Abstract. In the field of post-quantum public-key cryptography, a direction of development of practical algebraic signature algorithms with a secret group is of particular interest, the security of which is based on the computational difficulty of solving large systems of power equations. As an algebraic carrier of such algorithms, finite non-commutative associative algebras (FNAA) are used. A specific point related to ensuring security against attacks based on known signatures when developing digital signature schemes of this type is the presence of the problem of improving the randomization mechanism of the fitting element of the digital signature, which is a vector S , repeatedly included in the verification equation as a multiplier. A well-known solution to this problem based on the use of two commutative secret groups, such that the elements of one of them are non-commutative with the elements of the other, when using the hash function value from S in the verification equation leads to an increase in the size of the digital signature, due to the need to specify two auxiliary fitting signature elements. The article proposes a new mechanism for signature randomization, which is distinguished by calculating the value of S depending on two vectors randomly selected from one commutative secret group. The proposed mechanism provides an acceptable security level against attacks based on known signatures and can be used as a basis for developing digital signature schemes with one verification equation and one auxiliary fitting signature element. A new algebraic algorithm has been developed that is of interest as a prototype of a practical post-quantum digital signature standard. Estimates of its parameters are given when using FNAA and the algebra of 3×3 matrices defined over a finite field as an algebraic support.

Keywords: post-quantum cryptography, multivariate cryptography, digital signature algorithm, signature randomization, cryptoalgorithm on finite algebras, cryptoalgorithm on non-commutative algebras, secret group.

References

1. Saarinen M.J., Smith-Tone D. Post-Quantum Cryptography. 15-th International Conference, PQCrypto 2024, Oxford, UK, 2024, Proceedings Part I. Lecture Notes in Computer Science. Springer, Cham, 2024. 434 p. DOI: 10.1007/978-3-031-62743-9.
2. Johansson T., Smith-Tone D. Post-Quantum Cryptography. 14-th International Conference, PQCrypto 2023, College Park, MD, USA, Proceedings. Lecture Notes in Computer Science. Springer, Cham, 2023. 714 p. DOI: 10.1007/978-3-031-40003-2.
3. Yan S.Y. Quantum Computational Number Theory. Springer. 2015. 252 p. DOI: 10.1007/978-3-319-25823-2.
4. Yan S.Y. Quantum Attacks on Public-Key Cryptosystems. Springer. 2014. 207 p. DOI: 10.1007/978-1-4419-7722-9.
5. Battarbee C., Kahrobaei D., Perret L., Shahandashti S.F. SPDH-Sign: Towards Efficient, Post-quantum Group-Based Signatures. Post-Quantum Cryptography. PQCrypto 2023. Lecture Notes in Computer Science. Springer, Cham, 2023. vol. 14154. DOI: 10.1007/978-3-031-40003-2_5.
6. Alamelou Q., Blazy O., Cauchie S., Gaborit P. A code-based group signature scheme. Designs, Codes and Cryptography. 2017. vol. 82. pp. 469–493. DOI: 10.1007/s10623-016-0276-6.

7. Kosolapov Y.V., Turchenko O.Y. On the construction of a semantically secure modification of the McEliece cryptosystem. *Prikladnaya Diskretnaya Matematika*. 2019. no. 45. pp. 33–43. DOI: 10.17223/20710410/45/4.
8. Gärtner J. NTWE: A Natural Combination of NTRU and LWE. *Post-Quantum Cryptography. PQCrypto 2023. Lecture Notes in Computer Science*, Springer, Cham, 2023. vol. 14154. DOI: 10.1007/978-3-031-40003-2_12.
9. Lysakov I.V. Solving some cryptanalytic problems for lattice-based cryptosystems with quantum annealing method. *Matematicheskie Voprosy Kriptografii [Mathematical Aspects of Cryptography]*. 2023. vol. 14. no. 2. pp. 111–122. DOI: 10.4213/mvk441.
10. Hamlin B., Song F. Quantum Security of Hash Functions and Property-Preservation of Iterated Hashing. *Post-Quantum Cryptography. PQCrypto 2019. Lecture Notes in Computer Science*. Springer, Cham, 2019. vol. 11505. pp. 329–349. DOI: 10.1007/978-3-030-25510-7_18.
11. Agibalov G.P. ElGamal cryptosystems on Boolean functions. *Prikladnaya Diskretnaya Matematika*. 2018. no. 42. pp. 57–65. DOI: 10.17223/20710410/42/4.
12. Shuaiting Q., Wenbao H., Yifa Li, Luyao J. Construction of Extended Multivariate Public Key Cryptosystems. *International Journal of Network Security*. 2016. vol. 18. no. 1. pp. 60–67.
13. Ding J., Petzoldt A. Current State of Multivariate Cryptography. *IEEE Security and Privacy Magazine*. 2017. vol. 15. no. 4. pp. 28–36.
14. Ding J., Petzoldt A., Schmidt D.S. Multivariate Cryptography. *Multivariate Public Key Cryptosystems. Advances in Information Security*. Springer, New York, NY, 2020. vol. 80. pp. 7–23. DOI: 10.1007/978-1-0716-0987-3_2.
15. Cartor R., Cartor M., Lewis M., Smith-Tone D. IPRainbow. *Post-Quantum Cryptography. Lecture Notes in Computer Science*. Springer, Cham, 2022. vol. 13512. pp. 170–184. DOI: 10.1007/978-3-031-17234-2_9.
16. Moldovyan N.A. Finite algebras in the design of multivariate cryptography algorithms. *Bulletin of Academy of Sciences of Moldova. Mathematics*. 2023. no. 3(103). pp. 80–89. DOI: 10.56415/basm.y2023.i3.p80.
17. Moldovyan N.A. Parameterized method for specifying vector finite fields of arbitrary dimensions. *Quasigroups and related systems*. 2024. vol. 32. no. 2. pp. 299–312. DOI: 10.56415/qrs.v32.21.
18. Moldovyan N.A. Algebraic signature algorithms with a hidden group, based on hardness of solving systems of quadratic equations. *Quasigroups and Related Systems*. 2022. vol. 30. no. 2(48). pp. 287–298. DOI: 10.56415/qrs.v30.24.
19. Moldovyan D.N. A new type of digital signature algorithms with a hidden group. *Computer Science Journal of Moldova*. 2023. vol. 31. no. 1(91). pp. 111–124. DOI: 10.56415/csjm.v31.06.
20. Moldovyan A., Moldovyan D., Moldovyan N. [Post-quantum public-key cryptoschemes on finite algebras]. *Informatics and Automation*. 2024. vol. 23(4). pp. 1246–1276. DOI: 10.15622/ia.23.4.12. (In Russ.).
21. Moldovyan N.A., Petrenko A.S. [Algebraic signature algorithm with two hidden groups]. *Voprosy kiberbezopasnosti – Cibersecurity questions*. 2024. no. 6(64). pp. 98–107. DOI: 10.21681/2311-3456-2024-6-98-107. (In Russ.).
22. Moldovyan N.A., Petrenko A.S. [Typical verification equations in algebraic signature schemes with two hidden groups]. *Voprosy kiberbezopasnosti – Cibersecurity questions*. 2025. no 3(67). pp. 45–54. DOI: 10.21681/2311-3456-2025-3-XX-XX. (in print). (In Russ.).
23. Moldovyan N.A. Unified Method for Defining Finite Associative Algebras of Arbitrary Even Dimensions. *Quasigroups and Related Systems*. 2018. vol. 26. no. 2. pp. 263–270.

24. Moldovyan N.A., Moldovyan A.A. Finite Non-commutative Associative Algebras as carriers of Hidden Discrete Logarithm Problem. Bulletin of the South Ural State University. Ser. Mathematical Modelling, Programming and Computer Software (Bulletin SUSU MMCS). 2019. vol. 12. no. 1. pp. 66–81. DOI: 10.14529/mmp190106.
25. Moldovyan D.N., Moldovyan A.A., Moldovyan N.A. Structure of a finite non-commutative algebra set by a sparse multiplication table. Quasigroups and Related Systems. 2022, vol. 30. no. 1. pp. 133–140. DOI: 10.56415/qrs.v30.11.
26. Duong M.T., Moldovyan A.A., Moldovyan D.N., Nguyen M.H., Do B.T. Structure of quaternion-type algebras and a post-quantum signature algorithm. International Journal of Electrical and Computer Engineering (IJECE). 2025. vol. 15. no. 3. pp. 2965–2976. DOI: 10.11591/ijece.v15i3.pp2965-2976.

Moldovyan Alexander — Ph.D., Dr.Sci., Professor, Chief researcher, Laboratory of computer security problems, St. Petersburg Federal Research Center of the Russian Academy of Sciences (SPC RAS). Research interests: cryptography, post-quantum public-key cryptoalgorithms, digital signature, cryptographic protocols, computer security. The number of publications — 200. maa1305@yandex.ru; 39, 14-th Line V.O., 199178, St. Petersburg, Russia; office phone: +7(812)328-7181.

Moldovyan Dmitriy — Ph.D., Associate professor of the department, Department of information systems, Saint Petersburg Electrotechnical University «LETI». Research interests: cryptography, post-quantum public-key cryptoalgorithms on algebras and on non-linear mappings, digital signature, information security. The number of publications — 100. mdn.spectr@mail.ru; 5, Professor Popov St., 197022, St. Petersburg, Russia; office phone: +7(812)234-2772.

Kostina Anna — Researcher, Laboratory of computer security problems, St. Petersburg Federal Research Center of the Russian Academy of Sciences (SPC RAS). Research interests: public-key cryptoschemes, digital signature algorithms, cryptographic protocols, information security. The number of publications — 40. to.ann@inbox.ru; 39, 14-th Line V.O., 199178, St. Petersburg, Russia; office phone: +7(812)328-7181.

Acknowledgements. This research is supported by RSF (project No. 24-21-00225).

А.И. САВЕЛЬЕВ, Д.А. АНИКИН
**МЕТОД РАСЧЕТА КАПСУЛООБРАЗНЫХ ВОЗДУШНЫХ
КОРИДОРОВ БЕЗОПАСНЫХ МАРШРУТОВ ГРУППЫ
БЕСПИЛОТНЫХ ЛЕТАТЕЛЬНЫХ АППАРАТОВ**

Савельев А.И., Аникин Д.А. Метод расчета капсулообразных воздушных коридоров безопасных маршрутов группы беспилотных летательных аппаратов.

Аннотация. В работе рассматривается задача построения безопасных маршрутов группы беспилотных летательных аппаратов в ограниченном воздушном пространстве над сельскохозяйственным участком. Актуальность исследования обусловлена растущим применением групп БПЛА в агропромышленном комплексе для мониторинга, картографирования и обработки полей, что требует обеспечения безопасности полётов в условиях высокой плотности воздушного движения, ограниченной связи, а также воздействия внешних факторов. Особую сложность представляет необходимость автономного выполнения миссий при наличии навигационных погрешностей и природных воздействий. Предлагается метод планирования маршрутов, основанный на представлении траектории каждого аппарата в виде капсульного воздушного коридора – трёхмерного объёма фиксированного радиуса, сформированного вдоль отрезков траектории. Пространственное резервирование обеспечивает безопасное разнесение траекторий на этапе планирования, исключая конфликты при последующем автономном выполнении полётов без необходимости непрерывной координации между агентами. Радиус капсулы включает запас на возможные отклонения от запланированной траектории, что обеспечивает устойчивость к навигационным ошибкам. В основе метода лежит последовательное формирование маршрутов для каждого аппарата по четырёхфазной схеме, включающей вертикальный подъём от точки старта до рабочей высоты, горизонтальный переход к входу в зону обработки, возврат от выхода из зоны к точке начала снижения и вертикальный спуск к исходной позиции. Каждый новый маршрут строится с учётом уже зарезервированных воздушных коридоров через аналитическую проверку геометрических пересечений между капсулами различных траекторий и выпуклыми многогранниками зон обработки. Для повышения вычислительной эффективности применяется иерархическая пространственная фильтрация на основе ограничивающих параллелепипедов, позволяющая на предварительном этапе быстро отсекаать заведомо непересекающиеся объекты и выполнять точную геометрическую проверку только для потенциально конфликтующих сегментов маршрутов. Численные эксперименты проводились для групп от 2 до 32 аппаратов на типовом сельскохозяйственном участке площадью один квадратный километр. Установлено нелинейное возрастание времени планирования и количества итераций с увеличением числа агентов, что обусловлено необходимостью построения каждого последующего маршрута в уже частично занятом пространстве с возрастающим числом пространственных ограничений. Длина маршрутов демонстрирует тенденцию к росту, особенно выраженную на начальных этапах масштабирования, что связано с необходимостью обхода уже зарезервированных воздушных коридоров.

Ключевые слова: БПЛА, воздушные коридоры, сельскохозяйственные задачи, Rapidly-Exploring Random Tree*, Loose Octree, автономная навигация, групповое планирование.

1. Введение. Современное сельское хозяйство активно использует беспилотные летательные аппараты (БПЛА) для

мониторинга посевов [1, 2], картографирования местности [3, 4], агрохимической обработки [5], оценки урожайности и анализа состояния полей [6]. Чем больше дронов работает одновременно и чем сложнее становится организация воздушного движения, тем острее встает вопрос слаженного взаимодействия групп БпЛА. Когда множество аппаратов выполняют различные задачи в одном воздушном пространстве, критически важно обеспечить их пространственную координацию и предотвратить столкновения на пересекающихся маршрутах.

При выполнении сельскохозяйственных операций БпЛА работают в условиях плотной загрузки выделенного воздушного пространства, жестких временных рамок для выполнения задач и ограниченной вычислительной мощности бортовых систем [7]. Кроме того, в реальной эксплуатации на выполнение полётных заданий влияют внешние природные воздействия и внутренние технические отклонения, включая погрешности телеметрии, сбои навигационных систем, временные потери связи с наземной станцией, отказы отдельных аппаратов, а также порывистый ветер и неустойчивые погодные условия, способные отклонить фактическую траекторию от запланированной. Исходя из этого, маршруты полета должны обеспечивать безопасность движения и исключение конфликтов между аппаратами в условиях прерывистой связи с центром управления, одновременно компенсируя навигационные ошибки при возможных отклонениях от запланированного курса.

Существующие методы и алгоритмы многоагентного планирования обеспечивают различный уровень оптимальности, но, как правило, не адаптированы к требованиям и ограничениям практической эксплуатации в аграрной отрасли.

Алгоритм Conflict-Based Search (CBS) [8] и его производные формируют маршруты путём централизованного разрешения конфликтов между агентами в пространственно-временной области через построение дерева конфликтов. Хотя базовый CBS обеспечивает оптимальность решения, он предполагает дискретное пространственно-временное планирование и строгое следование фиксированным траекториям, что делает его непригодным для прямого применения в реальных сценариях с БпЛА [9]. Такие алгоритмы не предусматривают отклонений от запланированного поведения – при выходе одного БпЛА из строя или отклонении его траектории возникает риск несогласованности с действиями других агентов, что может потребовать полной переработки конфигурации маршрутов для сохранения безопасности. Модификация для работы с

непрерывным временем – Continuous-time CBS (CCBS) [10], также сталкивается с вычислительными ограничениями при масштабировании и усложняется необходимостью учёта физических характеристик и динамических параметров реальных летательных аппаратов. Алгоритм Multi-Conflict-Based Search (MCBS) [11] показывает, что при росте числа агентов и плотности взаимодействий структура дерева конфликтов значительно усложняется – количество возможных ветвлений растёт экспоненциально, что ограничивает применимость метода в сценариях с высокой насыщенностью полётов, особенно при необходимости учёта трёхмерного пространства и временных ограничений.

Алгоритм Multi-Agent Rapidly-Exploring Random Tree* (MA-RRT*) [12] осуществляют поиск траекторий в объединённом конфигурационном пространстве размерности $\mathbb{R}^3$, где m – число агентов. Базовый алгоритм Rapidly-Exploring Random Tree* (RRT*) [13] представляет собой стохастический подход выборочного построения деревьев, обеспечивающий асимптотическую сходимость к оптимальному решению. В многоагентной версии данная стратегия теоретически сохраняет глобальную сходимость, однако сталкивается с проклятием размерности – объём пространства поиска растёт экспоненциально с увеличением числа агентов. Кроме того, MA-RRT* предполагает точное выполнение заранее построенных траекторий и не учитывает внешние отклонения – например, сбой навигации, воздействие ветра или ошибки телеметрии. При отказе одного БПЛА или отклонении от плана нарушается согласованность всей конфигурации, что требует пересчёта маршрутов группы в целом.

Централизованные эвристические алгоритмы, такие как Cooperative A*, использующий приоритетное планирование и эвристики расстояния для координации агентов [14], и Enhanced Conflict-Based Search (ECBS), применяющий фокальный поиск и ограниченную субоптимальность для ускорения разрешения конфликтов [15], сокращают объем вычислений при планировании маршрутов. Данные подходы эффективны при фиксированном составе группы БПЛА и статичной среде, но не обладают устойчивостью к навигационным ошибкам и вынуждены инициировать глобальное перепланирование даже при незначительных отклонениях агентов от заданных траекторий.

Реактивные алгоритмы, такие как Optimal Reciprocal Collision Avoidance (ORCA) [16] и Reciprocal Velocity Obstacles (RVO) [17], обеспечивают локальное предотвращение столкновений через

корректировку векторов скорости в реальном времени без явной коммуникации между агентами. Однако данные подходы критически зависят от точности локализации и предполагают надёжное измерение скоростей окружающих агентов. Они особенно чувствительны к накапливающимся погрешностям навигации – даже небольшие ошибки в определении относительных положений и скоростей приводят к неэффективным или потенциально опасным траекториям. При временных сбоях в измерениях или задержках обновления телеметрии эти алгоритмы не имеют механизмов компенсации, что ограничивает их применение в условиях нестабильной связи и требует постоянной высокоточной локализации для безопасного функционирования.

Поведенческие и децентрализованные подходы, включая модели на основе потенциальных полей (Artificial Potential Field, APF) [18], алгоритмы имитации поведения роя (Boids, Cucker-Smale model) [19, 20], а также алгоритмы альтернативной коллективной адаптации [21] и роевого управления с элементами кластерного анализа [22] обладают высокой адаптивностью и подходят для работы в частично наблюдаемых или слабо структурированных средах. Однако такие методы, как правило, требуют тонкой эмпирической настройки параметров, их поведение сложно предсказать в новых условиях, а маршруты формируются в процессе исполнения, а не заранее. Это затрудняет проверку надёжности, воспроизводимости и безопасности движения до начала полёта.

С учётом изложенных ограничений предлагается альтернативный метод, в котором безопасность и устойчивость движения достигаются за счёт геометрически гарантированного резервирования пространства для каждого маршрута ещё до начала выполнения миссии. При этом траектория каждого БПЛА представляется в виде набора капсул, формирующих воздушный коридор – совокупность трёхмерных объёмов фиксированного радиуса, последовательно расположенных вдоль отрезков маршрута. Эти объёмы трактуются как зарезервированные безопасные зоны, исключающие пересечения с траекториями других агентов. Таким образом, безопасность полётов обеспечивается конструктивно, ещё на этапе планирования, без потребности в онлайн-координации или реактивном избегании конфликтов в процессе выполнения миссии.

Предложенная простая и строго геометрическая структура маршрутов группы БПЛА позволяет устранить ряд критичных для реального применения ограничений:

1. Нет необходимости в точной временной синхронизации: БПЛА могут стартовать независимо и выполнять задачи с различной продолжительностью.

2. Отсутствие зависимости от обмена данными в режиме реального времени с наземной станцией: все вычисления выполняются заранее, полёт проходит автономно.

3. Устойчивость к внешним воздействиям: выделенный безопасный объём формирует пространственный запас, позволяющий БПЛА сохранять корректность траектории даже при наличии внешних возмущений.

4. Отказоустойчивость: при потере связи или выходе из строя одного БПЛА конфликтные ситуации исключены, поскольку каждый агент движется в своём зарезервированном воздушном пространстве.

5. Простота реализации: метод не опирается на сложные динамические модели или ресурсоёмкие численные процедуры, а использует аналитическую геометрию.

Маршруты для агентов группы строятся на основе алгоритма RRT*[13] модифицированного для последовательного построения с учётом уже зарезервированных объёмов. Новый путь генерируется только в том случае, если он не пересекается с существующими коридорами. Для ускорения проверок используется структура данных «свободное октодерев» (Loose Octree) [23] – разновидность октодерева с расширенными ячейками, что позволяет эффективно индексировать пространственные объекты произвольного размера и выполнять локализованные запросы.

Построенные маршруты могут быть напрямую загружены в управляющее программное обеспечение БПЛА в виде контрольных точек. Это обеспечивает полную автономность миссии, исключает зависимость от связи в процессе выполнения и гарантирует воспроизводимость поведения.

Следует отметить, что предложенный метод не обеспечивает глобальной оптимальности маршрутов всей группы, поскольку маршруты строятся последовательно, с учётом уже зарезервированных пространственных объёмов. Для каждого агента используется алгоритм RRT*, который гарантирует асимптотическую сходимость к кратчайшему пути в рамках доступной области поиска. Поскольку эта область заранее ограничена ранее построенными маршрутами и зонами обработки, найденное решение оптимально лишь среди маршрутов, допустимых с точки зрения геометрических ограничений. Такой подход обеспечивает приемлемое качество отдельных

маршрутов при высокой вычислительной эффективности и устойчивости всей конфигурации.

Таким образом, метод объединяет пространственную независимость, отказоустойчивость, масштабируемость, автономность и простоту интеграции. Указанные характеристики обеспечивают его применимость в агропромышленной сфере, где критически важны надёжность исполнения, устойчивость к внешним воздействиям и эффективность в условиях ограниченных ресурсов на борту БПЛА.

2. Постановка задачи. Рассматривается задача построения безопасных маршрутов для группы из $m \in \mathbb{Z}$ БПЛА, выполняющих полёты к распределённым сельскохозяйственным участкам для проведения агротехнических работ. Каждый участок представлен в виде многогранника $A_j \subset \mathbb{R}^3$, $j = 1, \dots, m$, а воздушное пространство моделируется трёхмерной рабочей областью $\Omega \subset \mathbb{R}^3$.

Каждый БПЛА a_i , $i = 1, \dots, m$ стартует из точки p_1^i , выполняет миссию на выделенном участке и возвращается в точку $p_{n_i}^i$. Маршрут i -го БПЛА описывается как последовательность из $n \in \mathbb{Z}$ трёхмерных точек:

$$P_i = (p_1^i, p_2^i, \dots, p_{n_i-1}^i, p_{n_i}^i), \quad (1)$$

где n_i – количество точек маршрута i -го БПЛА.

Каждому отрезку маршрута $[p_k^i, p_{k+1}^i]$, $k = 1, \dots, n-1$ сопоставляется трехмерный объем в форме капсулы заданного радиуса R_i , объединение таких элементов определяет воздушный коридор C_i :

$$C_i = \bigcup_{k=1}^{n_i-1} \text{Capsule}(p_k^i, p_{k+1}^i, R_i), \quad (2)$$

где $\text{Capsule}(p_k^i, p_{k+1}^i, R_i)$ – объем в форме капсулы с радиусом R_i , образованный вокруг отрезка между точками p_k^i и p_{k+1}^i , моделирующий занимаемое воздушным коридором пространство; n_i – количество точек маршрута i -го БПЛА.

Сама капсула представляет собой множество точек, находящихся на расстоянии не более R_i от отрезка маршрута $[p_k^i, p_{k+1}^i]$ [24]:

$$\text{Capsule}(p_k^i, p_{k+1}^i, R_i) = \{x \in \mathbb{R}^3 : d(x, [p_k^i, p_{k+1}^i]) \leq R_i\}, \quad (3)$$

где $x \in \mathbb{R}^3$ – произвольная точка пространства; $d(x, [p_k^i, p_{k+1}^i])$ – расстояние от точки x до отрезка $[p_k^i, p_{k+1}^i]$.

Задача формируется как поиск такой конфигурации маршрутов группы БПЛА, при которой суммарная длина всех маршрутов минимальна, при этом соответствующие им воздушные коридоры не пересекаются между собой и не вторгаются в зоны активной обработки, закреплённые за другими агентами, а маршруты не выходят за пределы рабочего пространства. Математическая формализация данной задачи представляет собой оптимизацию с геометрическими ограничениями.

Целевая функция направлена на минимизацию суммарной длины всех маршрутов:

$$\min_{P_1, \dots, P_m} \sum_{i=1}^m \sum_{k=0}^{n_i-1} \|p_{k+1}^i - p_k^i\|_2, \quad (4)$$

где P_i – маршрут i -го БПЛА.

При этом на допустимые маршруты накладываются следующие пространственные ограничения:

Во-первых, воздушные коридоры различных БПЛА не должны пересекаться:

$$\forall i = 1, \dots, m, \forall j = 1, \dots, m, i \neq j \quad C_i \cap C_j = \emptyset. \quad (5)$$

Во-вторых, необходимо исключить попадание коридора одного БПЛА в активную зону обработки другого БПЛА:

$$\forall i = 1, \dots, m, \forall j = 1, \dots, m, i \neq j \quad C_i \cap Z_j = \emptyset. \quad (6)$$

В-третьих, все точки маршрутов должны находиться внутри допустимой рабочей области:

$$\forall i = 1, \dots, m, \forall k = 1, \dots, n_i \quad p_k^i \in \Omega. \quad (7)$$

3. Проверка пересечений воздушных коридоров. При построении нового маршрута требуется учитывать уже сформированные воздушные коридоры других БПЛА. Для этого на каждом шаге поиска следующей точки маршрута необходимо построить капсулу между текущей точкой и кандидатом на следующую точку, после чего выполнить проверку ее пересечения с капсулами существующих воздушных коридоров.

Пусть заданы две капсулы с несущими отрезками $[p, q]$ и $[r, s]$, а также радиусами R_1 и R_2 соответственно. Капсулы пересекаются, если выполняется условие:

$$d([p, q], [r, s]) \leq R_1 + R_2, \quad (8)$$

где $d([p, q], [r, s])$ – расстояние между отрезками.

Расстояние между отрезками определяется как наименьшее расстояние между произвольными точками, лежащими на этих отрезках.

Для аналитического вычисления расстояния удобно представить сами отрезки в параметрической форме с использованием линейной интерполяции между их конечными точками. Это позволяет выразить произвольную точку на каждом отрезке через скалярный параметр.

Точка на отрезке $[p, q]$ задаётся как:

$$x(\lambda_1) = (1 - \lambda_1)p + \lambda_1 q, \quad (9)$$

где $\lambda_1 \in [0, 1]$ определяет положение точки вдоль первого отрезка.

Точка на отрезке $[r, s]$ задается аналогично:

$$y(\lambda_2) = (1 - \lambda_2)r + \lambda_2 s, \quad (10)$$

где $\lambda_2 \in [0, 1]$ определяет положение точки вдоль второго отрезка.

Тогда поиск расстояния между отрезками сводится к минимизации евклидовой нормы между такими точками:

$$d([p, q], [r, s]) = \min_{\lambda_1, \lambda_2 \in [0, 1]} \|x(\lambda_1) - y(\lambda_2)\| = \min_{\lambda_1, \lambda_2 \in [0, 1]} \|(1 - \lambda_1)p + \lambda_1q - [(1 - \lambda_2)r + \lambda_2s]\| \quad (11)$$

Введем обозначения: $a = q - p$; $b = s - r$; $c = p - r$. Тогда точку на первом отрезке можно привести к виду:

$$x(\lambda_1) = p + \lambda_1 a, \quad (12)$$

а точку на втором отрезке к:

$$y(\lambda_2) = r + \lambda_2 b. \quad (13)$$

Для поиска оптимальных значений λ_1 и λ_2 удобно перейти к минимизации квадрата нормы, поскольку минимум нормы и минимум её квадрата достигаются при одном и том же значении параметров. Это позволяет избежать производной от квадратного корня и упростить дальнейшие вычисления. В свою очередь квадрат нормы удобно представить как скалярное произведение вектора, получаемого при вычитании векторов точек $x(\lambda_1)$ и $y(\lambda_2)$:

$$\min_{\lambda_1, \lambda_2 \in [0, 1]} \|x(\lambda_1) - y(\lambda_2)\|^2 = \min_{\lambda_1, \lambda_2 \in [0, 1]} \langle \lambda_1 a - \lambda_2 b + c, \lambda_1 a - \lambda_2 b + c \rangle \quad (14)$$

Минимум достигается в точке, где градиент функции по обоим параметрам обращается в нуль. Это условие приводит к системе линейных уравнений, полученной дифференцированием по λ_1 и λ_2 :

$$\begin{cases} \langle a, a \rangle \lambda_1 - \langle a, b \rangle \lambda_2 = -\langle a, c \rangle \\ \langle a, b \rangle \lambda_1 - \langle b, b \rangle \lambda_2 = -\langle b, c \rangle \end{cases} \quad (15)$$

Система решается по формуле Крамера [25]. Обозначим определитель:

$$\Delta = \langle a, b \rangle^2 - \langle a, a \rangle \cdot \langle b, b \rangle. \quad (16)$$

Если $\Delta > \varepsilon$, где ε – малое положительное число, решение считается устойчивым и оптимальные значения λ_1^* и λ_2^* вычисляются как:

$$\lambda_1^* = \frac{\langle b, b \rangle \cdot \langle a, c \rangle - \langle a, b \rangle \cdot \langle b, c \rangle}{\Delta}, \quad (17)$$

$$\lambda_2^* = \frac{\langle a, a \rangle \cdot \langle b, c \rangle - \langle a, b \rangle \cdot \langle a, c \rangle}{\Delta}. \quad (18)$$

Значения параметров ограничиваются отрезком $[0, 1]$, что соответствует проекции точек на соответствующие отрезки:

$$\lambda_i^{\min} = \min(\max(\lambda_i^*, 0), 1), \quad i = 1, 2. \quad (19)$$

После подстановки проецированных параметров $\lambda_1^{\min}$, $\lambda_2^{\min}$ значение расстояния между отрезками вычисляется по соответствующим точкам на каждом отрезке:

$$d([p, q], [r, s]) = \left\| (1 - \lambda_1^{\min})p + \lambda_1^{\min}q - [(1 - \lambda_2^{\min})r + \lambda_2^{\min}s] \right\|. \quad (20)$$

Если $\Delta < \varepsilon$, отрезки считаются почти параллельными, и аналитическое решение может быть численно неустойчивым. Также, если хотя бы одно из значений $\lambda_i^{\min}$, $i = 1, 2$ совпадает с границами интервала $[0, 1]$, то это указывает на граничный минимум, возникающий в результате проекции оптимальной точки за пределы отрезка и последующего приведения её к ближайшему концу.

В таком случае используется перебор всех пар граничных комбинаций: расстояние вычисляется между каждой из четырёх возможных пар граничных точек, и выбирается наименьшее значение:

$$d([p, q], [r, s]) = \min \{ \|p - r\|, \|p - s\|, \|q - r\|, \|q - s\| \}. \quad (21)$$

Таким образом, на каждом шаге построения маршрута выполняется проверка пересечений между формируемой капсулой и

всеми капсулами, ранее образованными отрезками воздушных коридоров других БПЛА. Эта процедура сводится к задаче минимизации квадрата евклидова расстояния между парой отрезков с учётом граничных и вырожденных случаев, после чего полученное расстояние сравнивается с суммой радиусов соответствующих капсул.

4. Определение пересечений воздушных коридоров с зонами обработки участков. При построении траекторий БПЛА необходимо учитывать ограничения, связанные с активными зонами обработки, в пределах которых нахождение других аппаратов недопустимо. Геометрически эти зоны аппроксимируются выпуклыми многогранниками в трёхмерном пространстве.

Каждая зона обработки $Z_j \subset \mathbb{R}^3$ задаётся в виде системы линейных неравенств, определяющих пересечение полупространств:

$$Z_j = \{x \in \mathbb{R}^3 : \langle n_i^j, x \rangle \leq c_i^j, \dots, \langle n_M^j, x \rangle \leq c_M^j\}, \quad (22)$$

где $n_i^j \in \mathbb{R}^3$ – нормаль к i -й грани многогранника (направленная наружу); $c_i^j \in \mathbb{R}$ – смещение соответствующей гиперплоскости, и M – общее число граней.

Проверка осуществляется для каждого сегмента маршрута $[p, q]$, которому соответствует капсула с радиусом R_i . Положение точки на осевой линии капсулы задаётся параметрически как:

$$x(\lambda) = (1 - \lambda)p + \lambda q, \quad (23)$$

где $\lambda \in [0, 1]$ определяет положение точки вдоль отрезка.

Для каждой i -й грани многогранника рассчитывается расстояние со знаком от текущей точки до соответствующей гиперплоскости:

$$d_i(\lambda) = \langle n_i^j, x(\lambda) \rangle - c_i^j = \alpha_i \lambda + \beta_i, \quad (24)$$

где $\alpha_i = \langle n_i^j, q - p \rangle$; $\beta_i = \langle n_i^j, p \rangle - c_i^j$.

Для фиксированного значения λ отклонение точки от многогранника определяется с использованием функции максимума расстояний до всех граней:

$$d_{\text{zone}}(\lambda) = \max_{1 \leq i \leq M} d_i(\lambda). \quad (25)$$

Геометрически это значение соответствует расстоянию от точки до границы зоны в том смысле, что:

- если $d_{\text{zone}}(\lambda) < 0$, то точка гарантированно лежит внутри многогранника;
- если $d_{\text{zone}}(\lambda) = 0$, то точка принадлежит его границе;
- если $d_{\text{zone}}(\lambda) > 0$, то точка вне зоны.

Функция $d_{\text{zone}}(\lambda)$ является кусочно-линейной, непрерывной и выпуклой на отрезке $[0, 1]$, поскольку максимум конечного набора линейных функций сохраняет выпуклость, что позволяет аналитически определить точку на осевой линии капсулы, наиболее близко расположенную к границам зоны обработки.

Минимум функции $d_{\text{zone}}(\lambda)$ достигается либо на концах отрезка, либо в точках пересечения двух функций $d_i(\lambda)$ и $d_j(\lambda)$, соответствующих смене активной грани, определяющей максимум:

$$\Lambda = \{0, 1\} \cup \left\{ \lambda = \frac{\beta_j - \beta_i}{\alpha_i - \alpha_j} \mid 1 \leq i < j \leq M, \alpha_i \neq \alpha_j, \lambda \in [0, 1] \right\}. \quad (26)$$

Для каждой $\lambda \in \Lambda$ вычисляется значение $d_{\text{zone}}(\lambda)$ и затем выбирается:

$$d_{\min} = \min_{\lambda \in \Lambda} d_{\text{zone}}(\lambda). \quad (27)$$

Полученное значение $d_{\min}$ интерпретируется как расстояние между осевой линией капсулы и поверхностью зоны обработки.

Если выполняется условие:

$$d_{\min} \leq R_i, \quad (28)$$

то это означает, что капсула проникает в зону обработки, и соответствующий сегмент маршрута считается конфликтным.

Описанный алгоритм применим в предположении выпуклости зоны. В случае, если зона имеет невыпуклую структуру, её геометрия должна быть предварительно разложена на конечное число выпуклых компонент. Для этого используются алгоритмы выпуклой

декомпозиции [26]. Тогда каждая компонентная зона проверяется независимо, а результирующее значение $d_{\min}$ принимается как минимум по всем подзонам.

5. Использование иерархической пространственной фильтрации для ускорения проверки пересечений. Вычислительная сложность проверки пересечений между сегментами маршрутов и зонами обработки растёт квадратично с увеличением числа БПЛА и элементов среды. Для снижения вычислительных затрат применяется иерархическая структура данных «свободное октодерево» (Loose Octree) [23] – модификация классического октодерева с расширенными границами ячеек.

Каждому сегменту маршрута, которому соответствует капсула, сопоставляется прямоугольный параллелепипед – ограничивающий объём, минимально охватывающий капсулу (Axis-Aligned Bounding Box, AABB) [27]. Аналогично, каждой зоне обработки сопоставляется прямоугольный ограничивающий параллелепипед, построенный по всем вершинам многогранника.

Координаты ограничивающего параллелепипеда определяются компонентно. Для сегмента маршрута $[p, q]$, которому соответствует капсула радиуса R_i , границы вычисляются как:

$$x_{\min}^{(ik)} = \min(p_k^x, p_{k+1}^x) - R_i, \quad (29)$$

$$x_{\max}^{(ik)} = \max(p_k^x, p_{k+1}^x) + R_i, \quad (30)$$

$$y_{\min}^{(ik)} = \min(p_k^y, p_{k+1}^y) - R_i, \quad (31)$$

$$y_{\max}^{(ik)} = \max(p_k^y, p_{k+1}^y) + R_i, \quad (32)$$

$$z_{\min}^{(ik)} = \min(p_k^z, p_{k+1}^z) - R_i, \quad (33)$$

$$z_{\max}^{(ik)} = \max(p_k^z, p_{k+1}^z) + R_i. \quad (34)$$

Для зоны обработки Z_j , заданной в виде выпуклого многогранника, координаты ограничивающего параллелепипеда определяются как минимум и максимум по координатам всех его вершин $v_1^j, \dots, v_N^j$:

$$x_{\min(j)} = \min(v_1^{j,x}, \dots, v_N^{j,x}), \quad (35)$$

$$x_{\max(j)} = \max(v_1^{j,x}, \dots, v_N^{j,x}), \quad (36)$$

$$y_{\min(j)} = \min(v_1^{j,y}, \dots, v_N^{j,y}), \quad (37)$$

$$y_{\max(j)} = \max(v_1^{j,y}, \dots, v_N^{j,y}), \quad (38)$$

$$z_{\min(j)} = \min(v_1^z, \dots, v_N^z), \quad (39)$$

$$z_{\max(j)} = \max(v_1^z, \dots, v_N^z). \quad (40)$$

В данном варианте октодеревя ограничивающие параллелепипеды объектов индексируются таким образом, что каждая ячейка покрывает определённую часть пространства, при этом вложенные объекты могут частично выходить за её границы. Такая организация позволяет избежать избыточного дробления пространства при работе с вытянутыми объектами, сохраняя эффективность индексации. Это обеспечивает быстрое отсеечение заведомо непересекающихся элементов и выполнение детальной проверки только для потенциально конфликтующих объектов.

Процесс отсеечения выполняется следующим образом: при добавлении нового сегмента маршрута сначала определяется набор ячеек октодеревя, с которыми пересекается его ограничивающий параллелепипед. Затем извлекаются только те объекты, которые находятся в этих ячейках или их родительских узлах. Объекты из непересекающихся ветвей дерева автоматически исключаются из проверки. Таким образом, вместо проверки всех пар объектов, алгоритм выполняет детальную проверку только для небольшого подмножества потенциально пересекающихся элементов.

Проверка пересечения между двумя ограничивающими объёмами выполняется по координатным интервалам. Два параллелепипеда считаются пересекающимися тогда и только тогда, когда выполнены следующие условия:

$$\text{Intersect}(\text{AABB}_1, \text{AABB}_2) \Leftrightarrow \begin{cases} x_{\min}^{(1)} \leq x_{\max}^{(2)} \wedge x_{\max}^{(1)} \geq x_{\min}^{(2)} \\ y_{\min}^{(1)} \leq y_{\max}^{(2)} \wedge y_{\max}^{(1)} \geq y_{\min}^{(2)} \\ z_{\min}^{(1)} \leq z_{\max}^{(2)} \wedge z_{\max}^{(1)} \geq z_{\min}^{(2)} \end{cases} \quad (41)$$

6. Алгоритм планирования маршрутов. Планирование маршрутов группы БПЛА организовано в два уровня. На верхнем уровне для каждого аппарата формируется полный маршрут, состоящий из четырёх фаз, соответствующих ключевым этапам движения: подъёму, входу в рабочую зону, выходу из неё и снижению. На нижнем уровне выполняется построение допустимого маршрута между заданной начальной и целевой точками, с учётом активных зон обработки и ранее зарезервированных маршрутов, проверяемых на пересечение с формируемыми капсулами. Такая организация обеспечивает независимость действий агентов, исключает необходимость глобальной синхронизации и позволяет реализовать пошаговое резервирование пространства.

Для каждого БПЛА маршрут формируется в виде последовательности четырех фаз, включающей:

1. Подъём от стартовой точки до рабочей высоты.
2. Горизонтальный переход к входу в зону обработки.
3. Возврат от выхода из зоны к точке начала снижения.
4. Спуск к исходной позиции.

Каждая из фаз строится как отдельный маршрут между двумя заданными точками с учётом геометрических ограничений. Построенные сегменты представляются в виде капсул фиксированного радиуса и регистрируются в октодереве. Псевдокод верхнего уровня представлен в Листинге 1.

Алгоритм реализует построение маршрутов для группы БПЛА по четырёхфазной схеме. На вход подаются: список агентов (UAVs), каждый из которых содержит координаты стартовой, входной и выходной точек, а также радиус безопасности капсулы; высота полёта (FlightAltitude); максимальное число итераций (MaxIterations); параметр точности достижения цели (Epsilon); вероятность выбора цели при генерации случайной точки (GoalBias) и зоны обработки (Zones) в виде многогранников.

```

function PlanPathsForUAVs(UAVs, FlightAltitude, MaxIterations, Epsilon,
GoalBias, Zones)
    LooseOctree := InitializeLooseOctree()

    // Индексация всех зон обработки
    for each Zone in Zones do
        ZoneAABB := ComputeAABB(Zone)
        InsertIntoLooseOctree(LooseOctree, Zone, ZoneAABB)
    end for

    for each UAV in UAVs do
        PBase := UAV.StartPoint
        PEntry := UAV.EntryPoint
        PExit := UAV.ExitPoint
        Radius := UAV.SafeRadius

        PTakeoff := (PBase.x, PBase.y, FlightAltitude)
        PEntryH := (PEntry.x, PEntry.y, FlightAltitude)
        PExitH := (PExit.x, PExit.y, FlightAltitude)

        // Фаза 1: Подъём
        Path1 := PlanPath(PBase, PTakeoff, MaxIterations,
Epsilon, GoalBias, Radius, LooseOctree)

        // Фаза 2: До входа в зону
        Path2 := PlanPath(PTakeoff, PEntryH, MaxIterations,
Epsilon, GoalBias, Radius, LooseOctree)

        // Фаза 3: Выход из зоны
        Path3 := PlanPath(PExitH, PTakeoff, MaxIterations,
Epsilon, GoalBias, Radius, LooseOctree)

        // Фаза 4: Снижение
        Path4 := Reverse(Path1)

        PathFinal := Concatenate(Path1, Path2, Path3, Path4)

        for each Segment in PathFinal do
            Capsule := CreateCapsule(Segment, Radius)
            CapsuleAABB := ComputeAABB(Capsule)
            InsertIntoLooseOctree(LooseOctree, Capsule,
CapsuleAABB)
        end for
        UAV.Path := PathFinal
    end for
    return UAVs
end function

```

Листинг 1. Псевдокод верхнего уровня планирования

На первом этапе все зоны индексируются в октодереве для последующей фильтрации. Далее для каждого агента последовательно формируются маршруты по четырём фазам, каждая из которых строится с использованием алгоритма нижнего уровня (PlanPath, Листинг 2). После завершения всех фаз полученная траектория преобразуется в набор капсул, регистрируемых в октодереве, и используется для проверки допустимости траекторий других агентов.

```

function PlanPath(PStart, PGoal, MaxIterations, Epsilon, GoalBias, Radius,
LooseOctree)
    Tree := {PStart}

    for i = 1 to MaxIterations do
        if Random() < GoalBias then
            PRand := PGoal
        else
            PRand := SampleUniform()
        end if

        PNear := FindNearest(Tree, PRand)
        PNew := Steer(PNear, PRand)

        Capsule := CreateCapsule(PNear, PNew, Radius)

        ZoneConflict := CheckZoneIntersection(LooseOctree,
Capsule)
        CorridorConflict := CheckCapsuleIntersection
(LooseOctree, Capsule)

        if not (ZoneConflict or CorridorConflict) then
            Tree := AddToTree(Tree, PNew)

            // Перестроение дерева: проверка новых рёбер
            Neighbors := FindNearbyNodes(Tree, PNew)
            for each PNeighbor in Neighbors do
                Cost(PNew) + Distance(PNew,
PNeighbor) < Cost(PNeighbor) then
                    NewCapsule :=
CreateCapsule(PNew, PNeighbor, Radius)
                    ZConflict :=
CheckZoneIntersection(LooseOctree, NewCapsule)
                    CConflict :=
CheckCapsuleIntersection(LooseOctree, NewCapsule)

```

```

                                if not (ZConflict or
CConflict) then
                                UpdateParent(Tree,
PNeighbor, PNew)
                                end if
                                end if
                                end for
                                if ReachedGoal(PNew, PGoal, Epsilon) then
                                return ExtractPath(Tree, PGoal)
                                end if
                                end if
                                end for
                                return EmptyPath
end function

```

Листинг 2. Псевдокод нижнего уровня планирования

На нижнем уровне используется алгоритм, основанный RRT [13] со следующими модификациями:

1. Использование капсульной модели маршрута, позволяющей формировать воздушные коридоры фиксированной толщины.

2. Проверка на конфликты с уже построенными маршрутами, основанная на пересечении капсул и ускоренная с применением октодерева.

3. Аналитическая проверка на пересечение с выпуклыми зонами обработки, реализуемая через оценку расстояния от капсулы до границ многогранника.

4. Локальное перестроение дерева RRT*, при котором каждая новая вершина дополнительно проверяется на допустимость при попытке переключения соседних узлов.

5. Механизм направленного семплирования [28]: с заданной вероятностью (GoalBias) в качестве следующей точки выбирается целевая, что способствует ускорению сходимости.

В результате формируется сеть маршрутов, обеспечивающих безопасные воздушные коридоры для каждого БПЛА, примеры которых приведены на рисунках 1 и 2.

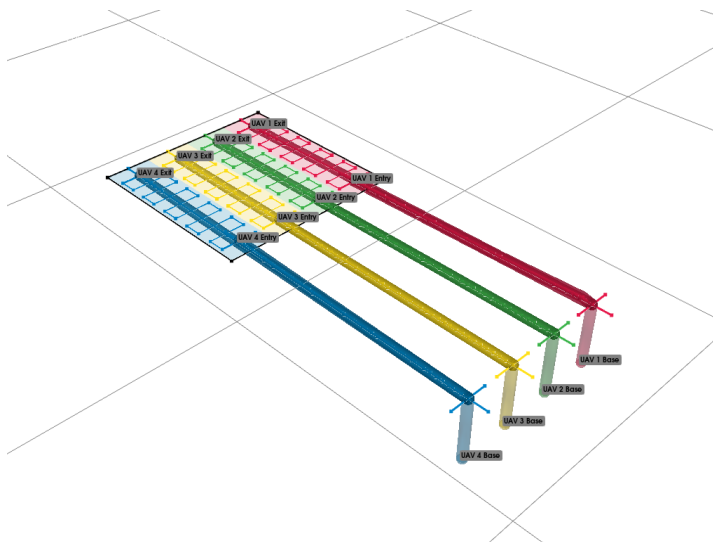


Рис. 1. Пример сети воздушных коридоров при разбиении поля на ряды для четырех БПЛА

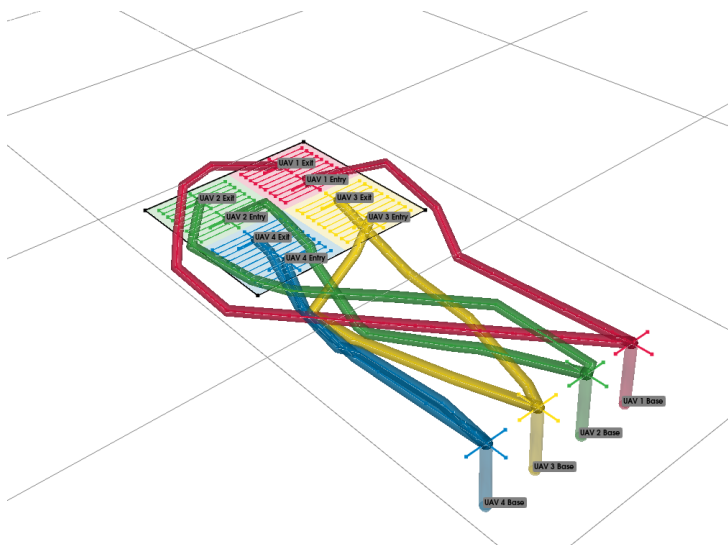


Рис. 2. Пример сети воздушных коридоров при сеточном разбиении поля для четырех БПЛА

Различие между рисунками демонстрирует влияние геометрии и параметра вероятностного выбора целевой точки на форму маршрутов: при линейном разбиении (рисунок 1) траектории близки к прямым благодаря упорядоченности расположения входов и выходов, тогда как при сеточной топологии (рисунок 2) из-за увеличения плотности и числа пересечений капсулы отклоняются от кратчайших направлений, формируя более разреженную структуру.

7. Вычислительные эксперименты. Для оценки работоспособности и вычислительной эффективности предложенного метода был реализован программный прототип на языке C++17 с использованием стандартной библиотеки и собственной реализации структуры октодеревя. Все эксперименты проводились на рабочей станции под управлением Ubuntu 22.04 с процессором Intel Core i7-12700H и 16 Гб оперативной памяти.

В качестве базовой сцены использовался прямоугольный сельскохозяйственный участок площадью 1 км², равномерно разбиваемый на прямоугольную сетку зон обработки по числу БПЛА. При этом количество строк в сетке выбирается как $\lfloor \sqrt{N} \rfloor$, а число

столбцов как $\left\lceil \frac{N}{\lfloor \sqrt{N} \rfloor} \right\rceil$, где N – общее число БПЛА. Каждому

аппарату сопоставляется отдельный подучасток, центр которого используется для задания входной и выходной точек.

Область планирования расширяется до 1.5 км по каждой горизонтальной оси, чтобы обеспечить необходимый запас пространства для формирования маршрутов с резервированием. Вертикальная граница фиксирована на уровне 25 м, что соответствует типичной высоте выполнения сельскохозяйственных задач.

Каждый маршрут строится по четырёхфазной схеме, описанной в разделе 6, а сегменты маршрута интерпретируются в виде капсул радиуса 1 м. Это значение отражает типичные габариты квадрокоптеров с добавленным зазором для безопасного пролёта.

Параметры алгоритма подобраны экспериментально для обеспечения баланса между качеством решения и вычислительной эффективностью. Экспериментальные параметры приведены в таблице 1.

Таблица 1. Экспериментальные параметры

Параметр	Значение	Обоснование
Площадь участка	1 км ²	Типичный размер сельскохозяйственного поля
Размер области планирования	1.5 × 1.5 × 0.025 км ³	Расширение для маневрирования
Радиус капсулы	1 м	Габариты БПЛА + запас безопасности
Шаг расширения RRT*	7 м	Компромисс точность/скорость
Радиус локального соединения	15 м	Охват соседних узлов без перегрузки
Максимум итераций на фазу	30000	Практический предел сходимости
Вероятность выбора цели	0.45	Баланс направленного/случайного поиска

Для обеспечения статистической достоверности результатов и оценки устойчивости алгоритма к случайным факторам инициализации каждая конфигурация тестировалась 25 раз.

В ходе численного эксперимента анализировались характеристики, отражающие поведение алгоритма в зависимости от размера группы агентов. В качестве метрик оценки эффективности выбраны:

1. Среднее время построения маршрута на одного агента, как показатель вычислительной нагрузки.
2. Среднее число итераций до завершения планирования, отражающее сложность конфигурационного пространства.
3. Длина полученного маршрута, характеризующая влияние ограничений на форму траектории.

Графики на рисунках 3 и 4 показывают, как изменяются эти показатели при увеличении числа агентов от 2 до 32.

Среднее и стандартное отклонение метрик (25 экспериментов)

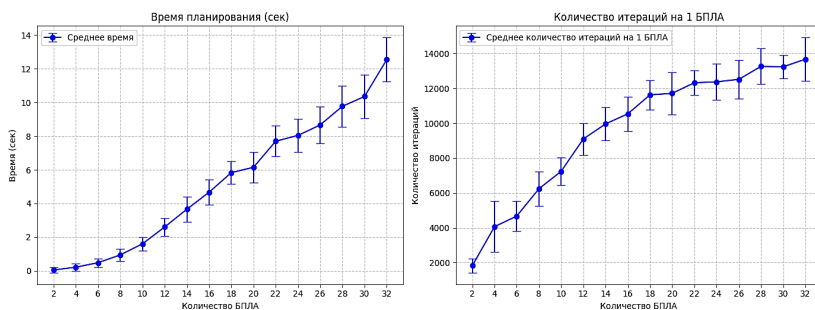


Рис. 3. Зависимости времени выполнения и количества итераций от количества БПЛА

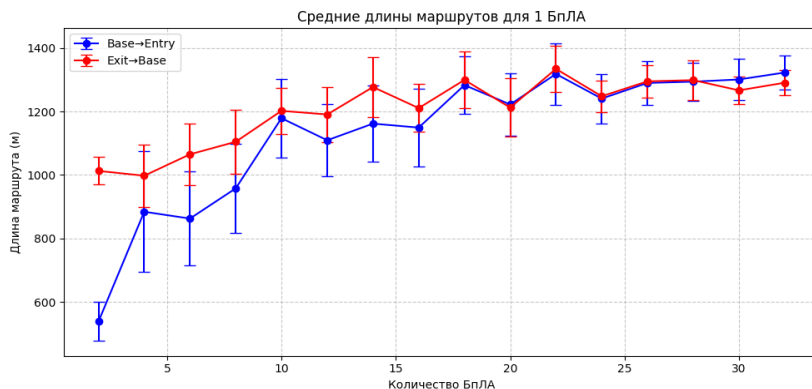


Рис. 4. Зависимость средних длин маршрутов для 1 БЛА от размера группы

Как и ожидалось, рост числа БЛА приводит к усложнению планирования: время увеличивается с 0.05 с до 12.5 с, а среднее число итераций – с 1800 до более чем 13 000. Это объясняется тем, что каждый новый маршрут строится в уже частично занятом пространстве, и каждый шаг планирования требует больше проверок на пересечения.

Длина маршрутов также постепенно увеличивается, так как новые БЛА вынуждены обходить уже построенные воздушные коридоры. Визуализации на рисунке 5 демонстрируют, как плотность маршрутов возрастает при увеличении числа агентов: свободное пространство постепенно вытесняется резервированными капсулами, оставляя всё меньше направлений для расширения дерева.

Важно отметить, что несмотря на рост вычислительной нагрузки, метод сохраняет устойчивость и успешно масштабируется до 32 агентов без существенного ухудшения качества маршрутов. При высокой плотности препятствий эффективность иерархического представления среды, основанного на октодереве, позволяет значительно сократить количество проверок на коллизии и обеспечить приемлемое вычислительное время даже в условиях высокой загруженности пространства.

Таким образом, проведённые эксперименты подтверждают применимость метода для групповой навигации БЛА в ограниченном воздушном пространстве. Результаты показывают, что предложенный подход сохраняет разумный баланс между точностью, стабильностью и вычислительной сложностью.

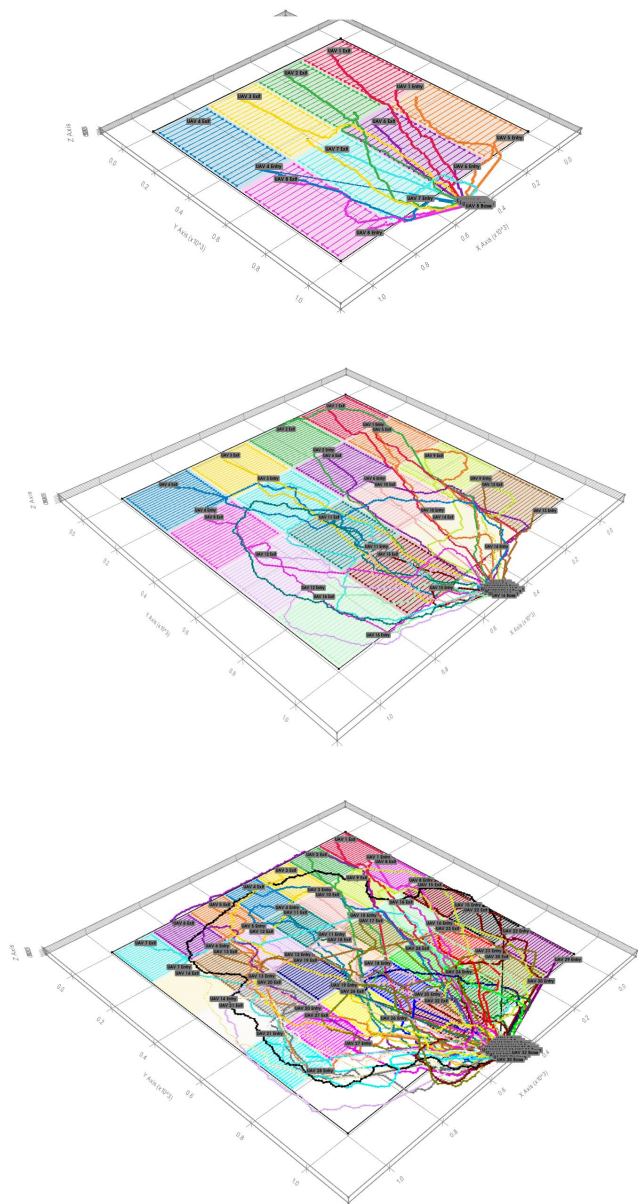


Рис. 5. Рост плотности коридорных маршрутов по мере увеличения числа БПЛА в группе на примере 8,16 и 32 БПЛА

8. Заключение. В данном исследовании представлен метод построения безопасных маршрутов группы БПЛА на основе капсульной модели воздушных коридоров, обеспечивающий пространственное резервирование траекторий в условиях плотного многоагентного движения. Предложенный метод решает ключевые проблемы групповой навигации БПЛА в агропромышленных применениях: обеспечивает автономность выполнения полётов, устойчивость к навигационным ошибкам и внешним воздействиям, независимость от централизованной координации в реальном времени.

Реализация на базе модифицированного алгоритма RRT* с интегрированными геометрическими проверками и иерархической пространственной индексацией демонстрирует вычислительную эффективность и масштабируемость.

Экспериментальные результаты подтвердили, что метод масштабируется до 32 БПЛА без значительной деградации качества маршрутов и при сохранении вычислительной эффективности. Время планирования растёт предсказуемо, без катастрофического ухудшения производительности.

Ключевым преимуществом метода является конструктивное обеспечение безопасности через геометрическое резервирование пространства, что исключает необходимость в сложных динамических моделях, онлайн-синхронизации и высокоточной навигации. Построенные маршруты могут быть напрямую загружены в управляющее программное обеспечение БПЛА в виде последовательности контрольных точек, обеспечивая полную автономность миссии.

Дальнейшее развитие метода связано с реализацией механизмов динамического освобождения воздушных коридоров после прохождения участков маршрута, что позволит повысить эффективность использования воздушного пространства при многократных полётах и увеличенных размерах групп БПЛА.

Рафаэль Мидхатович Юсупов навсегда останется в памяти как выдающийся учёный, мудрый наставник и человек высокой ответственности. Его путь – от военного инженера-исследователя до директора Санкт-Петербургского института информатики и автоматизации РАН – стал образцом преданности делу и служения стране. Более четверти века, возглавляя институт, он укреплял его научный потенциал и развивал новые направления в области информатики и теории управления. Обладая редкой глубиной мышления и организаторским даром, он создал научную школу, подготовил десятки учеников и внёс весомый вклад в развитие

информационного общества. Светлая память о Рафаэле Мидхатовиче – как о талантливом руководителе и наставнике – навсегда сохранится в сердцах коллег, учеников и всех, кому посчастливилось работать с ним.

Литература

1. Gubanov B., Lebedeva V., Lebedev I., Astapova M. Algorithms and software for evaluation of plant height in vertical farm using UAVs // *Agriculture Digitalization and Organic Production: Proceedings of the Second International Conference*. Singapore: Springer Nature Singapore, 2022. pp. 351–362. DOI: 10.1007/978-981-19-7780-0_31.
2. Шайтура Н.С. Мониторинг сельскохозяйственных земель при помощи беспилотных летательных аппаратов // *Практические аспекты применения современных беспилотных летательных аппаратов*. 2022. С. 46–57.
3. Nikolakopoulos I.A., Petropoulos G.P. Obtaining a Land Use/Cover Cartography in a Typical Mediterranean Agricultural Field Combining Unmanned Aerial Vehicle Data with Supervised Classifiers // *Land*. 2025. vol. 14. no. 3. DOI: 10.3390/land14030643.
4. Павлова А.И. Применение беспилотных и геоинформационных технологий для цифрового почвенного картографирования // *Информационные технологии, системы и приборы в АПК. Агроинфо-2021*. Т. 2021. С. 97–99.
5. Hassler S.C., Baysal-Gurel F. Unmanned aircraft system (UAS) technology and applications in agriculture // *Agronomy*. 2019. vol. 9. no. 10. DOI: 10.3390/agronomy9100618.
6. Velusamy P., Rajendran S., Mahendran R.K., Naseer S., Shafiq M., Choi J.G. Unmanned Aerial Vehicles (UAV) in precision agriculture: Applications and challenges // *Energies*. 2021. vol. 15. no. 1. DOI: 10.3390/en15010217.
7. Makam S., Komatineni B.K., Meena S.S., Meena U. Unmanned aerial vehicles (UAVs): an adoptable technology for precise and smart farming // *Discover Internet of Things*. 2024. vol. 4. DOI: 10.1007/s43926-024-00066-5.
8. Sharon G., Stern R., Felner A., Sturtevant N. R. Conflict-based search for optimal multi-agent pathfinding // *Artificial Intelligence*. 2015. vol. 219. pp. 40–66. DOI: 10.1016/j.artint.2014.11.006.
9. Wang Z., Zhang Z., Dou W., Hu G., Zhang L., Zhang M. Extending Conflict-Based Search for Optimal and Efficient Quadrotor Swarm Motion Planning // *Drones*. 2024. vol. 8. no. 12. DOI: 10.3390/drones8120719.
10. Andreychuk A., Yakovlev K., Boyarski E., Stern R. Improving continuous-time conflict based search // *Proceedings of the AAAI Conference on Artificial Intelligence*. 2021. vol. 35. no. 13. pp. 11220–11227. DOI: 10.1609/aaai.v35i13.17338.
11. Liu X., Su Y., Wu Y., Guo Y. Multi-conflict-based optimal algorithm for multi-UAV cooperative path planning // *Drones*. 2023. vol. 7. no. 3. DOI: 10.3390/drones7030217.
12. Cap M., Novak P., Vokrinek J., Pechoucek M. Multi-agent RRT*: Sampling-based cooperative pathfinding // *arXiv preprint arXiv:1302.2828*. 2013.
13. Karaman S., Frazzoli E. Sampling based algorithms for optimal motion planning // *The International Journal of Robotics Research*. 2011. vol. 30. no. 7. pp. 846–894. DOI: 10.1177/0278364911406761.
14. Silver D. Cooperative pathfinding // *Proceedings of the AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment*. 2005. vol. 1. no. 1. pp. 117–122. DOI: 10.1609/aiide.v1i1.18726.
15. Barer M., Sharon G., Stern R., Felner A. Suboptimal variants of the conflict based search algorithm for the multi agent pathfinding problem // *Proceedings of the International Symposium on Combinatorial Search*. 2021. vol. 5. no. 1. pp. 19–27. DOI: 10.1609/socs.v5i1.18315.
16. Van Den Berg J., Guy S.J., Lin M., Manocha D. Reciprocal n-body collision avoidance // *Robotics Research: The 14th International Symposium ISRR*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011. pp. 3–19.

17. Van den Berg J., Lin M., Manocha D. Reciprocal velocity obstacles for real-time multi-agent navigation // IEEE International Conference on Robotics and Automation. 2008. pp. 1928–1935. DOI: 10.1109/ROBOT.2008.4543489.
18. Yan X., Jiang D., Miao R., Li Y. Formation control and obstacle avoidance algorithm of a multi-USV system based on virtual structure and artificial potential field // Journal of Marine Science and Engineering. 2021. vol. 9. no. 2. DOI: 10.3390/jmse9020161.
19. Wu W., Zhang X., Miao Y. Starling behavior inspired flocking control of fixed wing unmanned aerial vehicle swarm in complex environments with dynamic obstacles // Biomimetics. 2022. vol. 7. no. 4. DOI: 10.3390/biomimetics7040214.
20. Alqudsi Y. Synchronous task allocation and trajectory optimization for autonomous drone swarm // 1st International Conference on Emerging Technologies for Dependable Internet of Things (ICETI). 2024. pp. 1–8. DOI: 10.1109/ICETI63946.2024.10777195.
21. Котов Д.В., Лебедев О.Б. Управление передвижением группы БПЛА с соблюдением геометрической структуры строя на основе альтернативной коллективной адаптации // Известия ЮФУ. Технические науки. 2024. Т. 1. С. 155–167.
22. Саенко И.Б., Лаута О.С., Митяков Е.С., Соколов А.П. Алгоритм роевого управления БПЛА с элементами кластерного анализа // Информация и космос. 2024. № 4. С. 68–75.
23. Ulrich T. Loose octrees // Game Programming Gems. 2000. vol. 1. pp. 434–442.
24. Pournin L., Weber M., Tsukahara M., Ferrez J.A., Ramaioli M., Liebling T.M. Three dimensional distinct element simulation of spherocylinder crystallization // Granular Matter. 2005. vol. 7. pp. 119–126. DOI: 10.1007/s10035-004-0188-4.
25. Bretscher O. Linear algebra with applications. NJ: Prentice Hall, 1997. 587 p.
26. Lien J.M., Amato N.M. Approximate convex decomposition of polyhedra and its applications // Computer Aided Geometric Design. 2008. vol. 25. no. 7. pp. 503–522.
27. Rendering W.P.B. Physically based rendering // Procedia IUTAM. 2015. vol. 13. pp. 3.
28. Kang G., Kim Y.B., You W.S., Lee Y.H., Oh H.S., Moon H., Choi H.R. Sampling based path planning with goal oriented sampling // IEEE International Conference on Advanced Intelligent Mechatronics (AIM). 2016. pp. 1285–1290. DOI: 10.1109/AIM.2016.7576947.

Савельев Антон Игоревич — канд. техн. наук, заведующий лабораторией, лаборатория автономных робототехнических средств, Федеральное государственное бюджетное учреждение науки «Санкт-Петербургский Федеральный исследовательский центр Российской академии наук» (СПб ФИЦ РАН). Область научных интересов: математические модели управления робототехническими средствами, методы и алгоритмы для построения траекторий движения робототехнических систем. Число научных публикаций — 151. saveliev@ias.spb.su; 14-я линия В.О., 39, 199178, Санкт-Петербург, Россия; р.т.: +7(812)328-4743.

Аникин Дмитрий Андреевич — младший научный сотрудник, лаборатория автономных робототехнических средств, Федеральное государственное бюджетное учреждение науки «Санкт-Петербургский Федеральный исследовательский центр Российской академии наук» (СПб ФИЦ РАН). Область научных интересов: алгоритмы группового управления наземными и воздушными роботами, алгоритмы для автоматической навигации, обнаружения препятствий и принятия решений в реальном времени. Число научных публикаций — 11. anikin.d@ias.spb.su; 14-я линия В.О., 39, 199178, Санкт-Петербург, Россия; р.т.: +7(812)328-4743.

Поддержка исследований. Работа выполнена в рамках бюджетной темы № FFZF-2025-0012.

A. SAVELIEV, D. ANIKIN
**METHOD OF CALCULATING CAPSULE-SHAPED AIR
CORRIDORS OF SAFE ROUTES FOR A GROUP OF UNMANNED
AERIAL VEHICLES**

Saveliev A., Anikin D. Method of Calculating Capsule-Shaped Air Corridors of Safe Routes for a Group of Unmanned Aerial Vehicles.

Abstract. The paper considers the problem of constructing safe routes for a group of unmanned aerial vehicles in a limited airspace over an agricultural area. The relevance of the study is due to the growing use of UAV groups in the agro-industrial complex for monitoring, mapping, and processing fields, which requires ensuring flight safety in conditions of high air traffic density, limited communication, and exposure to external factors. A particular challenge is the need for autonomous missions in the presence of navigation errors and natural impacts. A route planning method is proposed based on representing the trajectory of each device as a capsule air corridor – a three-dimensional volume of a fixed radius formed along the trajectory segments. Spatial redundancy ensures safe spacing of trajectories at the planning stage, eliminating conflicts during subsequent autonomous flight operations without the need for continuous coordination between agents. The capsule radius includes a reserve for possible deviations from the planned trajectory, which ensures resistance to navigation errors. The method is based on the sequential formation of routes for each device according to a four-phase scheme, including a vertical ascent from the starting point to the operating altitude, a horizontal transition to the entrance to the processing zone, a return from the exit from the zone to the starting point of the descent, and a vertical descent to the initial position. Each new route is built considering the already reserved air corridors through an analytical check of geometric intersections between the capsules of different trajectories and convex polyhedrons of the processing zones. To improve computational efficiency, hierarchical spatial filtering is used based on bounding parallelepipeds, which allows for the rapid cutting off of obviously non-intersecting objects at the preliminary stage and performing an accurate geometric check only for potentially conflicting route segments. Numerical experiments were carried out for groups of 2 to 32 devices on a typical agricultural plot of one square kilometer. A nonlinear increase in the planning time and the number of iterations with an increase in the number of agents was found, which is due to the need to build each subsequent route in an already partially occupied space with an increasing number of spatial constraints. The length of routes shows a tendency to increase, especially pronounced at the initial stages of scaling, which is associated with the need to bypass already reserved air corridors.

Keywords: UAVs, air corridors, agricultural tasks, Rapidly-Exploring Random Tree*, Loose Octree, autonomous navigation, group planning.

References

1. Gubanov B., Lebedeva V., Lebedev I., Astapova M. Algorithms and software for evaluation of plant height in vertical farm using UAVs. Agriculture Digitalization and Organic Production: Proceedings of the Second International Conference. Singapore: Springer Nature Singapore, 2022. pp. 351–362. DOI: 10.1007/978-981-19-7780-0_31.
2. Shaitura N.S. [Monitoring agricultural lands using unmanned aerial vehicles]. *Prakticheskie aspekty primeneniya sovremennykh bespilotnykh letatel'nykh apparatov* [Practical aspects of the use of modern unmanned aerial vehicles]. 2022. pp. 46–57. (In Russ.).

3. Nikolakopoulos I.A., Petropoulos G.P. Obtaining a Land Use/Cover Cartography in a Typical Mediterranean Agricultural Field Combining Unmanned Aerial Vehicle Data with Supervised Classifiers. *Land*. 2025. vol. 14. no. 3. DOI: 10.3390/land14030643.
4. Pavlova A.I. [Application of unmanned and geoinformation technologies for digital soil mapping]. *Informacionnye tehnologii, sistemy i pribory v APK. Agroinfo-2021 [Information technologies, systems and devices in agroindustrial complex. Agroinfo-2021]*. vol. 2021. pp. 97–99. (In Russ.).
5. Hassler S.C., Baysal-Gurel F. Unmanned aircraft system (UAS) technology and applications in agriculture. *Agronomy*. 2019. vol. 9. no. 10. DOI: 10.3390/agronomy9100618.
6. Velusamy P., Rajendran S., Mahendran R.K., Naseer S., Shafiq M., Choi J.G. Unmanned Aerial Vehicles (UAV) in precision agriculture: Applications and challenges. *Energies*. 2021. vol. 15. no. 1. DOI: 10.3390/en15010217.
7. Makam S., Komatineni B.K., Meena S.S., Meena U. Unmanned aerial vehicles (UAVs): an adoptable technology for precise and smart farming. *Discover Internet of Things*. 2024. vol. 4. DOI: 10.1007/s43926-024-00066-5.
8. Sharon G., Stern R., Felner A., Sturtevant N. R. Conflict-based search for optimal multi-agent pathfinding. *Artificial Intelligence*. 2015. vol. 219. pp. 40–66. DOI: 10.1016/j.artint.2014.11.006.
9. Wang Z., Zhang Z., Dou W., Hu G., Zhang L., Zhang M. Extending Conflict-Based Search for Optimal and Efficient Quadrotor Swarm Motion Planning. *Drones*. 2024. vol. 8. no. 12. DOI: 10.3390/drones8120719.
10. Andreychuk A., Yakovlev K., Boyarski E., Stern R. Improving continuous-time conflict based search. *Proceedings of the AAAI Conference on Artificial Intelligence*. 2021. vol. 35. no. 13. pp. 11220–11227. DOI: 10.1609/aaai.v35i13.17338.
11. Liu X., Su Y., Wu Y., Guo Y. Multi-conflict-based optimal algorithm for multi-UAV cooperative path planning. *Drones*. 2023. vol. 7. no. 3. DOI: 10.3390/drones7030217.
12. Cap M., Novak P., Vokrinek J., Pechoucek M. Multi-agent RRT*: Sampling-based cooperative pathfinding. *arXiv preprint arXiv:1302.2828*. 2013.
13. Karaman S., Frazzoli E. Sampling based algorithms for optimal motion planning. *The International Journal of Robotics Research*. 2011. vol. 30. no. 7. pp. 846–894. DOI: 10.1177/0278364911406761.
14. Silver D. Cooperative pathfinding. *Proceedings of the AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment*. 2005. vol. 1. no. 1. pp. 117–122. DOI: 10.1609/aiide.v1i1.18726.
15. Barer M., Sharon G., Stern R., Felner A. Suboptimal variants of the conflict based search algorithm for the multi agent pathfinding problem. *Proceedings of the International Symposium on Combinatorial Search*. 2021. vol. 5. no. 1. pp. 19–27. DOI: 10.1609/socs.v5i1.18315.
16. Van Den Berg J., Guy S.J., Lin M., Manocha D. Reciprocal n-body collision avoidance. *Robotics Research: The 14th International Symposium ISRR*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011. pp. 3–19.
17. Van den Berg J., Lin M., Manocha D. Reciprocal velocity obstacles for real-time multi-agent navigation. *IEEE International Conference on Robotics and Automation*. 2008. pp. 1928–1935. DOI: 10.1109/ROBOT.2008.4543489.
18. Yan X., Jiang D., Miao R., Li Y. Formation control and obstacle avoidance algorithm of a multi-USV system based on virtual structure and artificial potential field. *Journal of Marine Science and Engineering*. 2021. vol. 9. no. 2. DOI: 10.3390/jmse9020161.
19. Wu W., Zhang X., Miao Y. Starling behavior inspired flocking control of fixed wing unmanned aerial vehicle swarm in complex environments with dynamic obstacles. *Biomimetics*. 2022. vol. 7. no. 4. DOI: 10.3390/biomimetics7040214.

20. Alqudsi Y. Synchronous task allocation and trajectory optimization for autonomous drone swarm. 1st International Conference on Emerging Technologies for Dependable Internet of Things (ICETI). 2024. pp. 1–8. DOI: 10.1109/ICETI63946.2024.10777195.
21. Kotov D.V., Lebedev O.B. [Control of movement of a group of UAVs with observance of the geometric structure of the formation based on alternative collective adaptation]. *Izvestija JuFU. Tehnicheskie nauki – Bulletin of SFedU. Technical sciences*. 2024. vol. 1. pp. 155–167. (In Russ.).
22. Saenko I.B., Lauts O.S., Mityakov E.S., Sokolov A.P. [Algorithm for swarm control of UAVs with elements of cluster analysis]. *Informacija i kosmos – Information and Space*. 2024. no. 4. pp. 68–75. (In Russ.).
23. Ulrich T. Loose octrees. *Game Programming Gems*. 2000. vol. 1. pp. 434–442.
24. Pournin L., Weber M., Tsukahara M., Ferrez J.A., Ramaioli M., Liebling T.M. Three dimensional distinct element simulation of spherocylinder crystallization. *Granular Matter*. 2005. vol. 7. pp. 119–126. DOI: 10.1007/s10035-004-0188-4.
25. Bretscher O. *Linear algebra with applications*. NJ: Prentice Hall, 1997. 587 p.
26. Lien J.M., Amato N.M. Approximate convex decomposition of polyhedra and its applications. *Computer Aided Geometric Design*. 2008. vol. 25. no. 7. pp. 503–522.
27. Rendering W.P.B. Physically based rendering. *Procedia IUTAM*. 2015. vol. 13. pp. 3.
28. Kang G., Kim Y.B., You W.S., Lee Y.H., Oh H.S., Moon H., Choi H.R. Sampling based path planning with goal oriented sampling. *IEEE International Conference on Advanced Intelligent Mechatronics (AIM)*. 2016. pp. 1285–1290. DOI: 10.1109/AIM.2016.7576947.

Saveliev Anton — Ph.D., Head of the laboratory, Laboratory of autonomous robotic means, St. Petersburg Federal Research Center of the Russian Academy of Sciences (SPC RAS). Research interests: mathematical models of robotic means control, methods and algorithms for constructing trajectories of robotic systems. The number of publications — 151. saveliev@ias.spb.su; 39, 14th Line of V.O., 199178, St. Petersburg, Russia; office phone: +7(812)328-4743.

Anikin Dmitry — Junior researcher, Laboratory of autonomous robotic means, St. Petersburg Federal Research Center of the Russian Academy of Sciences (SPC RAS). Research interests: algorithms for group control of ground and air robots, algorithms for automatic navigation, obstacle detection and decision-making in real time. The number of publications — 11. anikin.d@ias.spb.su; 39, 14th Line of V.O., 199178, St. Petersburg, Russia; office phone: +7(812)328-4743.

Acknowledgements. This research was carried out within the framework of budget theme No. FFZF-2025-0012.

Д.К. ЛЕВОНЕВСКИЙ, А.И. МОТИЕНКО, А.В. БОГОМОЛОВ АВТОМАТИЗАЦИЯ КИБЕРФИЗИЧЕСКОГО МОНИТОРИНГА СОСТОЯНИЯ ПАЦИЕНТОВ В УМНЫХ ПАЛАТАХ

Левоневский Д.К., Мотиенко А.И., Богомолов А.В. Автоматизация киберфизического мониторинга состояния пациентов в умных палатах.

Аннотация. Цифровизация здравоохранения предполагает широкое внедрение технологий информатизации и автоматизации процессов мониторинга состояния человека, ориентированных на сохранение, укрепление и восстановление его здоровья. Для адаптации существующих решений в области электронного здравоохранения к медицинской инфраструктуре выполнено моделирование расширяемых программных компонентов автоматизированного киберфизического мониторинга состояния пациентов в умных палатах. Представлена математическая модель процессов киберфизического мониторинга состояния пациента, описывающая медицинский алгоритм как эстафету с альтернативными маршрутами с помощью аппарата параллельных полумарковских процессов. Охарактеризованы особенности программных компонентов киберфизического мониторинга состояния пациента как гибкого, расширяемого интегрируемого элемента умной медицинской палаты, позволяющего персонифицировать и оптимизировать лечебно-диагностические процедуры с помощью целевой функции, характеризующей отклонение показателей текущего состояния пациента от показателей его целевого состояния. Продемонстрированы потенциальные возможности автоматизации медицинских процессов при реализации методик диагностики и рискометрии здоровья, основанных на использовании балльных шкал. Результаты моделирования показали, что использование программных компонентов автоматизированного киберфизического мониторинга состояния пациентов в умных палатах позволяет не только ускорить расчёты, но и сократить время получения данных благодаря их автоматической загрузке из информационных систем и автоматизированному снятию показателей с медицинских устройств, открывая новые возможности персонифицированного управления оказанием стационарной медицинской помощи.

Ключевые слова: информатизация здравоохранения, медицинская информатика, автоматизация мониторинга состояния, умная палата, киберфизическая система, программный компонент, киберфизический мониторинг.

1. Введение. Здоровье человека характеризуется множеством показателей, отражающих активность функциональных систем организма. Это обуславливает насущную необходимость мониторинга здоровья, глубина которого (количество показателей и дискретность их мониторинга) определяется спецификой решаемых прикладных задач здравоохранения [1, 2]. В настоящее время наиболее часто применяется дистанционный мониторинг состояния здоровья и окружающей среды человека – разновидность телемедицинских технологий, позволяющая отслеживать во времени выбранные показатели здоровья человека и показатели окружающей среды, накапливать эти данные в цифровом виде и передавать их для анализа

врачу или медицинскому персоналу, чтобы реализовать персонифицированное управление здоровьем [3 – 5].

Автоматизация процессов мониторинга здоровья – это использование технологий и систем для сбора, анализа и интерпретации данных о состоянии здоровья человека без необходимости постоянного вмешательства медицинских специалистов [2]. Она использует широкий спектр устройств: от носимых и бесконтактных до мобильных приложений типа «лаборатория на чипе» и систем телемедицины. Применение интеллектуальных медицинских технологий сбора и обработки данных о состоянии человека, выработки предварительных результатов и рекомендаций с помощью систем поддержки принятия решений, включая технологии искусственного интеллекта, позволяет эффективно решать задачи персонифицированного управления здоровьем [6]. Кроме того, большие языковые модели, такие как ChatGPT, Deepseek и подобные им, могут быть использованы для интерпретации медицинских данных, генерации персонализированных рекомендаций и общения с пациентами в понятной для них форме.

В настоящее время мониторинг здоровья осуществляется с помощью технологий киберфизических систем, реализующих интеграцию вычислительных ресурсов в инженерно-технические системы [7, 8]. Вычислительная компонента любой киберфизической системы распределена по всей системе и синергетически увязана с составляющими ее элементами, которые взаимодействуют, самонастраиваются и адаптируются к изменениям с помощью типовых программных интернет-протоколов. Поэтому к приоритетным задачам относятся диагностика и прогнозирование состояния пациентов с помощью расширяемых программных компонентов автоматизированного киберфизического мониторинга. Решение подобных задач имеет особое значение в стационарных условиях, поскольку здесь пациенты зачастую находятся в более тяжёлом состоянии, а количество мониторируемых показателей и частота их мониторинга существенно больше, чем в других условиях.

Одним из приоритетов совершенствования технологий оказания медицинской помощи в стационарных условиях является создание умных медицинских палат – медицинских киберфизических систем, интегрирующих передовые технологии сбора и анализа данных, информатизации и автоматизации процессов лечения, повышения безопасности пациентов и оптимизации рабочего процесса медицинского персонала. Умные палаты объединяют ряд инновационных инструментов: носимые и бесконтактные устройства,

датчики и сенсоры мониторинга состояния пациента и микроклимата в палате; электронные медицинские карты и цифровые аптеки; инструменты автоматизированного предоставления медицинским работникам полной и актуальной информации о пациентах в реальном времени и т.п. Объединяемые инструменты связаны между собой посредством различных программных интерфейсов и протоколов, таких как интерфейсы прикладного программирования (API), функционирующие по протоколу TCP/IP, протоколы Интернета вещей (IoT), специальные медицинские протоколы типа HL7 и т.д. Данные, поступающие от этих компонентов, интегрируются и анализируются для решения задач здравоохранения и медицины [9].

Целью проведенного исследования было изучение возможностей автоматизации киберфизического мониторинга состояния пациентов в умных палатах, реализуемого в виде концептов мобильных приложений с использованием облачных технологий для выполнения обработки данных.

2. Обзор мобильных медицинских приложений. Цифровое здравоохранение активно развивается в последние годы и охватывает такие направления, как электронное здравоохранение, мобильное здравоохранение, телемедицина и данные здравоохранения [3, 5]. e-Health (электронное здравоохранение) представляет собой использование информационных и коммуникационных технологий в сфере здравоохранения. Оно охватывает широкий спектр услуг, включая электронные медицинские записи, телемедицину, дистанционный мониторинг, системы поддержки принятия клинических решений и платформы для управления состоянием здоровья. Главные цели e-Health – улучшение качества медицинского обслуживания, повышение доступности услуг и оптимизация процессов в здравоохранении.

m-Health (mobile health; мобильное здравоохранение) – отрасль электронного здравоохранения e-Health, использующая мобильные технологии и устройства, а также беспроводные технологии передачи данных для улучшения здравоохранения и медицинских услуг (мониторинг, профилактика заболеваний, диагностика и пр.). m-Health включает в себя приложения для смартфонов, носимые устройства, телемедицину и системы мониторинга состояния здоровья. Это направление помогает в управлении заболеваниями, повышении доступа к медицинской помощи, а также в сборе и анализе данных о здоровье. Под управлением заболеваниями понимается способность человека с хроническими заболеваниями управлять симптомами, лечением, физическими и социальными последствиями, а также

изменениями в образе жизни для снижения числа обострений, осложнений, потребности в стационарной и скорой медицинской помощи. В настоящее время существует большое количество решений в этой области.

Исследователи во Франции разработали онлайн-инструмент для выявления потенциальных рецидивов рака легких на основе отчетов пациентов [7, 8]. Пациентов с раком легких просили оценивать свои симптомы (например, усталость, боль или изменения веса) с помощью короткой онлайн-формы каждую неделю. Алгоритм обрабатывает баллы симптомов и автоматически отправляет электронное письмо онкологу пациента, когда эти баллы указывают на потенциальный рецидив. Этот процесс позволяет онкологам предпринимать немедленные действия (например, контрольные звонки или визуализационные тесты) при первых признаках рецидива [10]. В конечном итоге алгоритм был встроен в веб-приложение Moovcare®.

Исследование [11] ориентировано на систематический поиск и выявление мобильных приложений, способных решать некоторые задачи врачей общей практики. В процессе исследования авторы выделили 12 задач, 9 из которых могут быть потенциально реализованы средствами mHealth. Охарактеризуем возможности решения этих задач с помощью существующих мобильных приложений, упоминаемых в работах [7 – 11].

1. Применение структурированного подхода к сбору и исследованию данных.

Приложение Ada – анализирует симптомы и предоставляет персонализированную информацию о состоянии здоровья на основании медицинского словаря, включающего симптомы и синдромы тысяч расстройств и заболеваний. Пользователь получает индивидуальный отчет об оценке, в котором описываются риски здоровью и дальнейшие действия по их минимизации. Ключевой момент состоит в том, что приложение Ada не может поставить медицинский диагноз и не заменяет консультацию врача.

Приложение HealthDiary: Medical History помогает пользователям составлять персональную историю болезни перед посещением врача, таким образом, устраняя необходимость сбора этих данных медицинским специалистом.

2. Точная интерпретация результатов для постановки диагноза.

Приложение Doctor Diagnose Symptoms Check предназначено для информирования и повышения осведомленности пациентов об их состояниях (может помочь с анализом симптомов и диагностикой).

Приложение GB Diagnosis Medical App определяет предполагаемый диагноз на основе взаимодействия и ответов пользователей (предлагает смоделированный разговор между врачом и пациентом для сбора симптомов и индивидуальной информации для постановки диагноза): за четыре шага после ввода симптомов заболевания приложение предлагает наиболее вероятный медицинский диагноз или рейтинговый список вероятных диагнозов.

Приложение My diagnostic направлено на идентификацию заболеваний в базе данных на основе симптомов, введенных пользователями.

Приложение Rapid diagnosis-mental health предназначено для помощи в диагностике психических, эмоциональных или психологических состояний, которые можно дифференцировать на основе симптомов.

Приложение Symptomate – Symptom checker предоставляет оценку здоровья пользователя (указывает возможные причины симптомов, варианты лечения и предлагаемые лабораторные анализы), анализируя тысячи симптомов и связывая их с несколькими заболеваниями или физиологическими состояниями, выбирая ограниченное количество из сотен возможных. Symptomate подходит для предварительной оценки выраженности симптомов как у детей, так и у взрослых. Сервис не ставит диагноз.

Приложение WebMD предлагает проверенную врачом информацию о состоянии здоровья и интерактивные инструменты формирования перечня возможных диагнозов на основе основного симптома и краткий набор общих вопросов (данные анамнеза, принимаемые в настоящее время лекарства, характеристики качества жизни и т.п.).

Также в список входит приложение Ada, описанное выше в п. 1.

3. Демонстрация квалифицированного подхода к клиническому обследованию.

Приложение Runtastic Heart Rate используется для мониторинга частоты сердечных сокращений методом фотоплетизмографии (необходимо приложить палец к камере смартфона, после чего его «просвечивает» вспышка).

Приложение SkinVision проверяет наличие признаков рака кожи, используя камеру телефона для захвата изображения кожного поражения и оценки риска рака кожи у пользователя: пациент может сфотографировать пятна на коже и в течение 30 с получить информацию о риске рака кожи с рекомендациями, в том числе о необходимости посещения врача.

4. Демонстрация профессионального подхода к выполнению процедур (не решается с помощью приложений mHealth).

5. Внедрение соответствующих принципов принятия решений.

Приложение Gout Decision Aid предоставляет информацию пациентам с подагрой (может функционировать как инструмент, помогающий пользователю участвовать в принятии решений).

6. Принятие структурированного подхода к клиническому менеджменту.

Приложение Rapid diagnosis-mental health, описанное выше.

7. Использование надлежащим образом других специалистов и услуг (не решается с помощью приложений mHealth).

8. Оказание неотложной помощи.

Приложение Google Assistant предоставляет информационную поддержку пользователям в использовании функций телефона и при поиске информации в Интернете (в том числе, может предоставлять информацию о неотложных и экстренных состояниях и может направлять пользователей в ближайшее медицинское учреждение).

Приложение Siri работает на платформе iOS (устройства Apple) и функционирует как виртуальный помощник (с предоставлением информации о неотложных и экстренных состояниях и направлением пользователей в ближайшее медицинское учреждение).

9. Обеспечение возможности людям с хроническими заболеваниями улучшить свое здоровье.

Приложение Asthma Manager используется для наблюдения за астмой (может помочь отслеживать симптомы и управлять приемом лекарств). В настоящее время приложение интегрируется с приложением Health для iOS.

Приложение Blood Pressure Companion предназначено для мониторинга артериального давления (может регистрировать и анализировать артериальное давление, а также создавать графики, диаграммы и таблицы результатов).

Приложение mySugr – Diabetes Tracker Log представляет собой журнал самоконтроля при сахарном диабете 1-го, 2-го типа и гестационного сахарного диабета (может регистрировать и анализировать параметры заболевания, а также предоставлять обратную связь).

Приложение forDiabetes: diabetes selfmanagement app предназначено для управления диабетом (может отслеживать и контролировать основные данные по диабету, включая уровень

глюкозы в крови, HbA1c (гемоглобин A1c), артериальное давление и прием лекарств).

Приложение Pill Reminder and Medication Tracker by Medisafe используется для напоминания и контроля приёма лекарственных препаратов (напоминает пользователям о необходимости приёма и создаст ежедневные и ежемесячные отчеты о его соблюдении).

10. Управление сопутствующими проблемами со здоровьем у отдельного пациента.

Приложения Asthma Manager, Blood Pressure Companion, mySugr, forDiabetes: diabetes selfmanagement app, Pill Reminder and Medication Tracker by Medisafe.

11. Координация командного подхода к уходу за пациентами (не решается с помощью приложений mHealth).

12. Поддержка людей с учётом их индивидуального опыта в сфере здоровья, болезни и восстановления.

Приложение Appibuddy представляет собой платформу для здорового образа жизни (позволяет пользователям записывать поведение в отношении здоровья и предоставляет платформу, на которой они могут делиться своими действиями и учиться у других пользователей).

Приложение BECCA – Breast Cancer Support направлено на поддержку и помощь пользователям в полноценной жизни после рака груди (предоставляет советы по здоровью, информацию и записи в блогах, чтобы помочь пациентам обеспечить высокое качество жизни после лечения рака).

Приложение Food (lg) представляет собой пищевой журнал, а также трекер и анализатор питания (может анализировать рацион и калории по результатам обработки фотографий еды).

Приложение HealthHub: Track and Improve направлено на достижение индивидуальных целей в отношении здоровья с помощью инструментов, планов действий и руководств по здоровому образу жизни (может фиксировать поведение в отношении здоровья и формировать индивидуальные планы действий пользователя на основе его целей).

Приложение HealthWatch направлено на предоставление практических инструментов для поддержания и улучшения здоровья и качества жизни, а также снижения рисков стресс-индуцированных заболеваний (обеспечивает обучение пациентов, включая регистрацию уровня стресса и предоставляя обратную связь по его коррекции).

Приложение Healthy365 способствует ведению здорового образа жизни (позволяет пользователям отслеживать ежедневные шаги и рассчитывать количество сожженных калорий).

Приложение The circle of health направлено на укрепление здоровья сердечно-сосудистой системы (может измерять и оценивать показатели здоровья сердечно-сосудистой системы и мотивировать пользователей поддерживать здоровый образ жизни).

Большинство названных приложений не могут заменить консультацию врача общей практики и подходят только для сбора анамнеза, предоставления первичной информации и рекомендаций, связанных со здоровьем. Другие компетенции врачей общей практики включают личные качества, практические навыки и гибкие навыки, которые приложения mHealth в настоящее время заменить не могут. Тем не менее, приложения mHealth могут поддерживать врачей общей практики с точки зрения обучения и рекомендаций.

В работе [12] представлены результаты контент-анализа описаний, функций и отзывов пользователей для 168 приложений, доступных в коммерческих цифровых магазинах, которые предназначены для медицинских работников и студентов. Особое внимание уделено мобильным приложениям, включающим искусственный интеллект, помогающий принимать диагностические решения. В исследовании приложения классифицированы по девяти категориям на основе их содержания и полезности: большинство (28%) приложений были «Справочниками/руководствами», которые в основном использовались студентами-медиками для подготовки к сдаче экзаменов, и «Инструментами для дифференциальной диагностики» (20,8%), применяемыми врачами при решении задач клинической дифференциальной диагностики.

Классификация приложений на основе использования искусственного интеллекта (17 приложений) определила «Калькуляторы» и «Инструменты для управления пациентами» как наиболее популярные типы приложений искусственного интеллекта, предлагающие диагностические возможности и оптимизацию лечения, а также минимизирующие время клинических испытаний и стоимость лечения [13].

Авторы работы [14] предлагают систему голосового управления для взаимодействия с оборудованием в умной палате, что повышает качество и удобство оказания медицинской помощи [15 – 17]. Система голосового управления подходит для всех пациентов и особенно эффективна для послеоперационных и инвалидов [18]. Пациенты

могут лежать на кровати и управлять оборудованием, произнося команды для мобильных устройств.

В [19] представлена инновационная интеллектуальная система ухода за пациентами. Ее основная функция заключается в уведомлении медсестер о том, что пациенту нужна помощь. В этой системе все сигналы могут передаваться на мобильные телефоны медсестер, тем самым обеспечивая немедленную связь и трехэтапное оповещение об уходе с постели, поскольку падения пациентов представляют собой серьезную проблему, а их предотвращение является одной из важнейших ожидаемых функций системы ухода [20, 21].

Авторы [22] предлагают систему контроля уровня жидкости в капельницах для внутривенного вливания, которая может использоваться в отделениях интенсивной терапии и послеоперационных отделениях. Система обмена сообщениями на основе GSM используется для оповещения персонала о необходимости замены флакона. Для измерения веса жидкости используется тензодатчик, а микроконтроллер считывает данные и отправляет их в модуль GSM [23]. Уровень жидкости во флаконе отображается в режиме реального времени в приложении Android. На дисплее показывается соответствующий процент жидкости, оставшейся во флаконе, и, когда уровень опускается ниже 100 мл, появляется предупреждение и повторно отправляется SMS-сообщение обслуживающему персоналу с периодом в 10 секунд до тех пор, пока система не будет отключена, и одновременно раздается звуковой сигнал. Кроме того, группа медицинских работников может контролировать уровень жидкости в режиме реального времени на своем рабочем месте с помощью модуля IoT. В случае, если обслуживающий персонал не может немедленно добраться до пациента по тревоге, отток жидкости автоматически блокируется электромагнитным клапаном.

В исследовании [24] рассматривается система локализации пациента. Достижения в технологиях локализации без устройств (DFL), основанных на методах мягких вычислений и позволяющих определять местоположение целей, не несущих никакого оборудования, в сочетании с технологиями машинного обучения, сделали локализацию более точной. Авторы используют сигналы Wi-Fi в отделениях и осуществляют локализацию пациента, сохраняя конфиденциальность, применяя многомасштабные сверточные нейронные сети (CNN) и модели долгосрочного/краткосрочного запоминающего устройства (LSTM). Результаты демонстрируют

высокую точность локализации. Кроме того, система допускает масштабирование для обнаружения чрезвычайных ситуаций, требующих быстрого реагирования медицинского персонала.

В работах [25] предлагается мобильное приложение для диагностики, стратификации и подбора лечения для пациентов с синдромом пароксизмальной симпатической гиперактивности (ПСГА). Приложение опирается на медицинские методики, основанные на шкалах. Несмотря на то, что эта разработка решает частную задачу, она имеет потенциал для развития, так как обладает гибкостью для внедрения шкал тяжести и прогноза развития различных заболеваний и строится как интегрируемый компонент умной медицинской палаты, направленный на автоматизацию киберфизического мониторинга состояния пациентов.

Интеллектуальные приложения являются ценными инструментами медицинской диагностики, особенно для определения состояния пациента и реализации его мониторинга – однако они не заменяют медицинских работников и не освобождают от очной консультации врача.

Существующие решения не вполне адаптированы для интеграции с другими компонентами медицинской инфраструктуры, поскольку являются отдельными приложениями или устройствами; также нет расширяемых и гибких решений для диагностики и рискометрии заболеваний по методикам на шкалах, но существуют частные (например, приложение для диагностики ПСГА). В проведенном исследовании предлагается подход к автоматизации киберфизического мониторинга состояния пациентов в умных палатах, основанный на гибком интегрируемом и расширяемом приложении для мониторинга состояния пациента при наличии некоторых ограничений. Во-первых, предполагается существование детерминированных методик для диагностики и рискометрии состояния здоровья, приводимых к общему виду на основе шкал. Во-вторых, подобные решения не ставят целью замену медицинского персонала, что невозможно как по профессиональным, так и по юридическим причинам, а только частичную автоматизацию медицинских услуг и, как следствие, повышение их эффективности и качества.

3. Материалы и методы исследования. Проводимое исследование направлено на устранение противоречия между потенциальными возможностями киберфизических систем, определяемыми уровнем научно-технического прогресса, и отсутствием инструментария исследования автоматизации

киберфизического мониторинга состояния здоровья пациентов в умных палатах. Математическая формулировка решаемой задачи может быть представлена как формирование и реализация управляющих воздействий для i -го пациента (U_i) так, чтобы:

$$U_i \rightarrow \min_{\substack{t \rightarrow \min \\ r \rightarrow \min}} (X_i(t, r) - X_i^C(t, r)),$$

где X_i – текущее состояние i -го пациента, X_i^C – целевое состояние (индивидуальная норма) i -го пациента, t – время пребывания в стационаре, r – ресурсы медицинской организации. По сути это типовая задача оказания стационарной медицинской помощи, но ее отличительной особенностью является ориентация на применение для формирования и реализации управляющих воздействий U_i технологий киберфизических систем, обеспечивающих дополнительные возможности минимизации величин t и r .

При проведении исследования использовались методы системного анализа (абстрагирование и конкретизация, анализ и синтез, индукция и дедукция, формализация и конкретизация, композиция и декомпозиция, реинжиниринг и алгоритмизация), технологии киберфизических систем, теория вероятностей, теория систем массового обслуживания и теория полумарковских процессов. В отличие от ранее известных, полученные результаты ориентированы на автоматизацию киберфизического мониторинга состояния пациентов в умных палатах, отличительной особенностью которого является реализация управляющих воздействий, направленных на персонализированную оптимизацию стационарного лечения, основанную на мониторинге динамики показателей состояния.

Решаемая задача, в отличие от классической оптимизационной задачи, учитывает ситуацию параллельного назначения терапевтических процедур врачами различных специализаций. То есть мониторинг состояния пациента и управляющих воздействий (терапевтических процедур) осуществляется территориально распределенно и взаимоувязанно, что обеспечено возможностями киберфизических технологий, вычислительные компоненты которых распределены и синергетически увязаны с составляющими их элементами (датчики состояния пациента, цифровые результаты диагностики и терапии, датчики окружающей среды). Элементы киберфизических систем взаимодействуют, самонастраиваются и адаптируются с помощью программных интернет-протоколов.

4. Сценарий применения расширяемых программных компонентов автоматизированного киберфизического мониторинга здоровья пациентов в умных палатах. Одним из программных компонентов автоматизированного киберфизического мониторинга состояния пациентов, часто применяемых на практике, являются оценочные шкалы, позволяющие анализировать состояние пациентов с представлением результатов в виде количественной оценки, определяющей тяжесть состояния и/или выраженность симптомов или синдромов потенциально опасных состояний. Примерами являются шкала для оценки синдрома ПСГА [25], шкала GRACE для рискометрии летальности и развития инфаркта миокарда [26, 27], шкала комы Глазго [28, 29] и т.п.

Рассмотрим сценарий использования программных компонентов сбора, обработки и анализа информации при автоматизированном киберфизическом мониторинге состояния пациентов с помощью методик, основанных на шкалах. Этот сценарий включает сбор исходных данных (набора показателей жизненно важных функций, симптомов, признаков), выполнение расчётов (например, вычисление оценки каждого показателя в отдельности), расчет интегральной оценки, формирование ее интерпретации и рекомендаций на основе полученных результатов. Реализация типового сценария предполагает:

1. Построение базовой методики оценки состояния пациента, ориентированной на балльную оценку, получаемую с применением шкал.
2. Синтез алгоритмов расчета частных и интегральных оценок.
3. Создание программных компонентов сбора и обработки информации.
4. Проведение валидации и верификации полученных результатов.

Базовая методика описывается кортежем, объединяющим четыре множества:

$$M = \langle I, S, C, R \rangle.$$

Множество I – это набор используемых показателей состояния пациентов, которые разделяются на три группы:

1. Первичные показатели служат исходными данными для расчета оценки. Они могут измеряться (например, пульс, систолическое и диастолическое артериальное давление, температура),

оцениваться (степень покраснения кожи, потливость), загружаться из базы данных приложения или из внешних источников, таких как электронные медицинские карты, медицинские информационные системы (ретроспективные данные). Внесение значений показателей в компьютерное приложение может осуществляться вручную специалистом, автоматически через «умные» устройства (мониторы) или сетевые интерфейсы (при загрузке из баз данных).

2. Промежуточные показатели рассчитываются на основе первичных или других промежуточных показателей. Примерами могут быть локальная оценка гипергидроза в баллах или индекс Кердо, который является функцией артериального давления и пульса. Промежуточные показатели используются для расчёта результирующих.

3. Результирующие показатели используются непосредственно для оценки состояния пациента и принятия врачебного решения. Примерами могут быть сумма баллов по шкале или оценка тяжести симптоматики заболевания.

Множество S – это набор правил, задающих порядок действий для определения значений показателей из множества I . Каждое правило характеризуется набором необходимых и достаточных условий для его применения, приоритетом и непосредственно выполняемым действием. Такие правила разделяются на виды:

1. Правила сбора данных применяются для определения начальных показателей и предусматривают спецификацию источника данных (метода ввода информации). Доступ к источникам данных осуществляется через плагины приложения, детали запроса определяются параметрами (к примеру, для внешней базы данных это может быть SQL-запрос).

2. Правила расчёта применяются для расчёта промежуточных и результирующих показателей. Они определяют зависимость начисления баллов от значений показателей текущего состояния и истории болезни (анамнеза) пациента и могут содержать инструкции типа «прибавить к итоговой оценке 1 балл, если температура превышает 37°C», «прибавить к итоговой оценке 1 балл, если симптом регистрируется повторно».

Множество C – это набор условий, определяющих, требуется ли выполнить то или иное действие или выдать рекомендацию в определенных ситуациях (например, «подать сигнал тревоги, если оценка более 10 баллов», «рекомендовать обратиться к врачу, если оценка более 15 баллов»).

Множество R – это набор действий, результатов и рекомендаций (например, «отправить сообщение врачу», «зафиксировать, что вероятность заболевания высока», «оценить течение заболевания как тяжёлое»).

Реализация изложенной базовой методики оценки состояния пациента, ориентированной на использование балльных шкал, в виде программных компонентов киберфизических систем требует наличия модели данных. Разработанная модель данных в минимальной конфигурации изображена на рисунке 1.

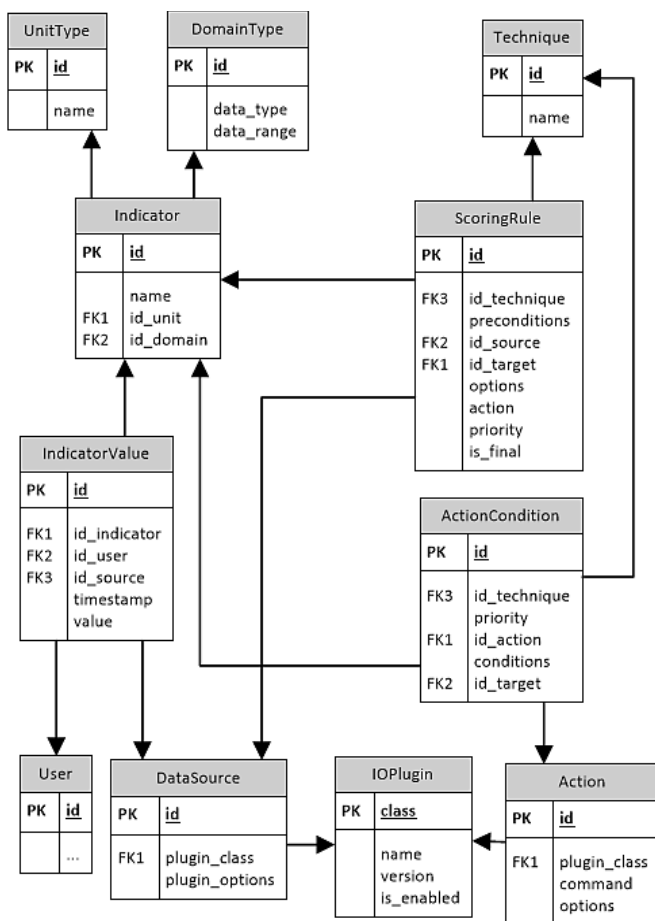


Рис. 1. Модель данных для методики оценки состояния здоровья пациента с применением шкал

В этой модели методика (Technique) непосредственно включает множество правил S (ScoringRule) и условий C (ActionCondition), множества показателей I (Indicator) и действий R (Action) задаются опосредованно. Итоговая оценка также рассматривается как один из показателей здоровья (Indicator). Правила расчёта (ScoringRule) определяются предварительными условиями, приоритетом и операциями, которые представлены математическими выражениями. Например, сложение баллов в зависимости от температуры по методике, изложенной в [25], описывается набором правил, которые строятся с использованием синтаксиса, подобного языку SQL (таблица 1).

Таблица 1. Пример набора правил ($[I_S]$ и $[I_T]$ означают исходный и целевой показатели соответственно)

Исходный показатель	Целевой показатель	Условие	Операция
[Температура тела] ($^{\circ}\text{C}$)	[Общий балл]	$[I_S] \geq 37 \text{ and } [I_S] < 38$	$[I_T] + 1$
		$[I_S] \geq 38 \text{ and } [I_S] < 39$	$[I_T] + 2$
		$[I_S] \geq 39$	$[I_T] + 3$
[Гипергидроз] (0=нет, 1=лёгкий, 2=умеренный, 3=тяжёлый)	[Общий балл]	True	$[I_T] + [I_S]$

Такая модель, помимо определения индикаторов, правил и действий, позволяет подключать плагины ввода и вывода. Это необходимо с точки зрения интеграции с датчиками, модулями умных помещений, базами данных и другими внешними системами и позволяет использовать приложение не только как отдельный модуль, но и как интегрированный компонент инфраструктуры умных медицинских палат. Основные типы входных плагинов:

1. Нет (по умолчанию): требуется ручной ввод данных.
2. Плагин расчета: значение индикатора определяется как производная от других индикаторов. Примером может служить индекс Кердо, который рассчитывается на основе пульса и артериального давления.
3. Плагины для подключения к датчикам: реализуют автоматический ввод данных из различных измерительных приборов.
4. Плагины для баз данных: реализуют сбор данных из внешних источников (баз данных и медицинских информационных систем).

Основные типы плагинов вывода:

1. Нет (по умолчанию): данные отображаются на экране пользовательского устройства.
2. Плагины для баз данных: передают данные во внешние источники (базы данных и медицинские информационные системы).
3. Плагины для отправки сообщений (например, по электронной почте).

Алгоритм оценки состояния пациента с использованием построенной модели данных представлен на рисунке 2.

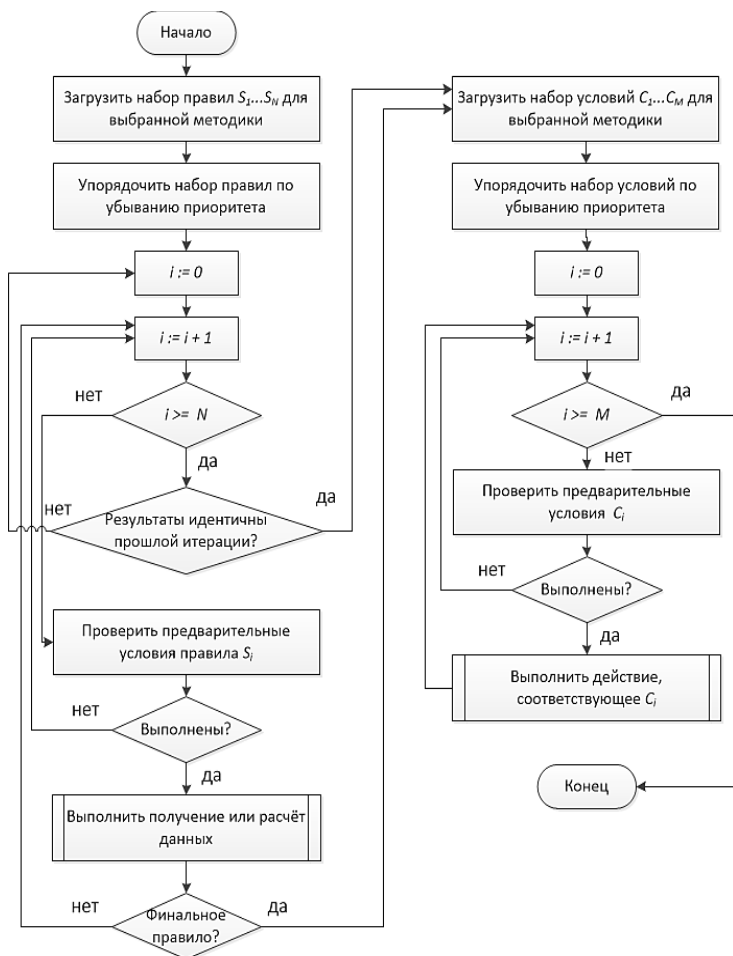


Рис. 2. Алгоритм оценки состояния пациента

Подпроцессы алгоритма связаны с операциями, которые выполняются через плагины программных компонентов киберфизического мониторинга здоровья (в частном случае, мобильного приложения) – сбор данных, обработка данных и вывод результатов.

Ввод данных и подключение плагинов производится на базе типового программного компонента автоматизированного киберфизического мониторинга состояния пациентов в умных палатах по сбору и обработке данных. Прототип такого программного компонента рассмотрен в работе [25]. Этот прототип разработан в виде мобильного приложения, написанного в среде Android Studio Electric Eel 2022.1.1 на языке Dart с использованием кроссплатформенного фреймворка Flutter, поэтому возможно выполнение сборки для различных типов хостинговых устройств и операционных систем. Серверная часть приложения реализована с применением стандартного стека LAMP (Linux, Apache, MySQL, PHP). Взаимодействие между клиентом и сервером выполняется с помощью библиотеки http, входящей в Dart. Тестирование приложения выполнялось на планшетном компьютере Samsung Galaxy Tab A7 Lite с операционной системой Android 11.

Учитывая вышеизложенные вспомогательные модели и концепции, результирующий метод создания программного компонента автоматизированного киберфизического мониторинга состояния пациентов в умных палатах состоит из следующих шагов:

1. Определить исходные данные: показатели; правила расчета; правила получения результатов.
2. Определить необходимые входные плагины для каждого индикатора, при необходимости их разработать.
3. Определить необходимые плагины для вывода данных, при необходимости их разработать.
4. Собрать типовый программный компонент (приложение) с подключенными плагинами для необходимой платформы.
5. Установить приложение.
6. Наполнить базу данных приложения информацией об индикаторах и правилах.

Для анализа эффектов от использования приложения программного компонента автоматизированного киберфизического мониторинга состояния пациентов выполнено моделирование его применения в умной палате при диагностике состояния пациентов с применением рассмотренных шкал.

5. Моделирование процессов автоматизированного киберфизического мониторинга состояния пациентов в умных палатах. В ходе исследования предложен подход к моделированию процессов автоматизированного киберфизического мониторинга состояния пациентов в умных палатах. Предложенный подход объединяет концепции персонализированного управления лечением, стохастического анализа и моделирования альтернативных маршрутов медицинских алгоритмов. Это, в свою очередь, позволяет реализовать непрерывный мониторинг состояния пациента и, при необходимости, скорректировать стратегию и тактику его лечения в масштабе времени, близком к реальному.

Процесс пребывания пациента в стационаре предполагает выполнение лечебно-диагностических мероприятий, направленных на восстановление, укрепление и сохранение его здоровья. Реализация таких мероприятий осуществляется в соответствии с врачебными предписаниями в виде последовательности воздействий, направленных на улучшение состояния пациента. Осуществление автоматизированного мониторинга состояния пациентов в умных палатах с применением технологий киберфизических систем открывает новые возможности для цифрового здравоохранения [2].

Наличие электронной медицинской карты пациента позволяет автоматически получить множество показателей, описывающих его состояние, интерпретируемое как «индивидуальная норма» [5]. При поступлении на стационарное лечение фиксируется описание начального состояния, которое (так как пациент нуждается в стационарном лечении) отличается от индивидуальной нормы. А цель стационарного лечения определяется как «приведение начального состояния пациента к состоянию индивидуальной нормы за минимальное время с достижением максимально выраженного длительного эффекта».

С учетом индивидуальных особенностей пациентов множества показателей состояния, характеризующих индивидуальную норму для разных пациентов, существенно различаются. И при выявлении у человека новых заболеваний множество показателей его состояния изменяется – как правило, расширяется.

После поступления пациента на стационарное лечение определяется диагноз и медицинский алгоритм, определяющий пошаговый протокол для решения задач врачебной практики. Большинство медицинских алгоритмов являются разветвленными – содержащими главный (описывающий наиболее желательное развитие событий или наиболее легкое и благоприятное течение болезни) и

побочные (описывающие нежелательное развитие событий или менее благоприятное, осложненное течение болезни) пути достижения целевого состояния. Каждый путь предполагает реализацию конвейера мероприятий, под которыми понимаются последовательно осуществляемые медицинские процедуры.

Формирование управляющих воздействий, направленных на персонализированную оптимизацию стационарного лечения, должно основываться на объективной динамике изменения показателей состояния. Для этого реализуется следующий подход:

- при поступлении пациента на стационарное лечение определяют модули разности каждого i -го показателя для начального (xn_i) и целевого состояния (xc_i): $xr_i = |xn_i - xc_i|$;
- все определенные разности делят на планируемый (в сутках) срок стационарного лечения (sl), получая величины: $\Delta_i = xr_i / sl$;
- целевое текущее состояние пациента в каждые сутки стационарного лечения (xt_i) определяется как сумма или разность между каждым показателем на момент поступления пациента и в текущие сутки лечения: $xt_i = xn_i \pm sl \times \Delta_i$ (знак «+» или «-» определяется в зависимости от направленности шкалы изменения показателя).

Разность между рассчитанным xt_i и реально достигнутым xt_i значением каждого показателя состояния определяет «штрафные санкции», которые определяют требуемые, согласно решению врача, изменения в стратегии и тактике стационарного лечения пациента, необходимые для достижения целевого состояния.

Таким образом, медицинский алгоритм можно представить как последовательную кооперацию звеньев роя (набора медицинских процедур и манипуляций), осуществляющего конвейерное взаимодействие, при котором следующее звено роя активизируется после того, как предыдущее звено закончит свою часть работы. На практике рой может действовать как последовательно, так и параллельно (например, пациенту выполняют массаж, и на него действует ранее принятое лекарство) с другими роями, также имеющими конвейерную организацию. Конвейерная организация роев характеризует достижение перехода от начального до целевого состояния пациента в результате стационарного лечения, при этом промежуточные состояния, как правило, определяются в каждые сутки лечения (например, в ходе ежедневных медицинских обходов) [30 – 32].

Для моделирования задач автоматизированного киберфизического мониторинга состояния пациентов в умных палатах вполне естественно потребовать, чтобы эффективность стационарного лечения оценивалась не только как результат мониторинга состояния пациента при поступлении и при выписке, а на основании ежесуточной динамики изменения состояния пациента. В противном случае нельзя, например, исключить манипуляций, направленных на экстренную медикаментозную нормализацию состояния пациента ради его плановой выписки.

Поскольку в большинстве случаев количество суток планового стационарного лечения известно, закон распределения времени нормализации состояния пациента является ключевым контролируемым параметром. То есть имеется тривиальное решение задачи стационарного лечения – привести начальное состояние пациента к целевому с максимальной скоростью (за минимальное время). Но на практике – вследствие сложности понимания механизмов регуляции гомеостаза и индивидуальных особенностей пациента – такое решение реализуется очень редко. Это приводит к необходимости постоянной коррекции реализуемого медицинского алгоритма, поэтому модель процесса киберфизического мониторинга состояния пациентов должна быть стохастической [33].

Наличие альтернативных стратегий и тактик управления здоровьем создает предпосылки для возникновения игровых ситуаций, в которых можно управлять не только временем стационарного лечения, но и вероятностями (в зависимости от сочетания диагнозов) или периодичностью выбора маршрута медицинского алгоритма. То есть при моделировании расширяемых программных компонентов автоматизированного киберфизического мониторинга состояния пациентов возникает задача анализа эстафет с альтернативными маршрутами [30, 34]. Начальным (стартовым) пунктом эстафеты является начальное состояние пациента (поступление на стационарное лечение), конечным пунктом – состояние пациента при выписке (в идеале – целевое состояние). Схему эстафеты с альтернативными маршрутами можно представить в виде графа, узлы которого представляют собой пункты эстафеты (как правило, сутки стационарного лечения), а дуги графа – альтернативные маршруты реализуемого медицинского алгоритма.

В общем случае в эстафетах участвуют M роев, которые функционируют независимо друг от друга. Рои должны преодолеть расстояние от поступления до выписки пациента за время, не превышающее время планового стационарного лечения. Это

расстояние делится на J этапов (как правило, сутки лечения $J=sI$), а j -й этап m -го роя включает $K(m, j)$ маршрутов. После завершения j -го этапа j -й единицей m -го роя, $(j+1)$ -я единица может выбрать один из $K(m, j)$ возможных маршрутов $(j+1)$ -го этапа. Первые единицы сразу стартуют по одному из возможных маршрутов первого этапа (врачебные назначения при поступлении пациента), а время прохождения маршрутов этапов единицами является случайным, определяется для них индивидуально и описывается плотностью вероятности. После завершения j -го этапа j -м звеном $(j+1)$ -е звено сразу приступает к реализации $(j+1)$ -го этапа. Победой в соревновании роев на конкретном этапе считается достижение показателями текущего состояния пациента значений, не хуже требуемых, а штраф определяется нормированной величиной разности между требуемым и реально достигнутым значением i -го показателя.

С учетом допущений показано, что модель эстафет с несколькими альтернативными маршрутами может быть представлена как M -параллельный полумарковский процесс [30, 31]:

$$\mu = \{A, \mathbf{h}(t)\},$$

где t – время; A – множество состояний; $\mathbf{h}(t)$ – полумарковская матрица;

$$A = \bigcup_{m=1}^M A^m, \quad A^m = \{a_j^m\},$$

$$A^m \cap A^n = \emptyset, \text{ если } m \neq n;$$

где a_0^m – являются исходными, а a_j^m – поглощающими состояниями M -параллельного процесса. Состояния модели можно соотнести с состоянием пациента, а маршруты на графе – с динамикой состояния.

В свою очередь m -я полумарковская матрица имеет вид:

$$\mathbf{h}^m(t) = [\mathbf{h}_{j,k}^m(t)],$$

$$\text{где } \mathbf{h}_{j,k}^m(t) = \begin{cases} \mathbf{h}_j^m(t), & \text{если } k = j+1; \\ 0, & \text{иначе;} \end{cases}$$

$$\mathbf{h}_j^m(t) = \left[h_{k(m,j)}(t) \right], \quad h_{k(m,j)}(t) = p_{k(m,j)}(t) f_{k(m,j)}(t),$$

$$p_{k(m,j)} = \int_0^\infty h_{k(m,j)}(t) dt, \quad p_{k(m,j)} = 0, \quad f_{k(m,j)}(t) = \lim_{\tau \rightarrow \infty} \delta(t - \tau),$$

$p_{k(m,j)}$ – вероятность выбора маршрута j -й единицей m -го роя;
 $f_{k(m,j)}(t)$ – плотность времени преодоления маршрута j -й единицей m -го роя. Процесс перехода из состояния S_{j-1}^m в состояние a_j^m по любому маршруту в полумарковском процессе описывается как [30]:

$$\mu^m = \{A^m, \mathbf{h}^m(t)\}.$$

В связи с тем, что все возможные переключения составляют полную группу несовместных событий, плотность времени пребывания процесса μ^m в состоянии a_{j-1}^m имеет следующий вид [32, 34]:

$$f_j^m(t) = \sum_{k(m,j)=1}^{k(m,j)} h_{k(m,j)}(t), \quad \sum_{k(m,j)=1}^{K(m,j)} p_{k(m,j)}(t) = 1.$$

Из всех возможных маршрутов блуждания по процессу μ^m выбирается маршрут:

$$s_{l(m)} = \left[k(m, j) \right],$$

а общее количество выбранных маршрутов равно

$$K(m) = \prod_{j=1}^J K(m, j).$$

Взвешенная плотность времени нахождения пациента в состоянии a_0^m до перехода в состояние a_j^m по выбранному маршруту выглядит следующим образом [34]:

$$h_{k(m)}(t) = L^{-1} \left[\prod_{j=1}^J L[h_{k(m,j)}(t)] \right],$$

а вероятность прохождения маршрута можно оценить как [30]:

$$p_{k(m)} = \int_0^{\infty} h_{k(m)}(t) dt = \prod_{j=1}^J p_{k(m,j)}.$$

Плотность вероятности для чистого времени перехода пациента из состояния a_0^m в состояние a_J^m по выбранному маршруту имеет вид [31]:

$$f_{k(m)}(t) = \frac{h_{k(m)}(t)}{p_{k(m)}},$$

а плотность времени такого перехода по любому возможному произвольному маршруту имеет вид [30]:

$$f_{\Sigma}^m(t) = \sum_{k(m)=1}^{K(m)} h_{k(m)}(t).$$

Представленные соотношения обеспечивают моделирование процессов автоматизированного киберфизического мониторинга состояния пациентов в умных палатах. Накопление массива репрезентативной информации о реализации такого мониторинга для различных групп пациентов обеспечит возможность формирования рекомендаций по совершенствованию стандартов оказания стационарной медицинской помощи пациентам с различными заболеваниями.

Практическая реализация предлагаемого подхода к моделированию процессов автоматизированного киберфизического мониторинга состояния пациентов в умных палатах требует решения комплекса задач, связанных со структурной и параметрической идентификацией, а также с верификацией, основанной на результатах практического применения. Это, в свою очередь, требует наличия сценариев моделирования с учетом применения расширяемых программных компонентов автоматизированного киберфизического

мониторинга состояния пациентов. В наиболее простом случае такие сценарии ориентированы на применение программных компонентов мониторинга состояния пациентов, основанных на использовании оценочных шкал.

6. Моделирование приложения для оценки состояния пациентов с помощью шкал. Наиболее простым сценарием использования приложения с помощью рассмотренных шкал является оценка степени нарушения сознания и комы для детей старше четырех лет и взрослых по шкале Глазго. Эта шкала предполагает определение трёх качественных показателей состояния пациента – открывание глаз (Eye response – E), речевая реакция (Verbal response – V) и двигательная реакция (Motor response – M), после чего выполняется подсчёт баллов и интерпретация результата в диапазоне от ясного сознания до глубокой комы и терминальной комы (смерти мозга). Все показатели оцениваются экспертно, среднее время тестирования согласно методике [35] составляет 1-2 минуты. В этом случае приложение может применяться только для расчёта баллов, оценки состояния и фиксации результата, но не для сбора данных. Процесс показан на рисунке 3, где серым цветом отмечены автоматизируемые действия.



Рис. 3. Сценарий использования шкалы Глазго

Более сложным примером является рискотметрия смерти при остром коронарном синдроме по данным, полученным при поступлении в стационар. Шкала GRACE (Global Registry of Acute Coronary Events) позволяет оценить риск смерти и развития инфаркта миокарда как на госпитальном этапе, так и в течение последующего полугодия, а также определить рациональный медицинский алгоритм лечения конкретного больного. Шкала GRACE предусматривает оценку 8 показателей, получаемых из разных источников: карточка пациента (возраст, результаты анализов), экспертная оценка (класс сердечной недостаточности по классификации Killip, девиация сегмента ST электрокардиограммы), измерение (частота сердечных сокращений, систолическое артериальное давление). Соответствующий процесс показан на рисунке 4.

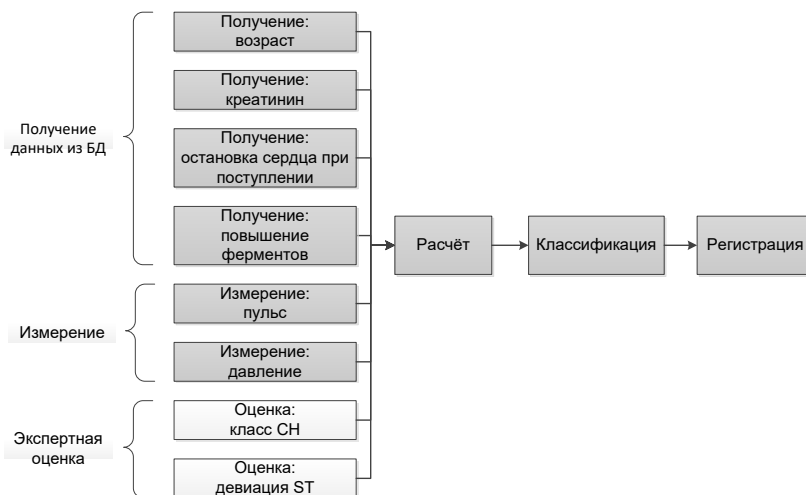


Рис. 4. Сценарий использования шкалы GRACE

Методика диагностики пароксизмальной симпатической гиперактивности также предполагает использование измеряемых и экспертно оцениваемых показателей, получаемых путём визуального осмотра пациента, а также загрузку ретроспективной информации о его состоянии. Применение этой методики рассмотрено в [36].

Для оценки времени, необходимого для выполнения указанных процессов, выполнено их имитационное моделирование.

7. Результаты моделирования. Моделирование выполнено для следующих вариантов диагностики состояния пациента по трём рассмотренным шкалам:

1. Без использования приложения: сбор данных и все расчёты выполняются специалистами вручную.
2. Приложение используется в автономном режиме: расчёты выполняются автоматически, но необходим ручной сбор и внесение всех показателей.
3. Приложение используется в режиме компонента умной палаты и способно получать значения некоторых показателей автоматически (например, с «умных» мониторов).

Для каждого случая проведено 1000 итераций моделирования, гистограммы распределения времени выполнения процессов для трёх шкал и трёх режимов использования приложения показаны на рисунках 5-7.

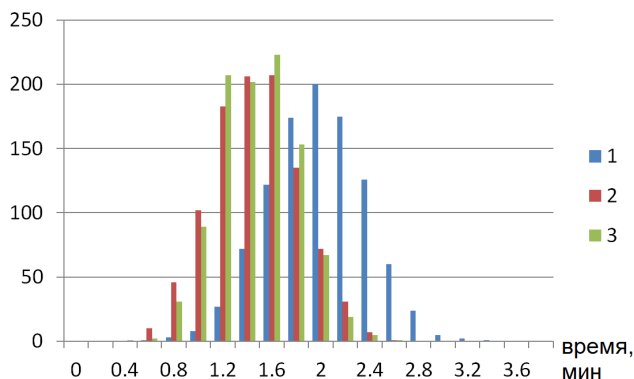


Рис. 5. Гистограмма времени, необходимого для оценки тяжести комы в трёх режимах

Из гистограммы (рисунок 5) видно, что использование приложения (режимы 2 и 3) позволяет сократить время определения тяжести комы со 110 до 80 с, то есть в среднем в 1,4 раза по сравнению с ручным способом расчёта (режим 1). При использовании шкалы Глазго оценки показателей определяются только экспертным путём, поэтому значимой разницы между режимами 2 (автономное использование приложения) и 3 (использование приложения в составе умной палаты) нет. В этом случае приложение используется только для обработки, хранения и визуализации данных.

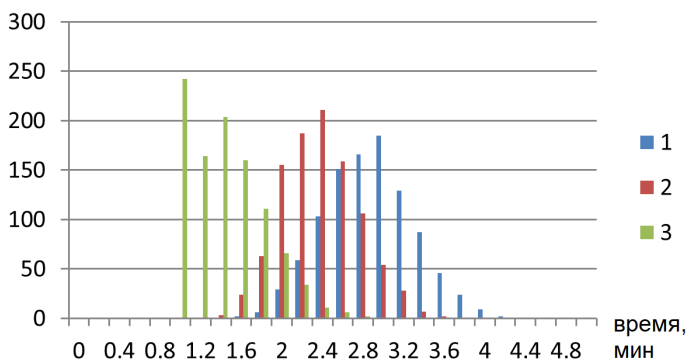


Рис. 6. Гистограмма времени, необходимого для оценки риска смерти при остром коронарном синдроме

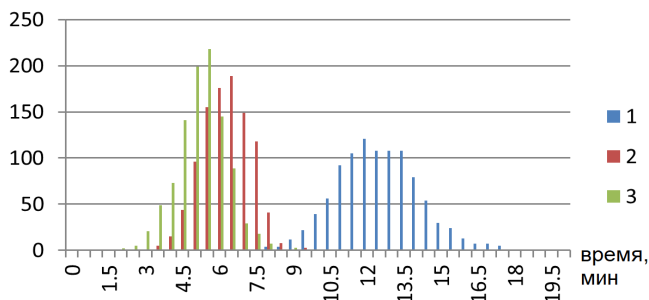


Рис. 7. Гистограмма времени, необходимого для оценки синдрома ПСГА

На рисунках 5-7 видно, что распределение времени для всех трёх режимов существенно отличается. В этом случае использование приложения позволяет сократить на несколько минут не только время расчётов, но и время получения данных за счёт их загрузки из информационных систем и получения из умных мониторов. Так, в задаче оценки ПСГА использование приложения в автономном режиме позволяет сократить время оценки в среднем примерно в 2 раза, при использовании мониторов - в 2,4 раза, т. е. с 12 минут до 6 и 5 минут соответственно. При использовании шкалы GRACE затрачиваемое время сокращается соответственно в 1,2 и 2,1 раза – с 3 минут до 1-2 минут. Такое отличие является значительным, поскольку время работы медицинских специалистов в процессе оценки состояния пациента сокращается более чем в 2 раза. Кроме того, повышается точность оценки за счёт снижения влияния человеческого фактора. Получена обратная связь от медицинских специалистов, подтверждающая результаты моделирования для диагностики ПСГА [25].

8. Заключение. Существующие решения в области m-Health, как правило, не адаптированы к медицинской инфраструктуре (медицинским информационным системам, компонентам умных палат и другим приложениям) и часто применяются изолированно. Кроме того, отсутствуют универсальные и расширяемые инструменты для диагностики и рискометрии заболеваний по шкальным методикам, за исключением узкоспециализированных приложений.

Предложенная математическая модель процессов автоматизированного киберфизического мониторинга состояния пациентов в умных палатах, описывающая медицинский алгоритм как эстафету с альтернативными маршрутами с помощью аппарата параллельных полумарковских процессов, позволяет реализовать персонифицированное управление здоровьем на основе оптимизации

лечебно-диагностических процедур с помощью целевой функции, характеризующей отклонение показателей текущего состояния от показателей целевого состояния пациента.

Исходя из этого, мобильные приложения описаны как элементы умной медицинской палаты и предложен подход к созданию гибкого, расширяемого и интегрируемого программного компонента автоматизированного киберфизического мониторинга состояния пациентов. Предложенный программный компонент базируется на детерминированных методиках диагностики и рискометрии здоровья, которые могут быть приведены к унифицированному формату на основе шкал. Это решение позволяет автоматизировать медицинские процессы, что способствует повышению эффективности и качества работы медицинского персонала, так как уменьшается время, затрачиваемое на рутинные операции, и снижается влияние человеческого фактора. Кроме того, использование облачных технологий позволит ускорить обработку данных и повысить гибкость программного компонента киберфизического мониторинга здоровья.

Для оценивания потенциальной эффективности приложения проведено моделирование его применения в умной палате при диагностике и рискометрии состояния пациентов с помощью методик, основанных на шкалах – оценка нарушения сознания и комы по шкале Глазго, риск летального исхода при остром коронарном синдроме по шкале GRACE и диагностика синдрома пароксизмальной симпатической гиперактивности. Моделирование выполнено для трех сценариев: без использования приложения, с его использованием в автономном режиме и с его использованием как компонента автоматизированного киберфизического мониторинга состояния пациентов в умных палатах. Результаты показали, что использование приложения в составе умной палаты позволяет не только ускорить расчёты, но и сократить время получения данных благодаря их автоматической загрузке из информационных систем и автоматизированному снятию показателей с медицинских мониторов.

Авторский коллектив считает важным почтить память выдающегося учёного в области информатики, информационных технологий и теории управления, доктора технических наук, члена-корреспондента РАН, заслуженного деятеля науки и техники РФ Рафаэля Мидхатовича Юсупова. Опыт личного общения с ним, оставивший глубокий и запоминающийся след, стал для авторов важной частью профессионального становления. Научная школа Р.М. Юсупова, отличавшаяся особой системностью и вниманием к строгому обоснованию решений, в значительной степени повлияла

на подходы, использованные при подготовке представленного исследования. Во многом благодаря этим методологическим ориентирам удалось обеспечить целостность и структурированность представленных результатов.

Литература

1. Шадеркин И.А. Дистанционный мониторинг состояния здоровья и окружающей среды человека: возможности и ограничения // Журнал телемедицины и электронного здравоохранения. 2022. Т. 8. № 3. С. 45–49. DOI: 10.29188/2712-9217-2022-8-3-45-54.
2. Богомолов А.В. Информационные технологии цифровой адаптивной медицины // Информатика и автоматизация. 2021. Т. 20. № 5. С. 1154–1182.
3. Ушаков И.Б., Богомолов А.В. Цифровая профилактическая медицина // Вестник Российской академии наук. 2024. Т. 94. № 11. С. 1003–1013. DOI: 10.31857/S0869587324110043.
4. Труханова И.Г., Гуреев А.Д., Лунина А.В. Влияние цифровой трансформации на повышение качества медицинского обеспечения и снижение врачебных ошибок // Биомедицинская радиоэлектроника. 2025. Т. 28. № 1. С. 83–90.
5. Зарубина Т.В. Актуальные вопросы внедрения информационных технологий в здравоохранении // Вестник Росздравнадзора. 2018. № 3. С. 20–25.
6. Zhukova N., Motienko A., Levonevskiy D., Subbotin A., Anh P.T. Smart room for patient monitoring based on IoT technologies // Proceedings of the 2023 6th Artificial Intelligence and Cloud Computing Conference. 2023. pp. 151–158. DOI: 10.1145/3639592.3639630.
7. Denis F., Viger L., Charron A., et al. Detection of lung cancer relapse using self-reported symptoms transmitted via an internet web-application: pilot study of the sentinel follow-up // Supportive Care in Cancer. 2014. vol. 22. pp. 1467–1473. DOI: 10.1007/s00520-013-2111-1.
8. Denis F., Viger L., Charron A., et al. Detecting lung cancer relapse using self-evaluation forms weekly filled at home: the sentinel follow-up // Supportive Care in Cancer. 2014. vol. 22(1). pp. 79–85. DOI: 10.1007/s00520-013-1954-9.
9. Лебедев Г.С., Фомина И.В., Артемова О.Р., и др. Обзор семейства стандартов по обмену данными с персональными медицинскими приборами // Вестник Росздравнадзора. 2022. № 6. С. 75–84.
10. Denis F., Lethrosne C., Pourel N., et al. Randomized trial comparing a web-mediated follow-up with routine surveillance in lung cancer patients // JNCI: Journal of the National Cancer Institute. 2017. vol. 109. no. 9. DOI: 10.1093/jnci/djx029.
11. Wattanapisit, A., Teo, C.H., Wattanapisit, S., et al. Can mobile health apps replace GPs? A scoping review of comparisons between mobile apps and GP tasks // BMC medical informatics and decision making. 2020. vol. 20. no. 1. DOI: 10.1186/s12911-019-1016-4.
12. Galetsi P., Katsaliaki K., Kumar S. Exploring benefits and ethical challenges in the rise of mHealth (mobile healthcare) technology for the common good: An analysis of mobile applications for health specialists // Technovation. 2023. vol. 121. DOI: 10.1016/j.technovation.2022.102598.
13. Хохлов А.Л., Зарубина Т.В., Котловский М.Ю., и др. Механизмы внедрения технологий искусственного интеллекта в здравоохранение: новые этические вызовы // Медицинская этика. 2024. Т. 12. № 3. С. 4–10.
14. Jian W.S., Wang J.Y., Rahmanti A.R., et al. Voice-based control system for smart hospital wards: a pilot study of patient acceptance // BMC health services research. 2022. vol. 22. no. 1. DOI: 10.1186/s12913-022-07668-1.

15. Brunete A., Gambao E., Hernando M., Cedazo R. Smart assistive architecture for the integration of IoT devices, robotic systems, and multimodal interfaces in healthcare environments // *Sensors*. 2021. vol. 21. no. 6. DOI: 10.3390/s21062212.
16. Azhiimah A.N., Khotimah K., Sumbawati M.S., Santosa A.B. Automatic Control Based on Voice Commands and Arduino // *Proceedings of the International Joint Conference on Science and Engineering (IJCSSE 2020)*. 2020. pp. 29–34. DOI: 10.2991/aer.k.201124.006.
17. Cronin S., Doherty G. Touchless computer interfaces in hospitals: A review // *Health informatics journal*. 2018. vol. 25. no. 4. pp. 1325–1342. DOI: 10.1177/1460458217748342.
18. Ali H., Cole A., Panos G. Transforming Patient Hospital Experience Through Smart Technologies // *Design, User Experience, and Usability. Case Studies in Public and Personal Interactive Systems: 9th International Conference (DUXU 2020), Held as Part of the 22nd HCI International Conference (HCII 2020)*. 2020. pp. 203–215. DOI: 10.1007/978-3-030-49757-6_14.
19. Wen M.H., Bai D., Lin S. et al. Implementation and experience of an innovative smart patient care system: A cross-sectional study // *BMC Health Services Research*. 2022. vol. 22. no. 1. DOI: 10.1186/s12913-022-07511-7.
20. Bouldin E.D., Andresen E.M., Dunton N.E. et al. Falls among adult patients hospitalized in the United States: prevalence and trends // *Journal of patient safety*. 2013. vol. 9. no. 1. pp. 13–17. DOI: 10.1097/PTS.0b013e3182699b64.
21. Hempel S., Newberry S., Wang Z., et al. Hospital fall prevention: a systematic review of implementation, components, adherence, and effectiveness // *Journal of the American Geriatrics Society*. 2013. vol. 61. no. 4. pp. 483–494. DOI: 10.1111/jgs.12169.
22. Yadav S., Manohar M., Jeba Shiney O., Priestly Shan B., Das G.J. A Smart System for Monitoring Flow in Drip Bottles in Healthcare // *Proceedings of the International Conference on Futuristic Communication and Network Technologies (VICFCNT 2020)*. 2022. pp. 663–669. DOI: 10.1007/978-981-16-4625-6_66.
23. Gayathri S., Ganesh C.S.S. Automatic indication system of glucose level in glucose trip bottle // *International Journal of Multidisciplinary Research and Modern Education*. 2017. vol. 3(1). pp. 148–151. DOI: 10.5281/zenodo.438622.
24. Feng Y.S., Liu H.Y., Hsieh M.H., Fung H.C., Chang C.Y., Yu C.C., Huang C.W. An RSSI-based device-free localization system for smart wards // *IEEE International Conference on Consumer Electronics-Taiwan (ICCE-TW)*. 2021. pp. 1–2. DOI: 10.1109/ICCE-TW52618.2021.9603249.
25. Ценципер Л.М., Мотиенко А.И., Терехов И.С., Левоневский Д.К., Самочерных К.А., Кондратьев А.Н. Цифровое решение для определения тяжести синдрома пароксизмальной симпатической гиперактивности у пациентов с повреждением головного мозга // *Вестник анестезиологии и реаниматологии*. 2023. Т. 20. № 6. С. 90–96. DOI: 10.24884/2078-5658-2023-20-6-90-96.
26. Tang E.W., Wong C.K., Herbison P. Global Registry of Acute Coronary Events (GRACE) hospital discharge risk score accurately predicts long-term mortality post acute coronary syndrome // *American heart journal*. 2007. vol. 153. no. 1. pp. 29–35.
27. Беззубцева Е.Н., Сахарова А.В., Лавлинская Л.И., Черных Е.А., Лавлинская Т.А. Оценка риска летальности по шкале GRACE у больных, перенесших инфаркт миокарда // *Молодежный инновационный вестник*. 2017. Т. 6. № 2. С. 302–303.
28. Пирадов М.А., Супонева Н.А., Рябинкина Ю.В. и др. Шкала комы Глазго (Glasgow Coma Scale, GCS): лингвокультурная адаптация русскоязычной версии // *Журнал им. НВ Склифосовского «Неотложная медицинская помощь»*. 2021. Т. 10. № 1. С. 91–99. DOI: 10.23934/2223-9022-2021-10-1-91-99.
29. Соловьева П.И., Синкин М.В., Талыпов А.Э. Шкала комы Глазго и шкала FOUR для оценки уровня сознания // *Нервные болезни*. 2023. № 3. С. 20–26. DOI: 10.24412/2226-0757-2023-13003.

30. Larkin E.V., Bogomolov A.V., Privalov A.N., Dobrovolsky N.N. Discrete model of paired relay-race // Bulletin of the South Ural State University. Ser. Mathematical Modelling, Programming & Computer Software. 2018. vol. 11. no. 3. pp. 72–84.
31. Larkin E.V., Privalov A.N., Bogomolov A.V. Discrete approach to simulating synchronized relay races // Automatic Documentation and Mathematical Linguistics. 2020. vol. 54. no. 1. pp. 43–51.
32. Larkin E.V., Bogomolov A.V., Privalov A.N., Dobrovolsky N.N. Relay races along a pair of selectable routes // Bulletin of the South Ural State University. Ser. Mathematical Modelling, Programming & Computer Software. 2018. vol. 11. no. 1. С. 15–26.
33. Кибзун А.И., Синицын И.Н. Современные проблемы теории оптимизации стохастических систем // Автоматика и телемеханика. 2020. № 11. С. 3–10.
34. Ларкин Е.В., Привалов А.Н., Богатырева Ю.И. Парное дискретное соревнование со свободным выбором маршрута // Чебышевский сборник. 2021. Т. 22. № 2(78). С. 145–159.
35. The Glasgow Coma Scale – GCS (Teasdale G.M., Jennett B.). URL: <https://cpd-program.ru/methods/gcsc.htm> (дата обращения 14.01.2025).
36. Левоневский Д.К., Мотиенко А.И. Моделирование и автоматизация процессов сбора и обработки данных в умной медицинской палате // Программная инженерия. 2023. Т. 14. № 10. С. 502–512. DOI: 10.17587/prin.14.502-512.

Левоневский Дмитрий Константинович — канд. техн. наук, доцент, старший научный сотрудник, лаборатория технологий больших данных социкиберфизических систем, Федеральное государственное бюджетное учреждение науки «Санкт-Петербургский Федеральный исследовательский центр Российской академии наук» (СПб ФИЦ РАН). Область научных интересов: технологии программирования и обработки данных, корпоративные социкиберфизические системы, моделирование и автоматизация корпоративных процессов. Число научных публикаций — 100. levonevskij.d@ias.spb.su; 14-я линия В.О., 39, 199178, Санкт-Петербург, Россия; р.т.: +7(812)508-3311.

Мотиенко Анна Игоревна — канд. техн. наук, старший научный сотрудник, лаборатория технологий больших данных социкиберфизических систем, Федеральное государственное бюджетное учреждение науки «Санкт-Петербургский Федеральный исследовательский центр Российской академии наук» (СПб ФИЦ РАН). Область научных интересов: математическая статистика, теория вероятности, применение методов математического моделирования в эпидемиологии. Число научных публикаций — 65. anna.gunchenko@gmail.com; 14-я линия В.О., 39, 199178, Санкт-Петербург, Россия; р.т.: +7(812)508-3311.

Богомолов Алексей Валерьевич — д-р техн. наук, профессор, главный научный сотрудник, лаборатория технологий больших данных социкиберфизических систем, Федеральное государственное бюджетное учреждение науки «Санкт-Петербургский Федеральный исследовательский центр Российской академии наук» (СПб ФИЦ РАН). Область научных интересов: цифровизация медицины, медицинская информатика, системы управления, диагностика состояния человека. Число научных публикаций — 350. a.v.bogomolov@gmail.com; 14-я линия В.О., 39, 199178, Санкт-Петербург, Россия; р.т.: +7(812)508-3311.

Поддержка исследований. Работа выполнена при финансовой поддержке государственного бюджета, номер проекта FFZF-2025-0019.

D. LEVONEVSKIY, A. MOTIENKO, A. BOGOMOLOV
**AUTOMATION OF CYBER-PHYSICAL MONITORING
OF PATIENTS' CONDITION IN SMART WARDS**

Levonevskiy D., Motienko A., Bogomolov A. Automation of Cyber-Physical Monitoring of Patients' Condition in Smart Wards.

Abstract. The digitalization of healthcare implies the widespread adoption of informatization and automation technologies for human health monitoring, aimed at maintaining, improving, and restoring health. To adapt existing e-health solutions to the medical infrastructure, we modeled expandable software components for automated cyber-physical patient monitoring in smart wards. A mathematical model of cyber-physical patient monitoring processes is presented, describing a medical algorithm as a relay race with alternative routes using a framework of parallel semi-Markov processes. The software components for cyber-physical patient monitoring are characterized as a flexible, expandable, and integrated element of a smart medical ward, enabling the personalization and optimization of treatment and diagnostic procedures using an objective function characterizing the deviation of the patient's current condition from their target condition. The potential for automating medical processes in the implementation of diagnostic and health risk assessment methods based on point scales is demonstrated. The simulation results showed that the use of software components for automated cyber-physical patient monitoring in smart wards not only accelerates calculations but also reduces the time required to obtain data through automatic downloading from information systems and automated readings from medical devices, opening up new opportunities for personalized management of inpatient medical care.

Keywords: healthcare informatization, medical informatics, automation of condition monitoring, smart ward, cyber-physical system, software component, cyber-physical monitoring.

References

1. Shaderkin I.A. [Remote monitoring of human health and the environment: opportunities and limitations]. *Zhurnal telemeditsiny i jelektronnogo zdravoohraneniya – Russian Journal of Telemedicine and E-Health*. 2022. vol. 8. no. 3. pp. 45–49. DOI: 10.29188/2712-9217-2022-8-3-45-54. (In Russ.).
2. Bogomolov A. [Information Technologies of Digital Adaptive Medicine]. *Informatics and Automation*. 2021. vol. 20. no. 5. pp. 1154–1182. (In Russ.).
3. Ushakov I.B., Bogomolov A.V. [Digital Preventive Medicine]. *Vestnik Rossijskoj akademii nauk – Bulletin of the Russian Academy of Sciences*. 2024. vol. 94. no 11. pp. 1003–1013. DOI: 10.31857/S0869587324110043.
4. Trukhanova I.G., Gureev A.D., Lunina A.V. [The impact of digital transformation on improving the quality of medical care and reducing medical errors]. *Biomeditsinskaja radioelektronika – Biomedicine Radioengineering*. 2025. vol. 28. no 1. pp. 83–90. (In Russ.).
5. Zarubina T.V. [Topical issues of the implementation of information technologies in health care]. *Vestnik Roszdravnadzora – Roszdravnadzor Bulletin*. 2018. no. 3. pp. 20–25. (In Russ.).
6. Zhukova N., Motienko A., Levonevskiy D., Subbotin A., Anh P.T. Smart room for patient monitoring based on IoT technologies. *Proceedings of the 2023 6th Artificial Intelligence and Cloud Computing Conference*. 2023. pp. 151–158. DOI: 10.1145/3639592.3639630.

7. Denis F., Viger L., Charron A., et al. Detection of lung cancer relapse using self-reported symptoms transmitted via an internet web-application: pilot study of the sentinel follow-up. *Supportive Care in Cancer*. 2014. vol. 22. pp. 1467–1473. DOI: 10.1007/s00520-013-2111-1.
8. Denis F., Viger L., Charron A., et al. Detecting lung cancer relapse using self-evaluation forms weekly filled at home: the sentinel follow-up. *Supportive Care in Cancer*. 2014. vol. 22(1). pp. 79–85. DOI: 10.1007/s00520-013-1954-9.
9. Lebedev G.S., Fomina I.V., Artemova O.R., et al. [Overview of a family of standards for data exchange with personal medical devices]. *Vestnik Roszdravnadzora – Vestnik Roszdravnadzora*. 2022. no 6. pp. 75–84. (In Russ.).
10. Denis F., Lethrosne C., Pourel N., et al. Randomized trial comparing a web-mediated follow-up with routine surveillance in lung cancer patients. *JNCI: Journal of the National Cancer Institute*. 2017. vol. 109. no. 9. DOI: 10.1093/jnci/djx029.
11. Wattanapisit, A., Teo, C.H., Wattanapisit, S., et al. Can mobile health apps replace GPs? A scoping review of comparisons between mobile apps and GP tasks. *BMC medical informatics and decision making*. 2020. vol. 20. no. 1. DOI: 10.1186/s12911-019-1016-4.
12. Galetsi P., Katsaliaki K., Kumar S. Exploring benefits and ethical challenges in the rise of mHealth (mobile healthcare) technology for the common good: An analysis of mobile applications for health specialists. *Technovation*. 2023. vol. 121. DOI: 10.1016/j.technovation.2022.102598.
13. Khokhlov A.L., Zarubina T.V., Kotlovsky M.Yu., et al. [Mechanisms for introduction of artificial intelligence in healthcare: new ethical challenges]. *Medicinskaja jetika – Medical Ethics*. 2024. vol. 12. no. 3. pp. 4–10. (In Russ.).
14. Jian W.S., Wang J.Y., Rahmanti A.R., et al. Voice-based control system for smart hospital wards: a pilot study of patient acceptance. *BMC health services research*. 2022. vol. 22. no. 1. DOI: 10.1186/s12913-022-07668-1.
15. Brunete A., Gambao E., Hernando M., Cedazo R. Smart assistive architecture for the integration of IoT devices, robotic systems, and multimodal interfaces in healthcare environments. *Sensors*. 2021. vol. 21. no. 6. DOI: 10.3390/s21062212.
16. Azhiimah A.N., Khotimah K., Sumbawati M.S., Santosa A.B. Automatic Control Based on Voice Commands and Arduino. *Proceedings of the International Joint Conference on Science and Engineering (IJCSE 2020)*. 2020. pp. 29–34. DOI: 10.2991/aer.k.201124.006.
17. Cronin S., Doherty G. Touchless computer interfaces in hospitals: A review. *Health informatics journal*. 2018. vol. 25. no. 4. pp. 1325–1342. DOI: 10.1177/1460458217748342.
18. Ali H., Cole A., Panos G. Transforming Patient Hospital Experience Through Smart Technologies. Design, User Experience, and Usability. *Case Studies in Public and Personal Interactive Systems: 9th International Conference (DUXU 2020), Held as Part of the 22nd HCI International Conference (HCII 2020)*. 2020. pp. 203–215. DOI: 10.1007/978-3-030-49757-6_14.
19. Wen M.H., Bai D., Lin S. et al. Implementation and experience of an innovative smart patient care system: A cross-sectional study. *BMC Health Services Research*. 2022. vol. 22. no. 1. DOI: 10.1186/s12913-022-07511-7.
20. Bouldin E.D., Andresen E.M., Dunton N.E. et al. Falls among adult patients hospitalized in the United States: prevalence and trends. *Journal of patient safety*. 2013. vol. 9. no. 1. pp. 13–17. DOI: 10.1097/PTS.0b013e3182699b64.
21. Hempel S., Newberry S., Wang Z., et al. Hospital fall prevention: a systematic review of implementation, components, adherence, and effectiveness. *Journal of the American Geriatrics Society*. 2013. vol. 61. no. 4. pp. 483–494. DOI: 10.1111/jgs.12169.

22. Yadav S., Manohar M., Jeba Shiney O., Priestly Shan B., Das G.J. A Smart System for Monitoring Flow in Drip Bottles in Healthcare. Proceedings of the International Conference on Futuristic Communication and Network Technologies (VICFCNT 2020). 2022. pp. 663–669. DOI: 10.1007/978-981-16-4625-6_66.
23. Gayathri S., Ganesh C.S.S. Automatic indication system of glucose level in glucose trip bottle. International Journal of Multidisciplinary Research and Modern Education. 2017. vol. 3(1). pp. 148–151. DOI: 10.5281/zenodo.438622.
24. Feng Y.S., Liu H.Y., Hsieh M.H., Fung H.C., Chang C.Y., Yu C.C., Huang C.W. An RSSI-based device-free localization system for smart wards. IEEE International Conference on Consumer Electronics-Taiwan (ICCE-TW). 2021. pp. 1–2. DOI: 10.1109/ICCE-TW52618.2021.9603249.
25. Tsentsiper L.M., Motienko A.I., Terekhov I.S., Levonevskij D.K., Samochernyh K.A., Kondrat'ev A.N. [A digital solution for determining the severity of paroxysmal sympathetic hyperactivity syndrome in patients with brain injury]. Vestnik anesteziologii i reanimatologii – Messenger of anesthesiology and resuscitation. 2023. vol. 20. no. 6. pp. 90–96. DOI: 10.24884/2078-5658-2023-20-6-90-96.
26. Tang E.W., Wong C.K., Herbison P. Global Registry of Acute Coronary Events (GRACE) hospital discharge risk score accurately predicts long-term mortality post acute coronary syndrome. American heart journal. 2007. vol. 153. no. 1. pp. 29–35.
27. Bezzubceva E.N., Saharova A.V., Lavlinskaja L.I., et al. [Mortality risk assessment using the GRACE scale in patients who have had myocardial infarction]. Molodezhnyj innovacionnyj vestnik – Youth Innovation Bulletin. 2017. vol. 6. no. 2. pp. 302–303. (In Russ.).
28. Piradov M.A., Suponeva N.A., Ryabinkina Yu.V., et al. [Glasgow Coma Scale: Linguistic-Cultural Adaptation of the Russian Version]. Zhurnal im. N.V. Sklifosovskogo «Neotlozhnaja medicinskaja pomoshh'» – Russian Sklifosovsky Journal of Emergency Medical Care. 2021. vol. 10. no. 1. pp. 91–99. DOI: 10.23934/2223-9022-2021-10-1-91-99. (In Russ.).
29. Solovieva P.I., Sinkin M.V., Talypov A.E. [Glasgow Coma Scale and FOUR Scale to Assess Level of Consciousness]. Nervnye bolezni – Nervous diseases. 2023. no. 3. pp. 20–26. DOI: 10.24412/2226-0757-2023-13003. (In Russ.).
30. Larkin E.V., Bogomolov A.V., Privalov A.N., Dobrovolsky N.N. Discrete model of paired relay-race. Bulletin of the South Ural State University. Ser. Mathematical Modelling, Programming & Computer Software. 2018. vol. 11. no. 3. pp. 72–84.
31. Larkin E.V., Privalov A.N., Bogomolov A.V. Discrete approach to simulating synchronized relay races. Automatic Documentation and Mathematical Linguistics. 2020. vol. 54. no. 1. pp. 43–51.
32. Larkin E.V., Bogomolov A.V., Privalov A.N., Dobrovolsky N.N. Relay races along a pair of selectable routes. Bulletin of the South Ural State University. Ser. Mathematical Modelling, Programming & Computer Software. 2018. vol. 11. no. 1. C. 15–26.
33. Kibzun A.I., Sinitsyn I.N. [Modern problems of optimization theory stochastic systems]. Avtomatika i telemekhanika – Automation and telemechanics. 2020. no. 11. pp. 3–10. (In Russ.).
34. Larkin E.V., Privalov A.N., Bogatyreva Yu.I. Paired discrete competition with a free route choice. Chebyshevskii sbornik – Chebyshevskii sbornik. 2021. vol. 22. no. 2(78). pp. 145–159. (In Russ.).
35. The Glasgow Coma Scale – GCS (Teasdale G.M., Jennett B.). Available at: <https://cpd-program.ru/methods/gcsc.htm> (accessed 14.01.2025).
36. Levonevskij D.K., Motienko A.I. [Modeling and Automation of Data Collection and Processing in a Smart Medical Ward]. Programmnaya Ingeneria – Programmnaya Ingeneria. 2023. vol. 14. no. 10. pp. 502–512. DOI: 10.17587/prin.14.502-512. (In Russ.).

Levonevskiy Dmitriy — Ph.D., Associate Professor, Senior researcher, Laboratory of big data technologies in socio-cyberphysical systems, St. Petersburg Federal Research Center of the Russian Academy of Sciences (SPC RAS). Research interests: programming technologies, data processing, corporate socio-cyber-physical systems, modeling and automation of corporate processes. The number of publications — 100. levonevskij.d@iiias.spb.su; 39, 14-th Line V.O., 199178, St. Petersburg, Russia; office phone: +7(812)508-3311.

Motienko Anna — Ph.D., Senior researcher, Laboratory of big data technologies in socio-cyberphysical systems, St. Petersburg Federal Research Center of the Russian Academy of Sciences (SPC RAS). Research interests: mathematical statistics, probability theory, application of mathematical modeling in epidemiology. The number of publications — 65. anna.gunchenko@gmail.com; 39, 14-th Line V.O., 199178, St. Petersburg, Russia; office phone: +7(812)508-3311.

Bogomolov Alexey — Ph.D., Dr.Sci., Professor, Chief researcher, Laboratory of big data technologies in socio-cyberphysical systems, St. Petersburg Federal Research Center of the Russian Academy of Sciences (SPC RAS). Research interests: digitalization of medicine, medical informatics, control systems, diagnostics of human condition. The number of publications — 350. a.v.bogomolov@gmail.com; 39, 14-th Line V.O., 199178, St. Petersburg, Russia; office phone: +7(812)508-3311.

Acknowledgements. This research is supported by the state budget, project No. FFZF-2025-0019.

Руководство для авторов

Взаимодействие автора с редакцией осуществляется через личный кабинет на сайте журнала «Информатика и автоматизация» <http://ia.spcras.ru/>. При регистрации авторам рекомендуется заполнить все предложенные поля данных. Подготовка статьи ведется с помощью текстовых редакторов MS Word 2007 и выше или LaTeX. Объем основного текста (до раздела Литература) - от 20 до 30 страниц включительно. Переносы запрещены. Номера страниц не проставляются. Основная часть текста статьи разбивается на разделы, среди которых являются обязательными: введение, хотя бы один «содержательный» раздел и заключение. Допускается также мотивированное содержанием и структурой материал а выделение подразделов. В основную часть опускается помещать рисунки, таблицы, листинги и формулы. Правила их оформления подробно рассмотрены на нашем сайте в разделе «Руководство для авторов».

Author guidelines

Interaction between each potential author and the Editorial board is realized through the personal account on the website of the journal "Informatics and Automation" <http://ia.spcras.ru/>. At the registration the authors are requested to fill out all data fields in the proposed form. The submissions should be prepared using MS Word 2007, LaTeX. The text of the paper in the main part should not exceed 30 pages. Pages are not numbered; hyphenations are not allowed. Certain figures, tables, listings and formulas are allowed in the main section, and their typography is considered in more detail at the journal web.

Signed to print 04.12.2025. Passed for print 04.12.2025.

Printed in Publishing center GUAP.

Address: 67 litera A, B. Morskaya, St. Petersburg, 190000, Russia

Founder and Publisher: SPC RAS.

Address: 39 litera A, 14th Line V.O., St. Peterburg, 199178, Russia.

The journal is registered in the Federal Service for Supervision of Communications,

Information Technology, and Mass Media,

Registration Certificate (registration number) ПИ № ФС77-79228 dated September 25, 2020

Subscription Index П5513, Russian Post Catalog

Подписано к печати 04.12.2025. Дата выхода в свет 04.12.2025.

Формат 60х90 1/16. Усл. печ. л. 21,4. Заказ № 393. Тираж 300 экз., цена свободная.

Отпечатано в Редакционно-издательском центре ГУАП.

Адрес типографии: Б. Морская, д. 67, лит. А, г. Санкт-Петербург, 190000, Россия

Учредитель и издатель: СПб ФИЦ РАН.

Адрес учредителя и издателя: 14-я линия В.О., д. 39, лит. А, г. Санкт-Петербург, 199178, Россия

Журнал зарегистрирован Федеральной службой по надзору в сфере связи, информационных технологий и массовых коммуникаций, свидетельство о регистрации (регистрационный номер) ПИ № ФС77-79228 от 25 сентября 2020 г.

Подписной индекс П5513 по каталогу «Почта России»