ISSN 2713-3192
DOI 10.15622/ia.2026.25.3
http://ia.spcras.ru

ТОМ 25 № 3

ИНФОРМАТИКА И АВТОМАТИЗАЦИЯ

INFORMATICS AND AUTOMATION

СПб ФИЦ РАН

Санкт-Петербург
2026

INFORMATICS AND AUTOMATION

Volume 25 № 3, 2026

A scientific and educational journal primarily specializing in computer science, automation, robotics, applied mathematics and interdisciplinary research
Founded in 2002

Frequency
6 times a year

Founder and Publisher:
St. Petersburg Federal Research Center of the Russian Academy of Sciences (SPC RAS)

Editorial Office and Publisher Address:
39, 14th Line V.O., St.Petersburg, 199178, Russia
e-mail: ia@spcras.ru, website: http://ia.spcras.ru

Editor-in-Chief:
A. L. Ronzhin, Dr. Sci., Prof., St. Petersburg, Russia

Managing Editor: V. A. Ganicheva
Translator: Ya. N. Berezina
Art Editor: Yu. S. Cherepanova

The journal is registered with the Federal Service for Supervision of Communications, Information Technology, and Mass Media,
Registration Certificate (registration number) ПИ № ФС77-79228 dated September 25, 2020

The journal is indexed in the Scopus international database

The journal is included in the «List of Leading Peer-Reviewed Scientific Journals and Publications, in which the main scientific results of a dissertation for the degree of Doctor and Candidate of Sciences must be published»

The journal is published under the scientific and methodological guidance of the Department of Nanotechnology and Information Technologies of the Russian Academy of Sciences

© St. Petersburg Federal Research Center of the Russian Academy of Sciences, 2026

Reproduction in the press, as well as broadcasting or transmission by cable, of articles published in the print periodical «Informatics and Automation» on current economic, political, social, and religious issues is permitted, provided that the author's name and the print periodical «Informatics and Automation» are clearly indicated

Editorial Council

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

Editorial Board

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

ИНФОРМАТИКА И АВТОМАТИЗАЦИЯ

Том 25 № 3, 2026

Научный, научно-образовательный журнал с базовой специализацией
в области информатики, автоматизации, робототехники, прикладной математики
и междисциплинарных исследований.
Журнал основан в 2002 году

Периодичность
6 раз в год

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

Адрес редакции и издателя:
Россия, 199178, г. Санкт-Петербург, 4-я линия В.О., д. 39, лит. А
e-mail: ia@spcras.ru, сайт: http://ia.spcras.ru

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

Выпускающий редактор: В. А. Ганичева
Переводчик: Я. Н. Березина
Художественный редактор: Ю. С. Черепанова

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.

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», 2026

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»

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
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. Сойфер	академик РАН, д-р техн. наук, проф., Самара, РФ

Редакционная коллегия

И. С. Азаров	д-р техн. наук, проф., Минск, Беларусь
М. В. Бобырь	д-р техн. наук, проф., Курск, РФ
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
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, Соединённое Королевство
М. Сечуйски	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
М. М. Татур	д-р техн. наук, проф., Минск, Беларусь
Л. В. Уткин	д-р техн. наук, проф., Санкт-Петербург, РФ
М. Н. Фаворская	д-р техн. наук, проф., Красноярск, РФ
М. Фу	д-р наук, проф., Ханчжоу, Китай
А. Д. Хомоненко	д-р техн. наук, проф., Санкт-Петербург, РФ
Л. Б. Шереметов	д-р техн. наук, Мехико, Мексика

CONTENTS

Towards 75th Anniversary of Alexander A. Moldovian 609

Towards 50th Anniversary of Ilya S. Lebedev 614

Towards 50th Anniversary of Nikolai G. Shilov 619

Artificial Intelligence, Knowledge and Data Engineering

I. Lebedev
ADAPTIVE DATA SAMPLES GENERATION FOR SELF-ORGANIZING CONTRAST-BASED LEARNING SYSTEMS 623

A. Satsyuk, S. Radkovsky, A. Shekhovtsov, S. Sonina, A. Vorobyov
INTEGRATION OF LIGHTWEIGHT CHANNEL ATTENTION MODULES INTO DEPTHWISE SEPARABLE CONVOLUTIONAL ARCHITECTURES FOR IMPROVING OBJECT LOCALIZATION ACCURACY IN EMBEDDED SYSTEMS 654

M. Maiza, C. Cherif, S. Chouraqui, A. Taleb-Ahmed
ACCURATE REAL-TIME URBAN WASTE DETECTION USING AN ENHANCED YOLOV9-EFFICIENTVIT FRAMEWORK 682

O. Lebedev, A. Shmeleva, N. Gezha
MEDICAL ELECTROPHYSIOLOGICAL DATA AUTOMATED INTERPRETATION DEEP LEARNING MODEL 720

I. Netay, E. Pankratov, K. Kornilov, A. Golubinskiy
IDENTIFICATION OF A PERSON BY GAIT KINEMATICS IN THREE-DIMENSIONAL SPACE BASED ON INVARIANT FEATURES 751

V. Pimeshkov, M. Nikonorova, M. Shishaev
INFORMATION RETRIEVAL IN SOCIAL MEDIA BASED ON A MULTI-DOMAIN DYNAMIC KNOWLEDGE SYSTEM 787

V. Kureychik, V. Danilchenko, E. Danilchenko
ADAPTIVE PARAMETERIZATION OF A METAHEURISTIC ALGORITHM IN VLSI COMPONENT PLACEMENT 825

Information Security

I. Kotenko, I. Saenko, E. Mityakov
USING A DIGITAL TWIN FOR CYBER THREAT DETECTION IN INTELLIGENT POWER GRID CONTROL SYSTEMS 860

I. Saenko, E. Novikova, A. Meleshko, V. Gardeners
AN ONTOLOGICAL APPROACH TO RISK MANAGEMENT IN WASTEWATER TREATMENT SYSTEM 903

N. Harina, M. Dolzhenkova, Z. Iminova
SECURE BIOMETRIC AUTHENTICATION SYSTEM FOR EMBEDDED SYSTEMS WITH LIMITED COMPUTING RESOURCES 931

A. Moldovyan, N. Moldovyan, D. Moldovyan, A. Kostina
POST-QUANTUM DIGITAL SIGNATURE ALGORITHM BASED ON THE DIFFICULTY OF SOLVING POWER EQUATIONS 958

СОДЕРЖАНИЕ

К 75-летнему юбилею Молдовяна Александра Андреевича 609

К 50-летнему юбилею Лебедева Ильи Сергеевича 614

К 50-летнему юбилею Шилова Николая Германовича 619

Искусственный интеллект, инженерия данных и знаний

И.С. Лебедев
АДАПТИВНОЕ ФОРМИРОВАНИЕ ВЫБОРОК ДАННЫХ ДЛЯ САМООРГАНИЗУЮЩИХСЯ СИСТЕМ КОНТРАСТНОГО ОБУЧЕНИЯ 623

А.В. Сацюк, С.А. Радковский, А.И. Шеховцов, С.Д. Сонина, А.А. Воробьев
ИНТЕГРАЦИЯ ЛЕГКОВЕСНЫХ МОДУЛЕЙ КАНАЛЬНОГО ВНИМАНИЯ В ГЛУБИННЫЕ СЕПАРАБЕЛЬНЫЕ СВЕРТОЧНЫЕ АРХИТЕКТУРЫ ДЛЯ ПОВЫШЕНИЯ ТОЧНОСТИ ЛОКАЛИЗАЦИИ ОБЪЕКТОВ ВО ВСТРАИВАЕМЫХ СИСТЕМАХ 654

М. Майза, Ч. Шериф, С. Шураки, А. Талеб-Ахмед
ТОЧНОЕ ОБНАРУЖЕНИЕ ГОРОДСКИХ ОТХОДОВ В РЕАЛЬНОМ ВРЕМЕНИ С ИСПОЛЬЗОВАНИЕМ УЛУЧШЕННОЙ АРХИТЕКТУРЫ YOLOV9-EFFICIENTVIT 682

О.Б. Лебедев, А.Г. Шмелева, Н.С. Гежа
МОДЕЛЬ ГЛУБОКОГО ОБУЧЕНИЯ ДЛЯ АВТОМАТИЗИРОВАННОЙ ИНТЕРПРЕТАЦИИ МЕДИЦИНСКИХ ЭЛЕКТРОФИЗИОЛОГИЧЕСКИХ ДАННЫХ 720

И.В. Нетай, Е.П. Панкратов, К.И. Корнилов, А.Н. Голубинский
ИДЕНТИФИКАЦИЯ ЛИЧНОСТИ ПО КИНЕМАТИКЕ ХОДЬБЫ В ТРЁХМЕРНОМ ПРОСТРАНСТВЕ НА ОСНОВЕ ИНВАРИАНТНЫХ ПРИЗНАКОВ 751

В.К. Пимешков, М.Л. Никонорова, М.Г. Шишаев
ИНФОРМАЦИОННЫЙ ПОИСК В СОЦИАЛЬНЫХ МЕДИА НА БАЗЕ МУЛЬТИДОМЕННОЙ ДИНАМИЧЕСКОЙ СИСТЕМЫ ЗНАНИЙ 787

В.В. Курейчик, В.И. Данильченко, Е.В. Данильченко
АДАПТИВНАЯ ПАРАМЕТРИЗАЦИЯ МЕТАЭВРИСТИЧЕСКОГО АЛГОРИТМА ПРИ РАЗМЕЩЕНИИ КОМПОНЕНТОВ СБИС 825

Информационная безопасность

И.В. Котенко, И.Б. Саенко, Е.С. Митяков
ИСПОЛЬЗОВАНИЕ ЦИФРОВОГО ДВОЙНИКА ДЛЯ ОБНАРУЖЕНИЯ КИБЕРУГРОЗ В СИСТЕМАХ УПРАВЛЕНИЯ ИНТЕЛЛЕКТУАЛЬНЫМИ ЭНЕРГОСЕТЯМИ 860

И.Б. Саенко, Е.С. Новикова, А.В. Мелешко, В.Е. Садовников
ОНТОЛОГИЧЕСКИЙ ПОДХОД К УПРАВЛЕНИЮ РИСКАМИ В СИСТЕМЕ УПРАВЛЕНИЯ ОЧИСТНЫМИ СООРУЖЕНИЯМИ 903

Н.Л. Харина, М.Л. Долженкова, З.Б. Иминова
ЗАЩИЩЕННАЯ СИСТЕМА БИОМЕТРИЧЕСКОЙ АУТЕНТИФИКАЦИИ ДЛЯ ВСТРАИВАЕМЫХ СИСТЕМ С ОГРАНИЧЕННЫМИ ВЫЧИСЛИТЕЛЬНЫМИ РЕСУРСАМИ 931

А.А. Молдовян, Н.А. Молдовян, Д.Н. Молдовян, А.А. Костина
ПОСТКВАНТОВЫЙ АЛГОРИТМ ЭЦП НА ОСНОВЕ ТРУДНОСТИ РЕШЕНИЯ СТЕПЕННЫХ УРАВНЕНИЙ 958

К 75-ЛЕТНЕМУ ЮБИЛЕЮ МОЛДОВЯНА АЛЕКСАНДРА АНДРЕЕВИЧА



Доктор технических наук, профессор Молдовян Александр Андреевич родился в селе Чок Майдан, Комратского района Молдавской ССР. В 1974 г. окончил Ленинградский электротехнический институт им. В.И. Ульянова (Ленина) по специальности «Автоматизированные системы управления». В период с 1974 г. по 1994 г. разрабатывал автоматизированные системы управления, из них 14 лет в ПКБ АСУ МИНПРИБОРА СССР. С 1995 г. главный конструктор НТЦ "СПЕКТР" НПК "Красная Заря". С 1996 г. перешел на работу в Специализированный центр программных систем «СПЕКТР», созданный по рекомендации ГОСКОМОБОРОНПРОМа для выполнения НИОКР в области защиты информации. С 2005 г. являлся директором Научного филиала ФГУП «НИИ» ВЕКТОР» – СЦПС «СПЕКТР». С марта 2009 г. работал заместителем директора СПИИРАН по информационной безопасности и научным руководителем научно-исследовательского отдела проблем информационной безопасности. С декабря 2016 г. руководил научно-исследовательским отделом проблем информационной безопасности СПИИРАН. Затем переведен на должность главного научного сотрудника лаборатории проблем компьютерной безопасности СПИИРАН – СПБ ФИЦ РАН.

Молдовян А.А. руководил и участвовал в разработке более 40 НИОКР в области защиты информации, а также создании ряда практически значимых систем защиты информации от несанкционированного доступа, получивших широкое внедрение

на межотраслевом уровне: «Кобра», «Спектр-Z» «Спектр-М», «СГУ-2», «Спектр-2000», «Ключ-П», «Ключ-Ш», «Вектор-01», «ЩИТ-РЖД», «Аура» сертифицированных Гостехкомиссией России, ФСТЭК России, МО РФ и ФСБ России.

Молдовян А.А. активный участник конференций международного и российского уровня. Сопредседатель секции «Информационная безопасность» V и XI Санкт-Петербургских международных конференций «Региональная информатика». Член оргкомитета регулярной Межрегиональной конференции «Информационная безопасность регионов России».

Молдовян А.А. известен как авторитетный ученый, участвовал в работе диссертационных советов при НИИ «Рубин», МГТУ им. Н.Э. Баумана, НИУ ИТМО, СПИИРАН, с 2021 г. член диссертационного совета при СПб ФИЦ РАН. Является экспертом Минобрнауки в научно-технической сфере, представитель СПб ФИЦ РАН в ТК №26 «Криптографическая защита информации». Разработал и читал авторские курсы лекций по проблемам информационной безопасности в НИУ ИТМО, ГУВК и СПГЭТУ-ЛЭТИ. Автор 6 монографий и 15 учебных пособий в области защиты информации.

Существенным вкладом Молдовяна А.А. в развитие науки является его деятельность по подготовке высококвалифицированных кадров. Им подготовлено 14 кандидатов и 2 доктора технических наук. Педагогический стаж составляет 27 лет.

Профессор Молдовян А.А. является руководителем ведущей научной школы Санкт-Петербурга «Криптография: методы, алгоритмы и протоколы для защиты информации в компьютерных системах», с 2022 г. председатель ГЭК НИУ ИТМО по направлению «Информационная безопасность». По результатам исследований в области защиты информации опубликовал более 250 научных работ, в том числе в ведущих журналах мира: «Автоматика и телемеханика» (Россия), «Кибернетика и системный анализ» (Украина), «Journal of Cryptology» (США), «Lecture Notes in Computer Science» (Германия). На созданные изобретения получил более 60 патентов России, США, Германии, Франции, Великобритании, Китая, Кореи и др. стран. Изобретения Молдовяна А.А. в области информационных технологий удостоены Гран-при, 10 золотых и 3 серебряных медалей, а также ряда дипломов на международных выставках в России, Японии, Бельгии, Болгарии, Венгрии и Польше. Правительством Санкт-Петербурга в 1997 г. награжден Дипломом за лучшую разработку военных ученых, внедренных в городское хозяйство города.

Указом Президента РФ награжден медалью «В память 300-летия Санкт-Петербурга» (2003 г.). Удостоен ведомственных наград за вклад в развитие методов, средств и систем защиты информации: значок «Почетный радист» (2002 г.), Почетная грамота МИНПРОМЭНЭНЕРГО (2004 г.), Благодарность ОАО «РЖД» (2004 г.), Памятный знак «90 лет ГРУ ГШ ВС РФ» (2008 г.). Памятная медаль Научного совета по информатизации Санкт-Петербурга «За вклад в развитие информационного общества» (2012 г.). Приказом Гостехкомиссии России награжден медалью «За укрепление государственной системы защиты информации» I степени (2004 г.). Приказом ФСТЭК России награжден знаком «За заслуги в защите информации» (2010 г.). Решением Российской академии наук награжден Почетной грамотой (2021 г.).

Сотрудники СПб ФИЦ РАН, коллеги из многих организаций, его ученики и последователи, а также редакционная коллегия журнала «Информатика и автоматизация» (Труды СПИИРАН) поздравляют Молдовяна Александра Андреевича с юбилеем и желают ему крепкого здоровья и дальнейших творческих успехов!

Список избранных публикаций

1. Молдовян А.А., Молдовян Н.А., Гуц Н.Д., Изотов Б.В. Криптография: скоростные шифры. СПб: БХВ-Петербург, 2002. 496 с.
2. Молдовян А.А., Молдовян Н.А., Еремеев М.А. Криптография: от примитивов к синтезу алгоритмов. СПб: БХВ-Петербург, 2004. 458 с.
3. Moldovyan A., Moldovyan N. Data-driven Block Ciphers for Fast Telecommunication Systems. New York: Auerbach Publications, 2008. 185 p.
4. Молдовян А.А., Молдовян Н.А. Новый принцип построения криптографических модулей в системах защиты ЭВМ. Кибернетика и системный анализ. 1993. №5. С. 42–49.
5. Молдовян А.А., Молдовян Н.А. Программно-ориентированная криптосистема с неопределенным алгоритмом шифрования. Управляющие системы и машины. 1995. №6. С. 38–43.
6. Молдовян А.А. Подход к созданию средств защиты информации массового применения. Управление защитой информации. 1998. Т. 2. №1. С. 26–27.
7. Moldovyan A., Moldovyan N. Software Encryption Algorithms for Transparent Protection Technology. Cryptologia. 1998. vol. 22. no. 1. pp. 56–68.
8. Moldovyan A. Fast Block Cipher Based on controlled permutations. Computer Science Journal of Moldova. 2000. vol. 8. no. 3. pp. 270–283.
9. Молдовян А.А. Построение скоростного программмного алгоритма поточного типа на базе псевдовероятностной выборки подключей. Вопросы защиты информации. 2000. №4. С. 53–56.
10. Moldovyan A., Moldovyan N. A Cipher Based on Data-Dependent Permutations. Journal of Cryptology. 2002. vol. 15. pp. 61–72.
11. Молдовян А.А. Новый способ синтеза переключаемых управляемых операционных блоков. Вопросы защиты информации. 2003. №3. С. 38–45.
12. Молдовян А.А., Молдовян Н.А., Еремеев М.И. Защитные преобразования информации в АСУ на основе нового примитива. Автоматика и телемеханика. 2002. №12. С. 147–165.
13. Изотов Б.В., Молдовян А.А. Перспективы создания шифраторов на основе управляемых криптографических примитивов. Известия высших учебных заведений. Приборостроение. 2003. №7. С. 73–79.
14. Moldovyan A., Moldovyan N., Sklavos N., Koufopavlou O. CHESS-64, a Block Cipher Based on Data-Dependent Operations: Design Variants and Hardware Implementation Efficiency. Asian Journal of Information Technology. 2005. vol. 4. no. 4. pp. 320–328.
15. Moldovyan A., Moldovyan N. Blind Collective Signature Protocol Based on Discrete Logarithm Problem. International Journal of Network Security. 2010. vol. 11. no. 2. pp. 106–113.
16. Молдовян А.А. Постквантовая схема ЭЦП, основанная на вычислительной сложности восстановления параметров векторного конечного поля. Вопросы защиты информации. 2023. №4. С. 20–26.
17. Молдовян А.А., Молдовян Н.А., Молдовян Д.Н. Постквантовые двухключевые криптосхемы на конечных алгебрах. Информатика и автоматизация. 2024. Т. 23. №4. С. 1246–1276.

18. Duong T.M., Moldovyan A., Moldovyan N., Nguyen H.M., Do T.B. Structure of 6-dimensional finite non-commutative algebras with many single-sided units. Bulletin of Electrical Engineering and Informatics. 2025. vol. 14. pp. 2017–2030.
19. Moldovyan A., Moldovyan N. A method for enhancing randomization in algebraic signature algorithms on non-commutative algebras. Quasigroups and related systems. 2025. vol. 33. pp. 71–83.
20. Молдовян А.А., Молдовян Д.Н., Костина А.А. Рандомизация в постквантовых алгоритмах ЭЦП с секретной группой. Информатика и автоматизация. 2025. Т. 24. №6. С. 1810–1835.
21. Молдовян А.А. Постквантовый алгебраический алгоритм ЭЦП с тремя скрытыми группами. Вопросы кибербезопасности. 2025. №5(69). С. 78–87.

К 50-ЛЕТНЕМУ ЮБИЛЕЮ ЛЕБЕДЕВА ИЛЬИ СЕРГЕЕВИЧА



Доктор технических наук, профессор Лебедев Илья Сергеевич, является выдающимся ученым в области самоорганизующихся систем искусственного интеллекта, интеллектуальных систем анализа и обработки слабоструктурированной информации. В 1998 г. с отличием окончил Санкт-Петербургское высшее училище радиоэлектроники противовоздушной обороны (СПВУРЭ ПВО), факультет математического и программного обеспечения автоматизированных систем специального назначения. В 2002 г. окончил очную адъюнктуру в филиале Военного университета противовоздушной обороны по специальности 20.02.12 «Системный анализ, моделирование боевых действий и систем военного назначения, компьютерные технологии в военном деле» и успешно защитил диссертацию на соискание ученой степени кандидата технических наук. За время прохождения службы в Вооруженных силах имеел награды МО РФ. Военную службу закончил в воинском звании подполковника.

В 2008 г. присвоено ученое звание доцента по кафедре информационных систем в экономике в Санкт-Петербургском государственном университете (СПбГУ). В 2012 г. успешно защитил докторскую диссертацию по специальности 05.13.19 «Методы и системы защиты информации, информационная безопасность» в Национальном исследовательском университете информационных технологий механики и оптике (Университет ИТМО), звание профессора по специальности 05.13.19 было присвоено в 2018 г.

В настоящее время является главным научным сотрудником Санкт-Петербургского Федерального исследовательского центра Российской академии наук (СПб ФИЦ РАН).

Профессор Лебедев И.С. обладает обширным опытом в области исследования, разработки и проектирования автоматизированных систем, систем искусственного интеллекта, управления знаниями, поддержки принятия решений и управления. За время своей деятельности занимался разработкой действующих и перспективных автоматизированных систем двойного назначения. Принимал участие в непосредственном создании элементов Единой системы организации воздушного движения РФ. Является одним из разработчиков ряда используемых в действующих АСУ устройств сопряжения, внедренных в серийное производство.

Лебедев И.С. работал над широким кругом проблем, таких как анализ и обеспечение состояний информационной безопасности систем и каналов связи на основе моделей и методов инженерии знаний и искусственного интеллекта. Полученные под его руководством результаты неоднократно использовались при проведении НИОКР и были включены в список важнейших результатов РАН.

Научные исследования профессора Лебедева И.С. нашли отражение в НИР и ОКР, поддержаны программами «Научно-технологического развития Российской Федерации», грантами РФФИ и РНФ. Основные результаты выполненных исследований являются значимыми в контексте приоритетного направления Стратегии научно-технологического развития РФ (утвержденной Указом Президента Российской Федерации от 28 февраля 2024 г. № 145) «Н1. Переход к передовым технологиям проектирования и создания высокотехнологичной продукции, основанным на применении интеллектуальных производственных решений, роботизированных и высокопроизводительных вычислительных систем, новых материалов и химических соединений, результатов обработки больших объемов данных, технологий машинного обучения и искусственного интеллекта».

Результаты исследований Лебедева И.С. неоднократно публиковались в ведущих отечественных и международных изданиях, а также обсуждались на различных научно-технических форумах.

Существенным вкладом Лебедева И.С. в развитие науки является его деятельность по подготовке высококвалифицированных кадров. Им подготовлено 8 кандидатов и 1 доктор технических наук. В настоящее время принимает активное участие в работе

диссертационных советов СПб ФИЦ РАН, СПб ПУ Петра Великого и ГУМРФ им. С.О. Макарова.

Профессор Лебедев И.С. ведет активную педагогическую деятельность в Санкт-Петербургском государственном университете (СПбГУ). Является автором более 200 научных публикаций, среди которых ряд работ выполнен в высокорейтинговых журналах мирового уровня. Является членом редколлегий ряда журналов и международных конференций. С 2016 г. по 2018 г. являлся научным руководителем образовательной программы Информационная безопасность (Факультет безопасных информационных технологии, Университет ИТМО).

Сотрудники СПб ФИЦ РАН, коллеги из многих организаций, его ученики и последователи, а также редакционная коллегия журнала «Информатика и автоматизация» (Труды СПИИРАН) поздравляют Лебедева Илью Сергеевича с юбилеем и желают ему крепкого здоровья и дальнейших творческих успехов!

Список избранных публикаций

1. Lebedev I., Sukhoparov M., Semenov V., Khasanov D. Adaptive Segmentation of Information Sequences for Machine Learning Modular Regression Models. Emerging Science Journal. 2025. vol. 9. no. 5. pp. 2420–2438.
2. Лебедев И.С. Адаптивное построение регрессионных моделей на основе анализа функционала качества обработки сегментов последовательности. Информатика и автоматизация. 2025. Т. 24. №2. С. 363–394.
3. Lebedev I., Sukhoparov M. Improving the Quality Indicators of Multilevel Data Sampling Processing Models Based on Unsupervised Clustering. Emerging Science Journal. 2024. vol. 8. no. 1. pp. 355–371.
4. Лебедев И.С. Обработка информационных последовательностей с использованием адаптивного анализа сегментов при оценке состояния систем. Информационно-управляющие системы. 2025. №3. С. 25–36.
5. Lebedev I., Sukhoparov M. Adaptive learning and integrated use of information flow forecasting methods. Emerging Science Journal. 2023. vol. 7. no. 3. pp. 704–723.
6. Лебедев И.С. Применение многоуровневых моделей в задачах классификации и регрессионного анализа. Информатика и автоматизация. 2023. Т. 22. №3. С. 487–510.
7. Лебедев И.С. Адаптивное применение моделей машинного обучения на отдельных сегментах выборки в задачах регрессии и классификации. Информационно-управляющие системы. 2022. №3. С. 20–30.
8. Лебедев И.С. Применение неконтролируемой кластеризации выборок для повышения качественных показателей многоуровневых моделей обработки данных. Вестник Томского государственного университета. Управление, вычислительная техника и информатика. 2024. №66. С. 44–54.
9. Лебедев И.С., Сикарев И.А., Сухопаров М.Е. Повышение качественных показателей многоуровневых моделей обработки выборок данных на основе неконтролируемой кластеризации. T-Comm: Телекоммуникации и транспорт. 2024. Т. 18. №4. С. 30–35.
10. Лебедев И.С. Сегментирование множества данных с учетом информации воздействующих факторов. Информационно-управляющие системы. 2021. №3. С. 29–38.
11. Сухопаров М.Е., Семенов В.В., Салахутдинова К.И., Лебедев И.С. Выявление аномального функционирования устройств индустрии 4.0 на основе поведенческих паттернов. Проблемы информационной безопасности. Компьютерные системы. 2020. №1. С. 96–102.
12. Лебедев И.С. Обучение на несбалансированных выборках ансамбля классификаторов при анализе состояния сетевых сегментов. Прикладная информатика. 2021. Т. 16. №3. С. 109–119.
13. Рзаев Б.Т., Лебедев И.С. Применение бэггинга при поиске аномалий сетевого трафика. Научно-технический вестник информационных технологий, механики и оптики. 2021. Т. 21. №2. С. 234–240.

14. Тихонов Д.Д., Лебедев И.С. Метод формирования сегментов информационной последовательности с использованием функционала качества моделей обработки. Научно-технический вестник информационных технологий, механики и оптики. 2024. Т. 24. №3. С. 474–482.
15. Sukhoparov M., Lebedev I., Tikhonov D. Segmenting input data to improve the quality of identification of information security events. Automatic Control and Computer Sciences. 2024. vol. 58. pp.1192–1203.

К 50-ЛЕТНЕМУ ЮБИЛЕЮ ШИЛОВА НИКОЛАЯ ГЕРМАНОВИЧА



Доктор технических наук, доцент Шилов Николай Германович, является выдающимся ученым в области производственных систем, управления знаниями, генеративного дизайна и интеллектуальных систем анализа психофизиологического состояния человека. В 1998 г. с отличием окончил Санкт-Петербургский государственный технический университет (н.в. СПбПУ Петра Великого), факультет экономики и менеджмента по специальности «Экономика и управление на машиностроительном предприятии». В 2001 г. окончил очную аспирантуру в Санкт-Петербургском институте информатики и автоматизации Российской академии наук по специальности 05.13.01 «Системный анализ, управление и обработка информации (технические системы)». В 2005 г. успешно защитил диссертацию на соискание ученой степени кандидата технических наук по теме «Разработка моделей для интеллектуальной поддержки принятия решений при конфигурировании виртуальных предприятий», специальность 05.13.01 «Системный анализ, управление и обработка информации (технические системы)» по техническим наукам. В 2025 г. успешно защитил докторскую диссертацию по теме «Контекстно-ориентированная поддержка принятия решений при конфигурировании производственных систем», специальность 05.13.01 «Системный анализ, управление и обработка информации (технические системы)». В настоящее время является ведущим научным сотрудником Санкт-Петербургского Федерального исследовательского центра Российской академии наук (СПб ФИЦ РАН).

Профессор Шилов Н.Г. обладает обширным опытом в области исследования, разработки и проектирования комплексных компьютерных систем, искусственного интеллекта, управления знаниями, поддержки принятия решений и управления производственными сетями. Николай Германович работал над широким кругом проблем, таких как корпоративное управление знаниями, онтологически-ориентированная интеграция информации, управление контекстом, оперативная поддержка принятия решений, конфигурирование виртуальных предприятий, динамическая логистика и мониторинг психофизиологического состояния человека. Полученные под его руководством результаты были неоднократно включены в список важнейших результатов РАН и позволили приступить к созданию персонализированных систем интеллектуальной поддержки принятия решений, имеющих большое практическое значение для решения прикладных проблем как в области научных исследований и образования, так и в бизнесе.

Научные исследования Шилова Н.Г. поддержаны многочисленными грантами РФФИ и РНФ. Основные результаты, полученные в ходе выполнения исследований, являются значимыми в контексте приоритетного направления Стратегии научно-технологического развития РФ (утвержденной Указом Президента Российской Федерации от 28 февраля 2024 г. № 145) «Н1. Переход к передовым технологиям проектирования и создания высокотехнологичной продукции, основанным на применении интеллектуальных производственных решений, роботизированных и высокопроизводительных вычислительных систем, новых материалов и химических соединений, результатов обработки больших объемов данных, технологий машинного обучения и искусственного интеллекта».

Шилов Н.Г. ведет активную педагогическую деятельность, работает в Санкт-Петербургском государственном электротехническом университете «ЛЭТИ» им. В.И. Ульянова (Ленина). Является автором более 300 научных публикаций, среди которых ряд работ выполнен в высокорейтинговых журналах мирового уровня. Является членом редколлегий ряда журналов и международных конференций.

Сотрудники СПб ФИЦ РАН, коллеги из многих организаций, его ученики и последователи, а также редакционная коллегия журнала «Информатика и автоматизация» (Труды СПИИРАН) поздравляют Шилова Николая Германовича с юбилеем и желают ему крепкого здоровья и дальнейших творческих успехов!

Список избранных публикаций

1. Smirnov A., Shilov N., Oroszi A., Sinko M., Krebs T. Towards Life Cycle Management for Product and System Configurations: Required Improvements in Business Processes and Information Systems. Proceedings of 23rd CIRP Conference on Life Cycle Engineering. 2016. vol. 48. pp. 84–89.
2. Шилов Н.Г. Методология построения проактивных рекомендующих систем для инфо-мобильных приложений. Информационно-управляющие системы. 2016. №6. С. 16–24.
3. Smirnov A., Shilov N., Oroszi A., Sinko M., Krebs T. Changing Information Management in Product-Service System PLM: Customer-Oriented Strategy. Product Lifecycle Management and the Industry of the Future. IFIP 14th International Conference on Product Lifecycle Management (PLM17). IFIP Advances in Information and Communication Technology, Springer. 2017. vol. 517. pp. 701–709.
4. Комзалов А.М., Шилов Н.Г. Применение современных технологий в системах помощи водителю автомобиля. Известия высших учебных заведений. Приборостроение. 2017. Т. 60. №11. С. 1077–1082.
5. Шилов Н.Г., Щекотов М.С. Онтологическое моделирование управляющих автоматов сервисов киберфизических систем. Информационно-управляющие системы. 2017. №5. С. 80–88.
6. Smirnov A., Ponomarev A., Shilov N., Kashevnik A., Teslya N. Ontology-Based Human-Computer Cloud for Decision Support: Architecture and Applications in Tourism. International Journal of Embedded and Real-Time Communication Systems (IJERTCS). 2018. vol. 9. no. 1. pp. 1–19.
7. Smirnov A., Shilov N. Multi-aspect Ontology for Semantic Interoperability in PLM: Analysis of Possible Notations. Proceedings of Lifecycle Management to Support Industry 4.0 (PLM 2018). IFIP Advances in Information and Communication Technology, Springer. 2018. vol. 540. pp. 314–323.
8. Зандкуль К., Смирнов А.В., Шилов Н.Г. Таргетированное предоставление информации на цифровых вывесках: возможные решения. Труды СПИИРАН. 2019. Т. 18. №4. С. 831–857.
9. Шилов Н.Г. Персонализация предоставления информации на электронных табло: основные требования и методология. Информационно-управляющие системы. 2019. №4. С. 19–28.
10. Kashevnik A., Othman W., Ryabchikov I., Shilov N. Estimation of Motion and Respiratory Characteristics during the Meditation Practice Based on Video Analysis. Sensors. 2021. vol. 21. no. 11. 3771 p.
11. Smirnov A., Levashova T., Ponomarev A., Shilov N. Methodology for Multi-Aspect Ontology Development: Ontology for Decision Support Based on Human-Machine Collective Intelligence. IEEE. 2021. vol. 9. pp. 135167–135185.
12. Смирнов А.В., Пономарев А.В., Шилов Н.Г., Левашова Т.В. Нейро-символический искусственный интеллект в коллаборативных системах поддержки принятия решений. Искусственный интеллект и принятие решений. 2022. №3. С. 36–50.

13. Шилов Н.Г., Пономарев А.В., Смирнов А.В. Анализ методов онтолого-ориентированного нейро-символического интеллекта при коллаборативной поддержке принятия решений. Информатика и автоматизация. 2023. Т. 22. №3. С. 576–615.
14. Hamoud B., Kashevnik A., Othman W., Shilov N. Neural Network Model Combination for Video-Based Blood Pressure Estimation: New Approach and Evaluation. Sensors. 2023. vol. 23. no. 4. 1753 p.
15. Kashevnik A., Shilov N., Teslya N., Hasan F., Kitenko A., Dukareva V., Abdurakhimov M., Blokhin D. An Approach to Engineering Drawing Organization: Title Block Detection and Processing. IEEE. 2023. pp. 1–15.
16. Смирнов А.В., Пономарев А.В., Шилов Н.Г., Левашова Т.В., Тесля Н.Н. Концепция построения коллаборативных систем поддержки принятия решений: подход и архитектура платформы. Информатика и автоматизация. 2024. Т. 23. №4. С. 1139–1172.
17. Шилов Н.Г. Разработка мультиаспектной онтологии для поддержки принятия решений в производственных системах. Информационные технологии и вычислительные системы. 2024. №2. С. 52–64.
18. Шилов Н.Г. Анализ вариантов интеграции символьных и нейросетевых знаний на примере задачи классификации моделей предприятий. Вестник Воронежского государственного университета. Серия: системный анализ и информационные технологии. 2024. №2. С. 140–151.
19. Шилов Н.Г. Мультиаспектное представление проблемной области для конфигурирования социотехнических систем. Известия высших учебных заведений. Приборостроение. 2025. Т. 68. №10. С. 852–861.
20. Sandkuhl K., Shilov N., Smirnov A., Seigerroth U. Multi-aspect Ontologies as Support for Quantified Products Lifecycle Management. IFIP Advances in Information and Communication Technology, Springer. 2025. vol. 740. pp. 15–29.

УДК 004.056 **DOI** 10.15622/ia.25.3.1

И.С. Лебедев

АДАПТИВНОЕ ФОРМИРОВАНИЕ ВЫБОРОК ДАННЫХ ДЛЯ САМООРГАНИЗУЮЩИХСЯ СИСТЕМ КОНТРАСТНОГО ОБУЧЕНИЯ

Лебедев И.С. **Адаптивное формирование выборок данных для самоорганизующихся систем контрастного обучения.**

Аннотация. Применение самообучения и самоорганизующихся моделей для задач классификации и прогнозирования при обработке временных рядов и информационных последовательностей сталкивается с рядом проблемных вопросов организации данных. Повышение показателей качества обработки таких систем вызывает необходимость совершенствования методов выбора объектов наблюдения для обучающих выборок данных. В статье предложен метод формирования и анализа выборок данных на основе определения патчей информационной последовательности, обладающих разными характеристиками длины и сдвига, отличающаяся использованием функционала качества моделей обработки. На основе эксперимента на модельных данных и выборках проведена оценка предлагаемого метода. Получены значения показателя качества accuracy для разных алгоритмов обработки при различных длинах и сдвигах последовательностей патча. Определены свойства полученных патчей с использованием метрик коэффициента силуэта и расстояния между центройдами. Проанализированы ошибки классифицирующих алгоритмов. Выделены доверительные интервалы ошибок. Определено, что на изменение длины и сдвига патча влияет на достигаемые значения accuracy классифицирующих алгоритмов. Предлагаемый метод дает возможность повысить показатель accuracy за счет выбора длины и сдвига при формировании патча и назначения моделей, которые имеют наилучшие показатели. Результаты показывают возможность увеличения на 6-10% для слабых моделей, в то время как для сильных моделей наблюдается улучшение на 1-5% в сценариях с ограниченными данными. Анализ предложенного решения показывает, что варьирование параметрами сдвига и длины последовательности при формирования обучающих выборок данных оказывает влияние на эффективность обработки данных.

Ключевые слова: машинное обучение, адаптивные модели, повышение качества обработки, формирование патчей.

1. Введение. Временные ряды и информационные последовательности играют важную роль в различных областях. Анализ финансовых рынков, состояния информационных систем, климатических изменений, генерации энергии, различных видов деятельности человека обуславливает необходимость разработки эффективных моделей и методов обработки, предполагающих решение задач классификации и прогнозирования.

В настоящее время одно из интенсивно развивающихся направлений обработки временных рядов и информационных последовательностей является применение различных методов самообучения для задач классификации [1 – 3] и прогнозирования [4 – 6].

Объемы накопленной информации настолько велики, что их разметка с помощью экспертов затруднительна.

Подходы к обработке на основе сверточных [7, 8], рекуррентных сетей [8, 9], трансформерных технологий [10 – 12] доказали свою эффективность при решении ряда прикладных задач. Ряд исследований показывает, что такие модели могут захватывать пространственную и временную информацию последовательностей. Они часто превосходят традиционные методы Dynamic Time Warping, Bag of Stochastic Frontier Analysis Symbols и Collective of Transformation-Based Ensembles [13]. Однако формирование обучающих выборок по-прежнему является одним из проблемных вопросов во многих практических приложениях, от которого зависит повышение качественных показателей. Это связано с рядом архитектурных ограничений моделей обработки.

Традиционные методы и модели глубокого обучения направлены на анализ наиболее значимой информации для классификации или прогнозирования. В большинстве случаев в задачах извлечения знаний последовательности данных ограничиваются подряд идущими последовательностями. Например, CNN при обработке временных рядов обычно оптимизируются для захвата шаблонов только в пределах ограниченного размера окна [14, 15]. Если последовательности, содержащие значимые информационные значения, имеют длительные временные интервалы, при обработке может наблюдаться понижение качественных показателей. Расширенные сверточные нейронные сети (Extended Convolutional Neural Networks, E-CNN) частично решают эту проблему за счет скользящих фильтров. Однако они по-прежнему ограничены своей внутренней архитектурой и зависят от количества слоев, которое может быть недостаточно большим для полного захвата долгосрочных зависимостей и приводить к исчезновению градиентов.

Аналогично, эффективность трансформеров [16] в значительной степени определяется множеством факторов, таких как длина последовательности, свойства обрабатываемых данных, особенностей разметки и определения токенов. Помимо этого, при создании на их основе моделей и методов обработки приходится решать ряд вопросов архитектурных ограничений, сложности оптимизации и настройки, снижения высоких вычислительных затрат при ограниченности ресурсов.

Подобными недостатками обладает большинство современных, недавно появившихся моделей, ориентированных на обработку временных рядов и информационных последовательностей. Например, Tiny Time Mixers (TTM) чувствительна к длине последовательностей [17]. В ряде задач это обуславливает необходимость обучения совокупности моделей для разных настроек длины. Предобученная нейросеть Moment при дообучении

небольшими наборами будет работать только для примеров в обучающих данных, но может не показать положительного результата в обобщении [18].

Модель Moirai для прогнозирования временных рядов имеет недостатки, связанные с ограничениями внутренней архитектуры, не позволяющими выявить частоту появления паттернов для «малых данных», что влияет на качество результата [19].

Производительность и внедрение существующих моделей и методов глубокого обучения, применяемых в анализе временных рядов и информационных последовательностей, ограничены доступностью хорошо аннотированных меток. Несмотря на то, что ежедневно появляется значительное количество новых данных, собираемых устройствами, датчиками, сенсорами, процесс маркировки последовательностей значительно отстаёт от их генерации. Ручная маркировка требует наличия экспертов со знанием предметной области и опытом, которые могут обработать только небольшую часть набора. Более того, в некоторых случаях, когда задействовано несколько экспертов, возникающие события в последовательностях объектов наблюдения часто сложно аннотировать из-за разногласий между ними. В совокупности эти проблемы приводят к дефициту размеченных объектов наблюдения, что является существенным препятствием для применения глубокого обучения в этой области.

Таким образом, возникает задача автоматического формирования обучающих выборок, где структура данных и свойства учитывает внутреннюю архитектуру, длину входящих последовательностей, и позволяет увеличить количество эталонных и уменьшить число шумовых примеров.

2. Модели и методы обработки последовательностей. Временные ряды имеют существенное значение для множества сфер деятельности и активно привлекают внимание научного сообщества. Это обуславливает непрерывное развитие и оптимизацию разнообразных моделей и методов глубокого обучения, применяемых в задачах классификации и прогнозирования.

В современной практике анализа и обработки данных широко используются свёрточные нейронные сети, трансформеры, а также предобученные модели. Однако одна из проблем эффективного применения классических моделей и методов машинного обучения состоит в том, что объем данных растет быстрее, чем возможности по разметке. В связи с этим, автоматическое формирование и повышение качества обучающих выборок данных является

направлением, позволяющим улучшить качество обработки информационных последовательностей.

Сегодня процесс создания размеченных данных и структур тесно связан с архитектурой используемых моделей обработки. В частности, эффективность свёрточных нейронных сетей напрямую определяется их способностью выявлять значимые локальные паттерны в данных [20]. Формирование обучающей выборки из содержащих их примеров позволяет достигать значительного качества обработки изображений, аудио и видео последовательностей. Тем не менее одним из проблемных вопросов является определение и выбор паттернов, позволяющих повысить качественные показатели обработки. Для его решения разрабатываются различные способы, чаще всего основанные на свертке. Например, в методах DSN [21] применяются динамические разреженные соединения между нейронами, чтобы определить значащие паттерны на различных уровнях для временных рядов. Другой подход, основанный на свертке, SciNet [22] фиксирует характеристики, разбивая каждую последовательность на подпоследовательности, позволяющие анализировать временную динамику данных.

Второе направление связано с использованием трансформеров. Изначально эти модели были разработаны для обработки естественного языка (NLP), но нашли применение и в обработке временных рядов и информационных последовательностей. Достижение качественного результата таких моделей зависит от выбора поступающих на вход подпоследовательностей, влияющих на эффективность применения «механизма внимания», определяющего наиболее значимые элементы.

Для этого используются разные решения, внутренние механизмы, направленные на определение свойств обрабатываемых последовательностей. Informer [23] улучшает возможности простого трансформера при работе с длинными входными последовательностями за счет анализа саморегуляции с использованием механизма самовнимания ProbSparse. В Autoformer [24] внедрен декомпозиционный подход и механизм автокорреляции, что позволяет более эффективно работать с длинными последовательностями данных и достигать более высокой точности прогнозирования. ContiFormer [25] использует непрерывное временное представление и модифицированный механизм внимания, что позволяет более эффективно работать с данными, собранными в произвольные моменты времени. PatchTST [26] улучшает прогнозирование временных рядов за счет применения подхода к сегментации данных и механизма самообучения, что позволяет более эффективно обрабатывать

и анализировать временные последовательности, сохраняя при этом глобальный контекст.

Описанные выше модели требуют больших объёмов данных для эффективного обучения (от сотен тысяч до миллиардов токенов). Это связано с большим числом параметров, необходимостью улавливать дальние зависимости в последовательностях, потребностью в обобщении на разнообразных примерах. Они имеют ограничения к выборкам. Для их обучения важна чистота и релевантность данных, что обуславливает необходимость фильтрации шума и некорректных примеров, удаление дубликатов, балансировка классов (для задач классификации). Точность трансформеров варьируется в зависимости от задачи, архитектуры модели, объёма и качества данных, а также других факторов. В целом считается, что трансформеры демонстрируют высокие показатели в различных областях от 91% и выше.

Другое направление определяется применением предварительно обученных моделей для захвата сложных временных паттернов. Например, решения на основе модели Moment [18], обученной на «глобальном датасете» временных рядов Time-series Pile, требуется только предварительная настройка на анализируемую выборку. Однако её качество при анализе временного ряда зависит от определения оптимального размера окна и учета временного сдвига последовательности.

Развиваются контрастные методы анализа временных рядов, использующие сопоставление данных для выявления закономерностей. В [27] представлен подход SoftCLT, который использует расстояния между выборками временных рядов. Он направлен на обнаружение корреляции между соседними последовательностями.

Для устранения дефицита данных применяется контрастное обучение [28]. Оно содержит шаги по расширению выборок временных рядов для создания объектов наблюдений разных классов, анализу расширенных выборок и оптимизацию модели обработки на основе полученных значений функций потерь. Самостоятельное контрастное обучение не требует меток образцов на этапе обучения модели [29 – 31]. Развитие контрастного обучения для временных рядов требует решения ряда проблемных вопросов формирования структур, позволяющих осуществлять эффективную обработку данных.

Несмотря на появление сильных моделей, использование в их составе различных механизмов анализа временных рядов, формирование и выбор последовательностей и паттернов, поступающих

на вход в процессе обучения, настройки и обработки по-прежнему остается сложной задачей, от которой зависят качественные показатели.

3. Предлагаемый метод.

3.1. Постановка задачи. В настоящее время применение методов интеллектуального анализа информационных последовательностей имеет ограничение, связанное с качеством разметки данных. Различные системы в ходе функционирования требуют оценки протекающих процессов. В связи с этим, возникает необходимость решения задачи классификации состояния системы, отнесения её текущего состояния к одному из заранее определённых классов на основе её наблюдаемых параметров и характеристик. Для её решения предлагается метод выделения репрезентативных фрагментов (патчей, паттернов, сегментов) из исходных информационных последовательностей и временных рядов, обеспечивающих эффективную адаптацию и самообучение интеллектуальных систем мониторинга состояния при соблюдении заданных показателей качества.

В предлагаемом методе в основе автоматической разметки лежит разделение последовательностей на патчи. Под «патчем» понимается подпоследовательность, которая используется в анализе временных рядов для обработки данных. Размер патча влияет на точность прогнозирования и классификации в моделях на основе машинного обучения. Несмотря на присутствие в трансформерах и моделях на основе нейронных сетей различных встроенных механизмов патчинга, разбиение временного ряда и формирование сегментов до их входа может оказывать положительное влияние на процессы обучения и качественные показатели обработки.

Формирование сегментов способно обеспечивать воздействие на вычисление свойств паттернов и захват «значащей информации» моделями обработки, сокращать длину последовательности, уменьшать вычислительную сложность. В случае наличия периодичности в последовательностях данных патчи дают возможность модели изучать более длинные временные периоды, что потенциально повышает качество прогнозов. Однако выбор размера патча зависит от задачи и особенностей данных.

Поступающие на вход моделей обработки выборки могут представлять собой данные с разнообразными структурами, имеющими неявные закономерности в распределениях, дисбаланс классов и частот появления событий, изменения диапазонов значений переменных под воздействием неопределенных факторов.

Формализованная постановка задачи может быть представлена следующим образом.

Пусть X – обучающая последовательность данных, состоящая из объектов наблюдения $x \in X$. Подпоследовательности $\{X^{c_1}, \ldots, X^{c_k}\} \in X$ получены в разных состояниях $\{c_1, \ldots, c_k\} \in C$.Подпоследовательности разделяются на патчи функцией μ . В результате X преобразуется в последовательность $X^{\mu} = \{X_1^{\mu_{c_1}}, \ldots, X_m^{\mu_{c_k}}\}$, состоящую из m патчей, где каждому патчу ставиться в соответствие состояние c_i . В процессе обработки выборки X^{μ} могут применяться процессы аугментации, выполняться сдвиг данных, изменяться длина патча. Пусть φ способ формирования обучающего подмножества X^{μ} . Разделяя последовательность X^{μ} на обучающую и тестовую часть, формируется обучающая выборка X_{φ}^{μ}.

Тогда возникает необходимость сформировать такое обучающее подмножество, которое позволило бы повысить функционал качества модели обработки последовательности a.

$$Q(a(x, X_{\varphi}^{\mu}), X) \to \max_{\mu,\varphi}. \quad (1)$$

Разделение информационной последовательности и временных рядов на патчи, применение различных способов формирования обучающего подмножества в рассматриваемом подходе (1) дает возможность автоматического выполнения процессов аугментации для систем машинного обучения. В результате образуется искусственный обучающий набор данных, где патчи повторяют свойства элементов подпоследовательности временного ряда. Правильно подобранные патчи по длине, информационным свойствам, содержащихся в них последовательностей данных дают возможность увеличить объём тренировочного набора, повысить устойчивости моделей к вариациям входных последовательностей. Дальнейшее применение классических технических приемов, например, кросс-валидации может служить для повышения обобщающей способности алгоритмов. Формирование выборки патчей дает возможность выполнять балансировку неравномерно представленных объектов наблюдения классов. Такой подход направлен на автоматическое создание выборки данных для систем самоконтролируемого обучения. Последовательность сформированных паттернов обучающей выборки с меткой состояния, в котором они были получены, подается на вход модели обработки. Автоматическая нарезка патчей, которые рассматриваются как объекты наблюдения, и проведение операций по выбору и анализу

способов формирования объектов наблюдения, позволяют анализировать результаты обучения моделей на предмет достижения лучших качественных показателей обработки.

Новизна предлагаемого метода заключается в комплексном подходе, объединяющем автоматическое формирование объектов обучения из временных рядов с адаптивным изменением длины и сдвига подпоследовательностей и анализ функции качества моделей для оценки эффективности обработки. В отличие от традиционных подходов, рассматривается обратная задача. Не модель настраивается и оптимизируется на выборке данных, а наоборот, объекты наблюдения формируются из данных «под модель». Достигаемый показатель качества моделей обработки позволяет выбирать обучающие выборки под конкретные алгоритмы, определять объекты наблюдения с информативными признаками для каждой модели и снижать влияние шумовых элементов.

Предлагаемый подход ориентирован на модели, не требующие значительных вычислительных ресурсов, и обладает относительно простой реализацией. В отличие от PatchTST, где длина подпоследовательности задаётся фиксированным параметром, в предлагаемом методе она определяется динамически на основе функционала качества модели. В Segment/Shuffle (S3) сегменты перемешиваются в оптимальном для задачи порядке, что может влиять на сохранение временных зависимостей, искажать значимые паттерны, зависеть от гипперпараметров.

3.2. Метод формирования и анализа выборок данных. В процессе применения самоконтролируемого обучения, когда на вход подаются последовательности разных классов состояний, возникает необходимость определения образцов для обучения. Большое количество моделей обработки используют поступающие последовательности одинаковой длины. Для их формирования применяется разбиение последовательности на сегменты (патчи). Для этого могут использоваться различные эвристики, отражающие особенности предметной области полученных последовательностей. Предлагаемый метод отличается автоматическим формированием и выбором патчей, на основе анализа качественных показателей моделей обработки, что позволяет осуществлять выбор моделей и способов формирования обучающих последовательностей, где достигаются лучшие значения качественных показателей. Это дает возможность определять локальные закономерности групп объектов наблюдения, настроить длину входной последовательности для

модели, обнаруживать значащие паттерны. Предлагаемый метод предполагает ряд шагов.

Шаг 1. На вход подаются исходные информационные последовательности, полученные в разных состояниях системы.

Шаг 2. Для разделения на патчи фиксированной длины определяется шаг разбиения. В результате образуется обучающая выборка, содержащая набор патчей одинакового размера, помеченных соответствующим классом состояния.

Шаг 3. Обучающая выборка подается на вход заранее выбранных моделей. Происходит формирование обучающего множества, определение совпадающих элементов, применение процессов кросс-валидации.

Шаг 4. Определяется модель, достигающая лучших качественных показателей.

Шаг 5. Далее для каждой длины происходит сдвижка на один объект наблюдения. В результате для каждой длины L формируется L выборок патчей. Осуществляется формирование выборки. Выполняется переход на шаг 3.

Шаг 6. Затем длина патча увеличивается для неё повторяются процессы формирования патчей и сдвигов. Осуществляется переход на шаг 2.

Шаг 7. В конце происходит выбор лучших результатов по значению показателей качества и выбор длины и сдвига для формирования патча.

Шаг 8. Определяются лучшие по значения качественных показателей модели.

На рисунке 1 представлена алгоритмическая последовательность шагов предлагаемого метода. Предлагаемый метод может быть использован в самообучающихся моделях. Поступающий на вход информационный поток подвергается разделению. Полученные патчи рассматриваются как отдельные независимые объекты наблюдений. Например, для задач классификации они могут рассматриваться как многомерные вектора, для прогнозирования – последовательностями, характеризующими элементы ряда. Из них формируется обучающая выборка для последующих процессов обучения и верификации.

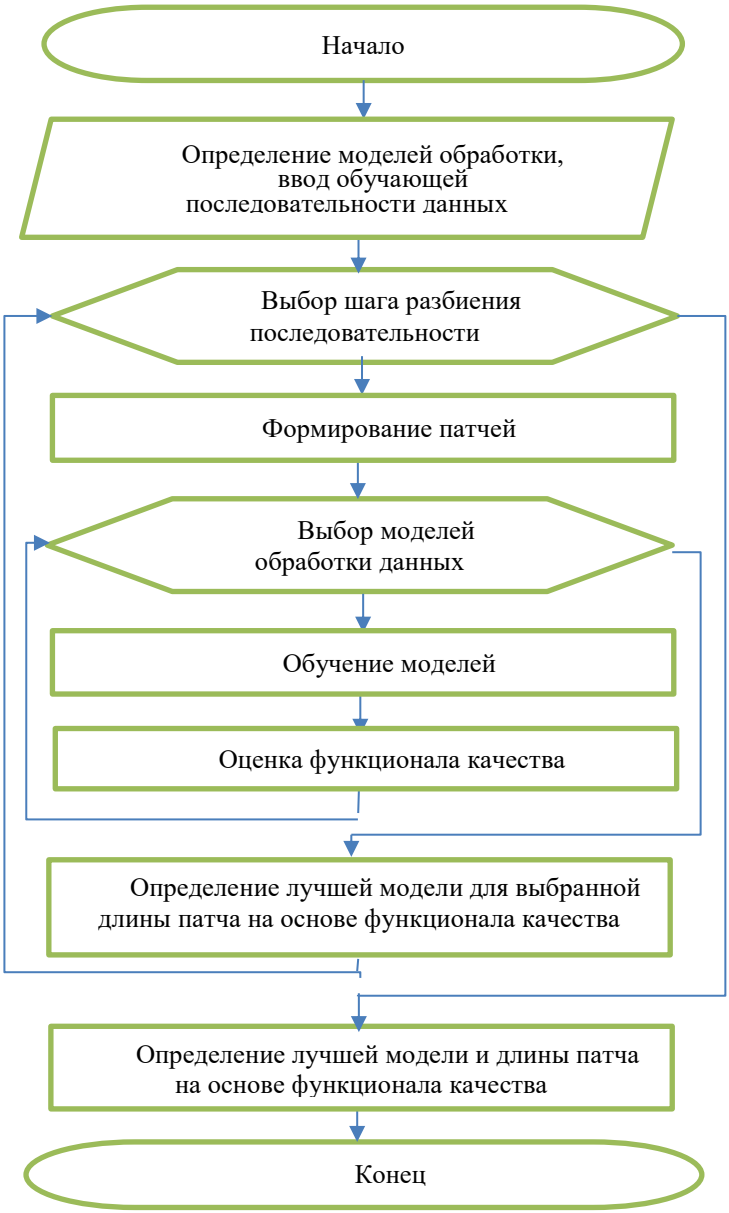


Рис. 1. Алгоритмическая последовательность шагов метода формирования и анализа выборок данных

Формирование патчей производится в автоматическом режиме. При этом возможно рассматривать различные способы расширения. Например, формирования с перекрытием, объединяя в одну выборку все полученные для последовательности одной длины выборки, так и без него. Изменение длины патча зависит от ограниченности вычислительных ресурсов. Для снижения вероятности пропуска значимых свойств информации в последовательности выполняется сдвижка. Эти шаги влияют на объем и количество анализируемых обучающих выборок. Для каждой полученной выборки производится обучение моделей обработки и, с использованием заранее определенного показателя качества осуществляется выбор модели с лучшими значениями.

Применяя многоуровневые модели обработки [32], возможно построение самоорганизующихся самообучающихся систем. Для этого результаты моделей должны сравниваться с реальными значениями объектов наблюдения, полученными от регистрирующих систем и устройств. В случае увеличения ошибок выше заранее определенного порога должно приниматься решение о формировании выборки данных. Над выборкой проводятся действия по формированию патчей, связанных с изменением длины и патчей, происходит очередная настройка и обучение модели.

Предлагаемый метод позволяет реализовывать процессы аугментации. Сегментируя последовательности данных, а затем объединяя выборки и перемешивая объекты наблюдений, полученных в разных условиях, возможно увеличение обучающих примеров.

В отличие от традиционных подходов, обрабатывающих информационные последовательности, этот метод формирует разные сгенерированные подпоследовательностей патчей таким образом, чтобы выбирать патчи, определяя значения функционала качества, свойства которых наилучшим образом соответствуют модели обработки.

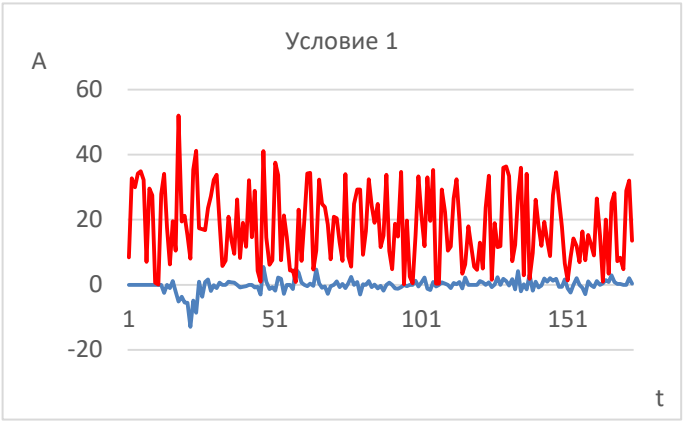
4. Эксперимент

4.1. Параметры эксперимента. В ходе эксперимента рассматривалась возможность создания выборок данных из информационных последовательностей в автоматическом режиме, так же влияние свойств содержащихся в них обучающих примеров на результат обработки различными моделями.

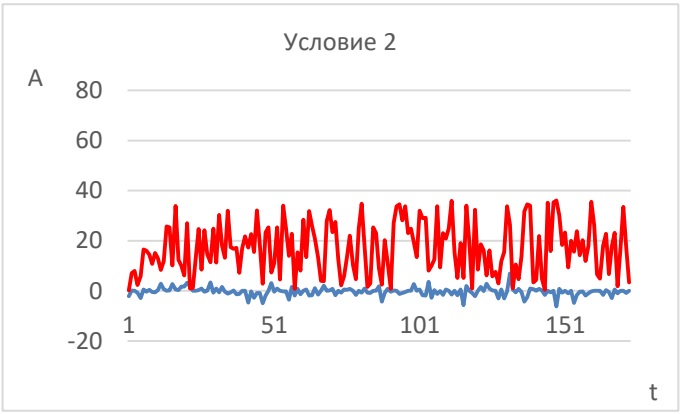
В качестве рассматриваемых данных в разделах 4.1-4.2 были выбраны последовательности датасета [33].

Цель эксперимента состояла в том, чтобы провести анализ влияния на качественные показатели моделей выборок, имеющих для последовательностей разные характеристики.

На вход самоорганизующейся самообучающейся модели подавались ряды, полученные в различных условиях. Их пример приведен на рисунке 2. Амплитуды радиосигналов во временной области двух каналов обозначены красным и синим цветом [34].



а)



б)

Рис. 2. Пример информационных последовательностей

Из поступившей на вход последовательности в автоматическом режиме формировались обучающие выборки с патчами разной длины и сдвига. Для всех выборок анализировались модели обработки на предмет достижения лучших показателей качества.

Состав вектора x определяется размерностью временного ряда и длинной подпоследовательности, выбранной для формирования объектов наблюдения.

Проверка достигаемых показателей качества выполнялась на группе моделей обработки. Их выбор связан с ограничениями на ресурсоемкость. Поэтому в эксперименте приоритет отдавался алгоритмам, имеющим относительно высокую скорость обучения и не требующим больших ресурсов.

На выборках патчей различной длины, при разных сдвигах сравнивались результаты наивного байесовского классификатора (NB), линейного дискриминанта (LD), машины опорных векторов (SVM), метода K-ближайших соседей (KNN), деревьев решений (DT) и ансамбля алгоритмов (ENS), включающего в себя деревья классификации, допускающих максимум 10 разбиений. Гиперпараметры настройки моделей определены Matlab в [34] и использовались «заданными по умолчанию». Реализация процедур предлагаемого метода представлена в [35]

Формирование патчей информационных последовательностей происходило на основе неперекрывающихся окон. Функция разбиения разделяла последовательность на дискретные блоки. Образовывались обучающие выборки примеров, имеющих различные длины, а для каждой длины выполнялся сдвиг, т.е. для длины n формировалось n обучающих выборок. Оценка предлагаемого метода осуществлялась для задачи бинарной классификации. Формируемые выборки делились в соотношении 90:10 и 70:30 на обучающие и тестовые части. Длина объектов наблюдения, образуемых последовательностями, в выборке изменялась от 2 до 70 отсчетов. Один сдвиг осуществлялся на один отчет. Количество сдвигов равнялось длине.

Каждая полученная выборка подавалась на вход всем моделям. Оценивался заранее заданный показатель качества полноты:

$$accuracy = (TP + TN)/n, \quad (2)$$

где n – количество наблюдений, используемых для оценки модели, TP и TN – верно распознанные положительные и отрицательные объекты наблюдения.

В процессе обучения анализировались кортежи, определяющие из множества моделей A модель обработки $a_i \in A$, обучающую выборку, полученную методом разделения, и метод разделения на патчи $< a_i(x), X^{\mu_j}, \mu_j >$. Каждому кортежу в процессе обучения определялось значение функционала качества (2).

Фреймворк эксперимента представлен на рисунке 3.

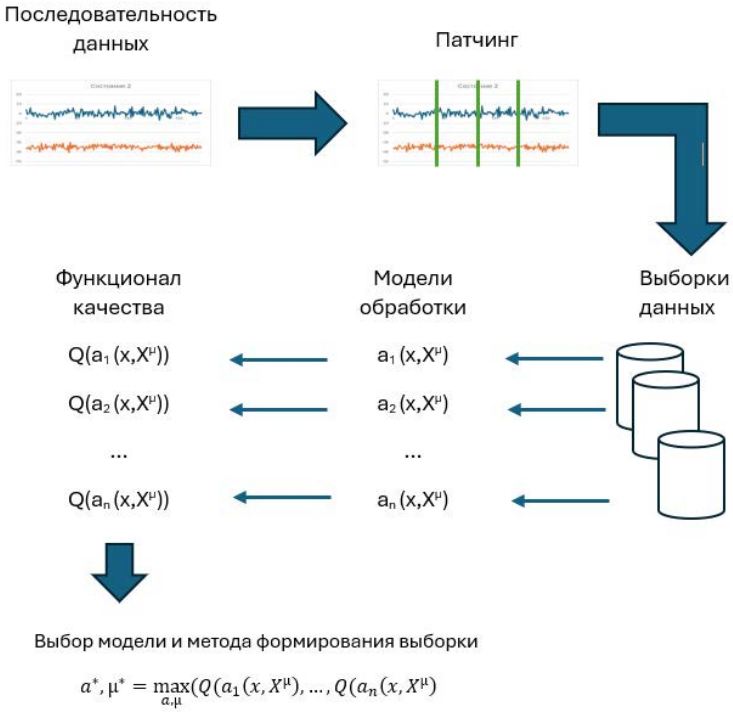


Рис. 3. Фреймворк проведения эксперимента

В результате обработки всех выборок становиться возможным выделить модель, метод формирования патчей и его характеристики, где достигаются лучшие значения выбранного функционала качества.

4.2. Проведение эксперимента. При проведении эксперимента на вход подавалась выборка временных рядов, полученных различных состояниях. На следующем шаге выполнялось разделение последовательности на патчи.

Способ формирования патчей для эксперимента представлен на рисунке 4.

Последовательности, полученные в состояниях 1 и 2, разбивались на патчи длиной *w*. Для каждой длины формировалось *w* сдвигов. Рассматривалось количество w последовательностей для каждой длины.



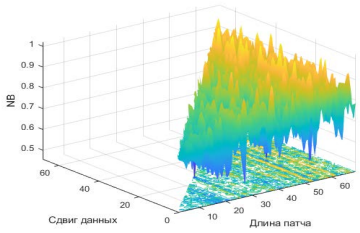
Рис. 4. Формирование патчей (w-ширина окна, 1 и 2 в верхнем индексе – состояние, нижний индекс – сдвиг)

Из полученных патчей составлялась выборка, которая разделялась на обучающую и тестовую части. Множеству патчей ставился в соответствие класс состояния последовательности, из которого оно было получены. Одна часть была помечена классом «состояние 1», а другая – классом «состояние 2». Последовательно изменяя длину шага разбиения, формировались патчи различной длины, получались выборки, содержащие обучающие примеры подпоследовательностей с разным количеством отсчетов и с различным сдвигом.

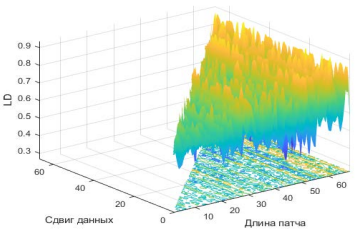
В результате для рассматриваемой задачи классификации в автоматическом режиме формировалась выборка данных, где каждый патч может рассматриваться как многомерный вектор, описывающий объект наблюдения.

Далее каждая обучающая выборка подавалась на вход классифицирующих алгоритмов, которые обучались и тестировались на ней. В результате для каждой модели и каждой выборки получались значения качественных показателей.

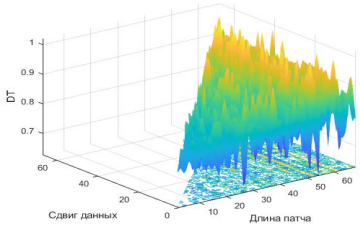
На рисунке 5 в осях «длина патча-сдвиг данных-точность(accuracy)» представлены значения для рассматриваемых моделей при соотношении 90:10 обучающих и тестирующих примеров. Графики показывают наличие областей, где достигается различные значения показателей качества. Для классифицирующих алгоритмов при одних способах формирования патчей значения достигаемых качественных показателей лучше, на других – хуже. Лучшие значения accuracy для рассматриваемых данных получаются на длинах 52 отсчета.



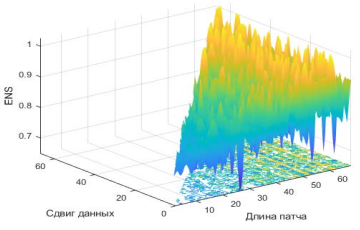
а) байесовский классификатор (NB)



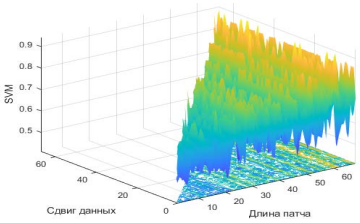
б) линейный дискриминант (LD)



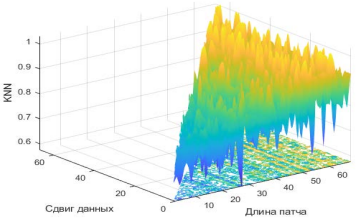
в) метод деревьев решений (DT)



г) ансамбль алгоритмов (ENS)



д) машина опорных векторов (SVM)

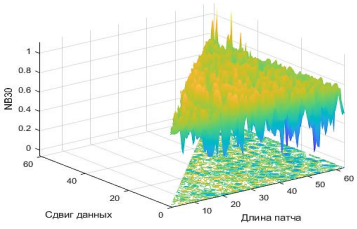


е) метод к-ближайших соседей (KNN)

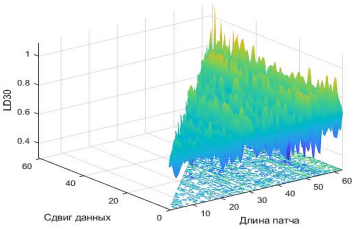
Рис. 5. Точность accuracy моделей при соотношении 90:10 для обучающих и тестирующих примеров

Увеличение размера объекта наблюдения приводит к появлению дополнительных признаков, которые могут содержать «значимую» информацию, улучшающую работу алгоритмов обработки. Дальнейший рост размера объекта наблюдения может привносить в его состав «шумовые» признаки и ухудшать обобщающую способность моделей. Изменение сдвига на фиксированной длине при формировании обучающей выборки оказывает влияние на количество дублирующихся, очень похожих примеров. Увеличение их количества может ухудшать, а уменьшение – улучшать достигаемые качественные показатели.

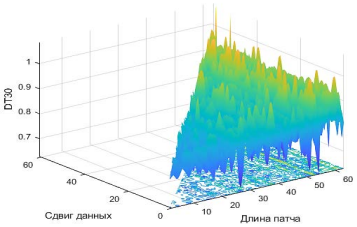
Такие же области получаются при изменении соотношений обучающих и тестовых примеров. На рисунке 6 показаны значения для тех же выборок данных в тех же осях при соотношении 70:30 для обучающих примеров и тестирующих примеров.



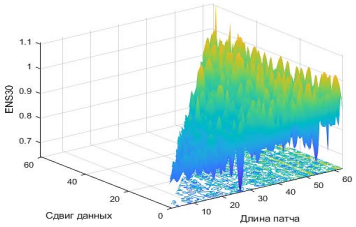
а) байесовский классификатор (NB)



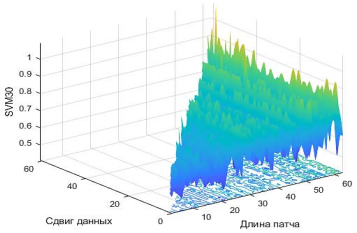
б) линейный дискриминант (LD)



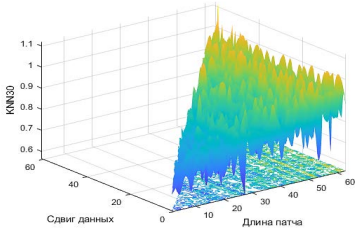
в) метод деревьев решений (DT)



г) ансамбль алгоритмов (ENS)



д) машина опорных векторов (SVM)



е) метод к-ближайших соседей (KNN)

Рис. 6. Точность accuracy моделей при соотношении 70:30 для обучающих и тестирующих примеров

Графики показывают, что для анализируемых данных для соотношений 90:10 и 70:30 области высоких и низких значений показателей качества accuracy прослеживаются и совпадают. Судя по графикам определенные длины и сдвиги оказывают влияние на достигаемые значения качества моделей. На рисунке 6 показана

развёртка значений accuracy по значениям «длина патча-сдвиг» при соотношении 70:30 обучающих и тестирующих примеров.

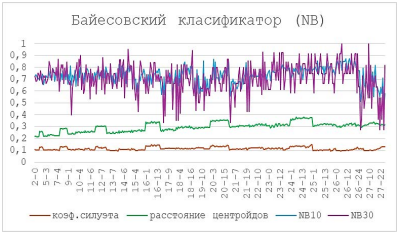
На рисунке 7 для оценки свойств обучающих примеров, полученных для каждой длины и сдвига выборки, были определены значения коэффициента силуэта и расстояния центройдов. Условно объекты наблюдения обучающей выборки разбивались на две группы, описывающие «состояние 1» и «состояние 2». Тогда, чем больше расстояние между центроидами групп обучающих примеров, помеченных разными классами, тем сильнее различаются группы. Это показывает на графиках рост и всплески зеленой линии расстояния центройдов. С другой стороны, для оценки свойств таких групп можно применить «коэффициент силуэта» (красная линия). Он показывает для группы объектов наблюдения насколько точно каждый объект соответствует своей группе и насколько он отличается от других групп.

Представленные в нижней части графиков линии коэффициента силуэта и расстояния центройдов показывают, что получаемые выборки объектов наблюдения в результате сдвига и изменения размера патча неоднородны.

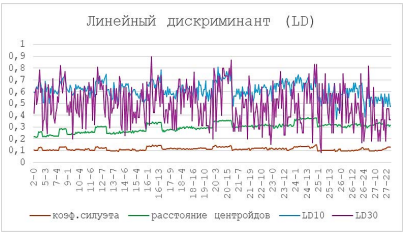
В обучающей выборке объекты наблюдения демонстрируют разнородность свойств. Среди них находятся как эталонные объекты, выполняющие функцию репрезентативных представителей класса, так и неинформативные, находящиеся на периферии границ классов или окруженные другими объектами, принадлежащими разным классам. Большое количество неинформативных, периферийных, шумовых объектов наблюдения негативно отражается на качестве классификации.

При формировании выборки изменения размера патчей и сдвиги внутри них могут влиять на относительное количество появления дублирующих или схожих примеров, затрудняющих процесс обучения модели.

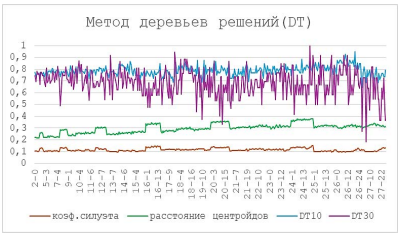
Тем не менее, графики показывают, что для различных соотношений обучающих и тестовых примеров, где наблюдается рост расстояния центройдов и увеличивается коэффициент силуэта прослеживается повышения значений качества accuracy, а при их падении – происходит понижение значений этого показателя качества.



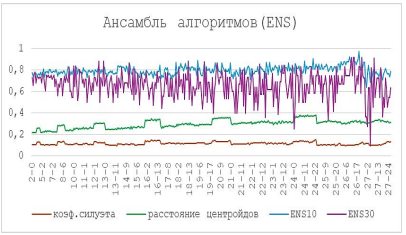
а) байесовский классификатор (NB)



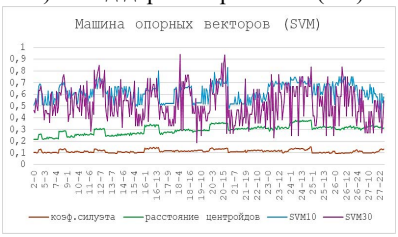
б) линейный дискриминант (LD)



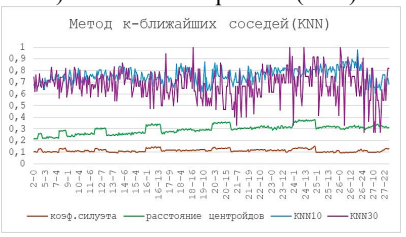
в) метод деревьев решений (DT)



г) ансамбль алгоритмов (ENS)



д) машина опорных векторов (SVM)



е) метод к-ближайших соседей (KNN)

Рис. 7. Значения accuracy (ось ординат) по совмещенной оси абсцисс «длина патча-сдвиг» для соотношений 90:10 и 70:30 обучающей и тестовой выборки

В рассматриваемом эксперименте анализируется бинарная классификация. Оценка качества обработки последовательности ошибки классифицирующих моделей определяются выражением:

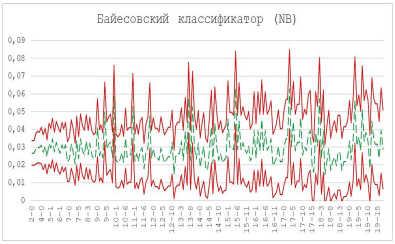
$$\rho = 1 - accuracy. \tag{3}$$

Предполагая гауссовское распределение, доверительный интервал ошибки классификаторов (3) (наивного Байеса, линейного дискриминанта, деревьев решений, ансамблевых классификаторов, KNN с гауссовском ядром) рассчитывается по формуле:

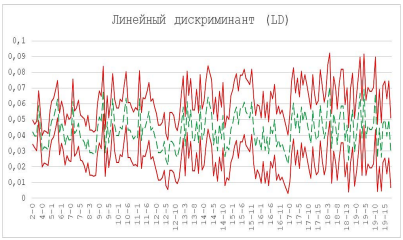
$$p \pm 1.96\sqrt{\frac{p(1-p)}{n}}, \quad (4)$$

где p – ошибки классификации, 1,96 – критическое значение гауссова распределения для вероятности 95%, n – количество наблюдений, используемых для оценки модели.

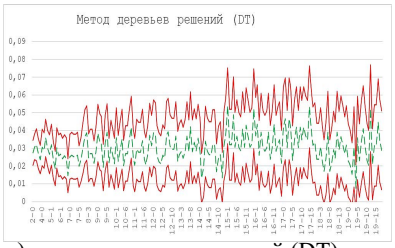
На рисунке 8 представлены изменения доверительных интервалов (4) классификаторов.



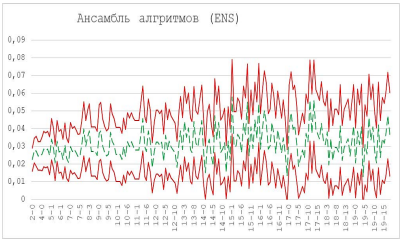
а) байесовский классификатор (NB)



б) линейный дискриминант (LD)



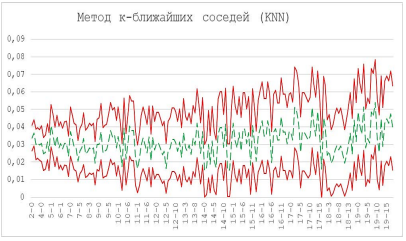
в) метод деревьев решений (DT)



г) ансамбль алгоритмов (ENS)



д) машина опорных векторов (SVM)



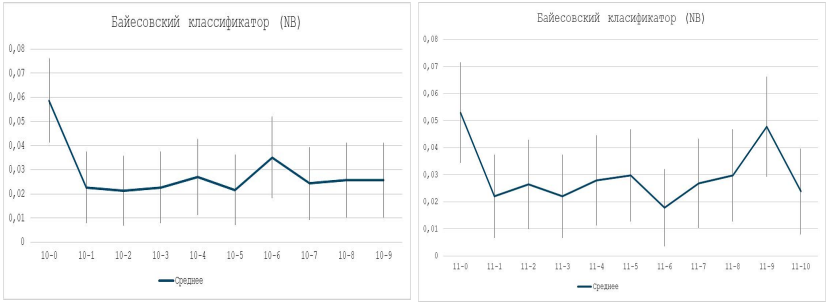
е) метод к-ближайших соседей (KNN)

Рис. 8. Доверительные интервалы классификаторов. Ось ординат – ошибка классификатора (зеленая линия) и его доверительный интервал (красная линия). Ось абсцисс – значения «длина патча-сдвиг»

Графики показывают, что по мере увеличения размера патча заметен рост доверительного интервала ошибки. Это связано

с ограниченностью объема выборки данных. Увеличение размера длины патча приводит к тому, что сокращается количество наблюдений, используемых для оценки модели. При формировании обучающей выборки необходимо оценивать её объем. Небольшой размер может преувеличивать точность модели из-за переобучения или случайных эффектов. Ограничения статистических методов оценки рекомендуют значения $n>=30$ для достижения значимых результатов, что является определенной границей при формировании множества обучающих примеров.

Тем не менее, на отдельных значениях длины патча и сдвига наблюдаются тенденции к улучшению или ухудшению результатов, где диапазоны доверительных интервалов не перекрываются. Рассматривая отдельные длины и сдвиги, образуемые для них, можно видеть неперекрывающиеся доверительные интервалы ошибок классификаторов. Например, на рисунке 9 для наивного Байеса представлены значения, ошибок для разных длин и сдвигов.



а) Ошибка на длине патча 10 б) Ошибка на длине патча 11

Рис.9 Доверительные интервалы на патчах разной длины. Ось ординат – ошибка классификатора и его доверительный. Ось абсцисс – значения «длина патча-сдвиг»

Графики рисунка 9 показывают существование неперекрывающихся доверительных интервалов (4), что позволяет говорить о влиянии длины и сдвига при формировании патча на достигаемые значения показателя качества.

4.3. Проверка на сторонних выборках данных.

В качестве анализируемых могут рассматриваться временные ряды, описывающие любые процессы, где, в зависимости от задачи классификации состояния, необходимо определить объекты из подпоследовательности, обладающими лучшими характеристиками

для достижения показателей качества. Для проверки были выбраны три набора.

Набор данных для обучения и тестирования, собранный компанией Ford в типичных условиях эксплуатации с минимальным шумовым загрязнением. Задача классификации состоит в диагностике наличия или отсутствия определенного симптома в автомобильной подсистеме [36].

Набор данных PowerCons содержит данные об индивидуальном потреблении электроэнергии домохозяйствами за один год, распределенные по двум сезонным классам: теплому (класс 1) и холодному (класс 2) [37].

Набор данных компании Sony для робота, оснащенного акселерометрами для измерения крена, тангажа и рыскания. Данные представлены только по оси X. Задача состоит в определении поверхности, по которой ходит робот (цемент или ковер) [38].

Для наборов данных проводилась нормировка, удаление дубликатов, неправильных значений выбросов, осуществлялось формирование последовательностей, которым ставилась метка состояния/класса для проверки метода.

Проверка ошибки алгоритмов классификации (4) вычислялась при формировании патчей и сдвигов на последовательностях других датасетов [36 – 38].

Определялись подпоследовательности, полученные для различных длин окна и сдвигов, затем выбирались средние значения и диапазоны доверительных интервалов. Размер патча для рассматриваемых датасетов варьировался от 2 до 20 отсчетов временного ряда.

В таблице 1 приведены достигаемые значения ошибок алгоритмов при формировании выборок данных описанным способом.

Значения ошибок алгоритмов, представленных в таблице 1 показывают, что выбор длины патча и сдвига оказывает влияние на результаты обработки последовательностей рассмотренными классифицирующими алгоритмами. Применяя описанный способ формирования обучающих выборок данных возможно повышение качественного показателя. Сравнение значений ошибок классификации позволяет определить, что сегментация выборки в большинстве случаев улучшает показатели качества обработки. Результаты, приведенные в таблице 1, показывают, что выбор размера патча и сдвига данных последовательности, в основном, улучшает показатели на 3-8%.

Применение патчинга со сдвигом изменяет свойства данных. Его влияние на слабые модели проявляется в большей степени, чем

на сильные. Результаты показывают возможность увеличения accuracy на 6-10% для слабых моделей, в то время как для сильных моделей наблюдается улучшение на 1-5% в сценариях с ограниченными данными.

Диапазоны лучших значений показателя (2) классифицирующих алгоритмов в таблице 1 в большинстве выборок для рассматриваемых классификаторов не пересекаются, что говорит о статистических значимых результатах

Таблица 1. Ошибки классифицирующих алгоритмов на выборках [36 – 38]

	Алгоритм	**Средняя ошибка алгоритма на датасете**	**Лучшее значение ошибки алгоритма**		
			Ошибка с доверительным интервалом	**Длина**	**Сдвиг**
Датасет 1	NB	0,11	0,016±0,012	9	1
	LD	0,09	0,013±0,011	9	8
	SVM	0,12	0,029±0,016	9	7
	DT	0,06	0,005±0,004	5	1
	KNN	0,07	0,005±0,004	5	3
	ENS	0,05	0,005±0,004	5	3
Датасет 2	NB	0,1	0,014±0,007	10	0
	LD	0,08	0,012±0,006	10	0
	SVM	0,13	0,016±0,009	10	2
	DT	0,09	0,013±0,003	4	1
	KNN	0,07	0,012±0,003	5	2
	ENS	0,06	0,012±0,003	4	0
Датасет 3	NB	0,16	0,019±0,014	11	4
	LD	0,11	0,013±0,012	9	7
	SVM	0,21	0,028±0,015	10	5
	DT	0,06	0,007±0,002	6	0
	KNN	0,05	0,009±0,007	6	2
	ENS	0,05	0,007±0,003	10	2

Проведённый эксперимент показывает, что варьирование параметрами сдвига и длины последовательности при формировании обучающих выборок оказывает влияние на эффективность обработки данных. При этом формируемые патчи должны адекватно отражать значащие свойства последовательностей, чтобы создать предпосылки для достижения классифицирующими алгоритмами более высоких показателей качества обработки информации.

Предлагаемый метод, использующий автоматическое формирование и выбор патчей на основе анализа достигаемых значений моделями обработки выбранного функционала качества, способствует снижению влияния шумовых компонентов в исходных данных.

Полученные результаты позволяют утверждать, что внедрение данного подхода увеличивает робастность алгоритмов классификации к вариациям входных данных.

4.4 Обсуждение результатов эксперимента. Формирование обучающей выборки приводит к тому, что объекты наблюдения в ней обладают разными свойствами. Часть из них являются эталонными экземплярами, позволяющими характеризовать класс, а другая часть – неинформативные, периферийные, шумовые, которые могут негативно влиять на качество классификации.

Эксперименты показывают, что оптимальное значение количества патчей сильно зависит от таких факторов, как сложность набора данных, длина, сдвиг. В зависимости от размера разделенных подпоследовательностей можно искать сдвиги, где захватываются значимые признаки, благодаря которым модель повышает качество обработки.

Изменение сдвига при формировании обучающей выборки приводит к тому, что могут появляться дублирующиеся или очень похожие примеры, что мешает обучению. Увеличение количества таких объектов приводит к настройке модели на отдельные характеристики данных в ущерб другим. Кроме того, может возникать ситуация, когда неконтролируемый осуществляемый в автоматическом режиме сдвиг формирует большое количество похожих или одинаковых объектов в одном классе, что может приводить к дисбалансу обучающих данных. Варьирование параметром сдвига может приводить как к возникновению эффекта шума, когда появляются значительное количество примеров на границах классов, так и к исчезновению таких объектов.

Изменение длины последовательности также оказывает влияние на качественные показатели моделей обработки. При рассмотрении

объекта наблюдения в виде вектора увеличение его размера приводит к появлению дополнительных признаков, которые могут как улучшать работу алгоритмов классификации путем внесения значимой информации, так и ухудшать обобщающую способность. В случае равномерного распределения объектов в признаковом пространстве добавление признака, может повысить удаленность объектов разных классов.

Значительное увеличение количества характеристик объектов наблюдения обучающей выборки может влиять на результат классификации из-за проклятия размерности. Появление лишних признаков, слабо влияющих на отклик модели снижает показатели качества обработки.

5. Заключение. Несмотря на значительные успехи и перспективы автоматического формирования анализируемых объектов наблюдения для самоорганизующихся систем, их обучение требует решения ряда проблемных вопросов для обработки информационных последовательностей.

Автоматическое определение структур данных и формирование обучающих, тестовых, валидационных выборок является одним из важнейших процессов, который сильно влияет на достигаемые значения показателей качества моделями обработки. Однако этим вопросам уделяется недостаточно внимания. В большинстве исследований формирование выборок данных осуществляется эмпирически. Для повышения качества обработки данных по-прежнему применяются различные эвристики, подходящие под рассматриваемые информационные свойства последовательностей, но не позволяющие производить обобщение в рамках теоретических и практических подходов.

Рассматриваемый в статье метод направлен на автоматическое формирование выборки на основе значений функционала качества. Применение разделения временного ряда на патчи и осуществление сдвигов при формировании объектов наблюдения является относительно универсальным способом и позволяет осуществлять построение данных под модель обработки.

Однако необходимо отметить, что эффективное применение описанного в статье метода зависит от ряда условий. Информационные последовательности обладают различными свойствами, влияющими на качество процессов обработки моделями. В результате такой «нарезки» могут возникать ситуации, когда размер патча не позволит выделить паттерн со значащей для обработки информации, или размер паттерна окажется слишком большим и будет

содержать много шумовых компонент, негативно влияющих на качество обработки.

На достигаемые показатели качества модели влияет размер обучающей выборки, но при автоматическом разбиении на патчи это влияние может быть разным. Изменение размера патча и манипуляции сдвигом при формировании выборки, повышающие различимость объектов наблюдения, принадлежащих разным классам, увеличивающие количество эталонные и уменьшающих число шумовых данных повышает точность обработки. В то же время эти манипуляции могут не только не приводить к существенным изменениям точности, но и ухудшать результаты. Это происходит, когда полученные таким образом обучающие объекты наблюдения оказываются нерепрезентативными или неинформативными. Возможность возникновения подобных ситуаций обуславливает необходимость валидационной поверки моделей обучения.

При формировании обучающей выборки необходимо оценивать её объем. Небольшой размер может преувеличивать точность модели из-за переобучения или случайных эффектов.

Тем не менее автоматические процессы разделение информационных последовательностей, формирование обучающих примеров, и их оценка на основе функционала качества в большинстве случает дают возможность повысить качество обработки.

Литература

1. Jarantow S.W., Pisors E.D., Chiu M.L. Introduction to the Use of Linear and Nonlinear Regression Analysis in Quantitative Biological Assays // Current Protocols. 2023. vol. 3. 801 p. DOI: 10.1002/cpz1.801.
2. He Y., Zhang X., Kong X., Yao L., Song Z. Causality-driven sequence segmentation assisted soft sensing for multiphase industrial processes // Neurocomputing. 2025. vol. 631 p. 129612. DOI: 10.1016/j.neucom.2025.129612.
3. Ци Д., Буре В.М. Исследование методов прогнозирования временных рядов для предсказания качества воздуха: объяснительный сравнительный анализ // Вестник Санкт-Петербургского университета. Прикладная математика. Информатика. Процессы управления. 2024. Т. 20. №2. С. 206–219. DOI: 10.21638/spbu10.2024.206.
4. Man T., Osipov V.Yu., Zhukova N., Subbotin A.N., 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. 102427 p. DOI: 10.1016/j.inffus.2024.102427.
5. Belitser E., Ghosal S. Bayesian uncertainty quantification and structure detection for multiple change points models // Bernoulli. 2025. vol. 31. no. 2. pp. 1181–1205. DOI: 10.3150/24-BEJ1766.
6. Xiao Z., Xing H., Qu R., Feng L., Luo S., Dai P., Zhao B., Dai Y. Densely Knowledge-Aware Network for Multivariate Time Series Classification // IEEE Transactions on Systems, Man, and Cybernetics: Systems. 2024. vol. 54. no. 4. pp. 2192–2204. DOI: 10.1109/TSMC.2023.3342640.

7. Liu M., Zeng A., Chen M., Xu Z., Lai Q., Ma L., Xu Q. Scinet: Time Series Modeling and Forecasting With Sample Convolution and Interaction // Advances in Neural Information Processing Systems. 2022. vol. 35. pp. 1–13.
8. Lebedev I.S., Sukhoparov M.E. Adaptive Learning and Integrated Use of Information Flow Forecasting Methods // Emerging Science Journal. 2023. vol. 7. no. 3. pp. 704–723.
9. Chen D., Chen L., Zhang Y., Wen B., Yang C. A Multiscale Interactive Recurrent Network for Time-Series Forecasting // IEEE Transactions on Cybernetics. 2021. vol. 52. no. 9. pp. 8793–8803. DOI: 10.1109/TCYB.2021.3055951.
10. Лебедев И.С. Адаптивное применение моделей машинного обучения на отдельных сегментах выборки в задачах регрессии и классификации // Информационно-управляющие системы. 2022. №3. С. 20–30. DOI: 10.31799/1684-8853-2022-3-20-30.
11. Vaswani A., Shazeer N., Parmar N., Uszkoreit J., Jones L., Gomez A.N., Kaiser L., Polosukhin I. Attention is All you Need // Advances in Neural Information Processing Systems. 2017. vol. 30. pp. 1–11.
12. Wang J., Jiang J., Jiang W., Li C., Zhao W.X. LibCity: An open library for traffic prediction // Proceedings of the 29th International Conference on Advances in Geographic Information Systems, SIGSPATIAL'21. 2021. pp. 145–148. DOI: 10.1145/3474717.3483923.
13. Grover S., Jalali A., Etemad A. Segment, Shuffle, and Stitch: A Simple Layer for Improving Time Series Representations // Proceedings of the 38th Conference on Neural Information Processing Systems (NeurIPS). 2024. pp. 4878–4905. DOI: 10.52202/079017-0158.
14. Salehi A., Balasubramanian M. DDCNet: Deep dilated convolutional neural network for dense prediction // Neurocomputing. 2023. vol. 523. pp. 116–129. DOI: 10.1016/j.neucom.2022.12.024.
15. Chen Y., Lu X., Xie Q. Collaborative networks of transformers and convolutional neural networks are powerful and versatile learners for accurate 3D medical image segmentation // Computers in Biology and Medicine. 2023. vol. 164 p. 107228. DOI: 10.1016/j.compbiomed.2023.107228.
16. Woo G., Liu C., Kumar A., Xiong C., Savarese S., Sahoo D. Unified Training of Universal Time Series Forecasting Transformers // Proceedings of the 41st International Conference on Machine Learning (ICML). 2024. pp. 1–25.
17. Ekambaram V., Ati A., Nguyen N., Sinthong P., Kalagnanam J. TSMixer: Lightweight MLP-Mixer Model for Multivariate Time Series Forecasting // Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD '23). 2023. pp. 459–469. DOI: 10.1145/3580305.3599533.
18. Goswami M., Szafer K., Choudhry A. et al. Moment: A Family of Open Time-Series Foundation Models // Proceedings of the 41st International Conference on Machine Learning. vol. 235. 2024. pp. 16115–16152.
19. Vishwas B.V.K., Macharla S.R. MOIRAI: A Time Series LLM for Universal Forecasting // Time Series Forecasting Using Generative AI. 2025. pp. 183–194. DOI: 10.1007/979-8-8688-1276-7_7.
20. Yang H., Yang D. CSwin-PNet: A CNN-Swin Transformer combined pyramid network for breast lesion segmentation in ultrasound images // Expert Systems with Applications. 2023. vol. 213. part b. 119024 p. DOI: 10.1016/j.eswa.2022.119024.
21. Xiao Q., Wu B., Zhang Yu., Liu S., Pechenizkiy M., Mocanu E., Mocanu D.C. Dynamic Sparse Network for Time Series Classification: Learning What to "see" // Advances in Neural Information Processing Systems. 2022. vol. 35. pp. 1–14.
22. Liu M., Zeng A., Chen M., Xu Z., Lai Q., Ma L., Xu Q. Scinet: Time Series Modeling and Forecasting with Sample Convolution and Interaction // Advances in Neural Information Processing Systems. 2022. vol. 35. pp. 5816–5828.

23. Zhou H., Zhang S., Peng J., Zhang S., Li J., Xiong H., Zhang W. Informer: Beyond efficient transformer for long sequence time-series forecasting // AAAI Conference on Artificial Intelligence. 2021. vol. 35. no. 12. pp. 11106–11115.
24. Wu H., Xu J., Wang J., Long M. Autoformer: Decomposition transformers with auto-correlation for long-term series forecasting // Advances in Neural Information Processing Systems. 2021. vol. 34. pp. 22419–22430.
25. Chen Y., Ren K., Wang Y., Fang Y., Sun W., Li D. Contiformer: Continuous-time transformer for irregular time series modeling // Advances in Neural Information Processing Systems. 2024. vol. 36. pp. 47143–47175.
26. Zeng A., Chen M., Zhang L., Xu Q. Are transformers effective for time series forecasting? // Proceedings of the AAAI Conference on Artificial Intelligence. 2023. vol. 37. no. 9. pp. 11121–11128.
27. Lee S., Park T., Lee K. Soft contrastive learning for time series // The International Conference on Learning Representations. 2024. vol. 37. pp. 11121–11128.
28. Tallman E., West M. Bayesian predictive decision synthesis // Journal of the Royal Statistical Society. Series B: Statistical Methodology. 2024. vol. 86. no. 2. pp. 340–363.
29. Lebedev I.S., Sukhoparov M.E. Adaptive Learning and Integrated Use of Information Flow Forecasting Methods // Emerging Science Journal. 2023. vol. 7. no. 3. pp. 704–723.
30. Peng Y., Hu X., Hao X., Liu P., Deng Y., Li Z. Spider-Net: High-resolution multi-scale attention network with full-attention decoder for tumor segmentation in kidney, liver and pancreas // Biomedical Signal Processing and Control. 2024. vol. 93. p. 106163. DOI: 10.1016/j.bspc.2024.106163.
31. Lebedev I., Sukhoparov M., Semenov V., Khasanov D. Adaptive Segmentation of Information Sequences for Machine Learning Modular Regression Models // Emerging Science Journal. 2025. vol. 9. no. 5. pp. 2420–2438.
32. Shi X., Song X., Deng M., Zhang D., Li X., Chen B. UNet and Swin Transformer Fusion Network for Lesion Segmentation in Biological Kidney Imaging // International Journal of Pattern Recognition and Artificial Intelligence. 2025. vol. 39. no. 12. 2550021 p. DOI: 10.1142/s0218001425500211.
33. Kaggle. E-Commerce Data. URL: https://www.kaggle.com/datasets/carrie1/ecommerce-data (дата обращения: 01.12.2025).
34. Kaggle. Hourly energy demand generation and weather. URL: https://www.kaggle.com/nicholasjhana/energy-consumption-generation-prices-and-weather/data (дата обращения: 01.12.2025).
35. Kaggle. Pima Indians Diabetes Database. URL: https://www.kaggle.com/datasets/uciml/pima-indians-diabetes-database (дата обращения: 01.12.2025).

Лебедев Илья Сергеевич — д-р техн. наук, профессор, главный научный сотрудник, лаборатория интеллектуальных систем, Федеральное государственное бюджетное учреждение науки «Санкт-Петербургский Федеральный исследовательский центр Российской академии наук» (СПб ФИЦ РАН). Область научных интересов: методы машинного обучения, представление и обработка слабоструктурированных данных, применение методов искусственного интеллекта в системах информационной безопасности. Число научных публикаций — 200. isl_box@mail.ru; 14-я линия В.О., 39, 199178, Санкт-Петербург, Россия; р.т.: +7(812)328-3311.

Поддержка исследований. Исследование выполнено за счет гранта Российского научного фонда № 25-21-00269, https://rscf.ru/project/25-21-00269/.

DOI 10.15622/ia.25.3.1

I. LEBEDEV

ADAPTIVE DATA SAMPLES GENERATION FOR SELF-ORGANIZING CONTRAST-BASED LEARNING SYSTEMS

Lebedev I. **Adaptive Data Samples Generation for Self-Organizing Contrast-Based Learning Systems.**

Abstract. The application of contrast-based learning and self-organizing models for classification and forecasting tasks in processing time series and information sequences faces a number of data organization challenges. Improving the processing quality of such systems necessitates improving the methods for generating observation objects and training data samples. This article proposes a method for generating and analyzing data samples based on selecting information sequence patches with different length and shift characteristics, distinguished by the use of the quality functional of processing models. An experiment on simulated data and samples is used to evaluate the proposed method. Accuracy values are obtained for different processing algorithms and with different patch lengths and shifts. The properties of the resulting data patches are determined using the silhouette coefficient and inter-centroid distance metrics. The errors of the classification algorithms are analyzed, and confidence intervals for the errors are determined. It is determined that at a statistical significance level of $p = 0.05$, changing the patch length and shift affects the achieved accuracy values of the classification algorithms. The proposed method improves accuracy by selecting the length and shift during patch generation and assigning the models with the best performance. Results show a potential improvement of 6–10% for weak models, while strong models achieve an improvement of 1–5% in data-constrained scenarios. Analysis of the proposed solution shows that varying the shift and sequence length parameters when generating training data sets impacts data processing efficiency.

Keywords: machine learning, adaptive models, processing quality improvement, patch generation.

References

1. Jarantow S.W., Pisors E.D., Chiu M.L. Introduction to the Use of Linear and Nonlinear Regression Analysis in Quantitative Biological Assays. Current Protocols. 2023. vol. 3. 801 p. DOI: 10.1002/cpz1.801.
2. He Y., Zhang X., Kong X., Yao L., Song Z. Causality-driven sequence segmentation assisted soft sensing for multiphase industrial processes. Neurocomputing. 2025. vol. 631 p. 129612. DOI: 10.1016/j.neucom.2025.129612.
3. Qi D., Bure V.M. [Research on Time Series Forecasting Methods for Air Quality Prediction: An Explanatory Comparative Analysis]. Vestnik Sankt-Peterburgskogo universiteta. Prikladnaya matematika. Informatika. Protsessy upravleniya – Vestnik of Saint Petersburg University. Applied Mathematics. Computer Science. Control Processes. 2024. vol. 20. no. 2. pp. 206–219. DOI: 10.21638/spbu10.2024.206. (In Russ.).
4. Man T., Osipov V.Yu., Zhukova N., Subbotin A.N., 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. 102427 p. DOI: 10.1016/j.inffus.2024.102427.
5. Belitser E., Ghosal S. Bayesian uncertainty quantification and structure detection for multiple change points models. Bernoulli. 2025. vol. 31. no. 2. pp. 1181–1205. DOI: 10.3150/24-BEJ1766.

6. Xiao Z., Xing H., Qu R., Feng L., Luo S., Dai P., Zhao B., Dai Y. Densely Knowledge-Aware Network for Multivariate Time Series Classification. IEEE Transactions on Systems, Man, and Cybernetics: Systems. 2024. vol. 54. no. 4. pp. 2192–2204. DOI: 10.1109/TSMC.2023.3342640.
7. Liu M., Zeng A., Chen M., Xu Z., Lai Q., Ma L., Xu Q. Scinet: Time Series Modeling and Forecasting With Sample Convolution and Interaction. Advances in Neural Information Processing Systems. 2022. vol. 35. pp. 1–13.
8. Lebedev I.S., Sukhoparov M.E. Adaptive Learning and Integrated Use of Information Flow Forecasting Methods. Emerging Science Journal. 2023. vol. 7. no. 3. pp. 704–723.
9. Chen D., Chen L., Zhang Y., Wen B., Yang C. A Multiscale Interactive Recurrent Network for Time-Series Forecasting. IEEE Transactions on Cybernetics. 2021. vol. 52. no. 9. pp. 8793–8803. DOI: 10.1109/TCYB.2021.3055951.
10. Lebedev I.S. [Adaptive application of machine learning models on individual sample segments in regression and classification tasks]. Informatsionno-upravliaiushchie sistemy – Information and Control Systems. 2022. no. 3. pp. 20–30. DOI: 10.31799/1684-8853-2022-3-20-30. (In Russ.).
11. Vaswani A., Shazeer N., Parmar N., Uszkoreit J., Jones L., Gomez A.N., Kaiser L., Polosukhin I. Attention is All you Need. Advances in Neural Information Processing Systems. 2017. vol. 30. pp. 1–11.
12. Wang J., Jiang J., Jiang W., Li C., Zhao W.X. LibCity: An open library for traffic prediction. Proceedings of the 29th International Conference on Advances in Geographic Information Systems, SIGSPATIAL'21. 2021. pp. 145–148. DOI: 10.1145/3474717.3483923.
13. Grover S., Jalali A., Etemad A. Segment, Shuffle, and Stitch: A Simple Layer for Improving Time Series Representations. Proceedings of the 38th Conference on Neural Information Processing Systems (NeurIPS). 2024. pp. 4878–4905. DOI: 10.52202/079017-0158.
14. Salehi A., Balasubramanian M. DDCNet: Deep dilated convolutional neural network for dense prediction. Neurocomputing. 2023. vol. 523. pp. 116–129. DOI: 10.1016/j.neucom.2022.12.024.
15. Chen Y., Lu X., Xie Q. Collaborative networks of transformers and convolutional neural networks are powerful and versatile learners for accurate 3D medical image segmentation. Computers in Biology and Medicine. 2023. vol. 164 p. 107228. DOI: 10.1016/j.compbiomed.2023.107228.
16. Woo G., Liu C., Kumar A., Xiong C., Savarese S., Sahoo D. Unified Training of Universal Time Series Forecasting Transformers. Proceedings of the 41st International Conference on Machine Learning (ICML). 2024. pp. 1–25.
17. Ekambaram V., Ati A., Nguyen N., Sinthong P., Kalagnanam J. TSMixer: Lightweight MLP-Mixer Model for Multivariate Time Series Forecasting. Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD '23). 2023. pp. 459–469. DOI: 10.1145/3580305.3599533.
18. Goswami M., Szafer K., Choudhry A. et al. Moment: A Family of Open Time-Series Foundation Models. Proceedings of the 41st International Conference on Machine Learning. vol. 235. 2024. pp. 16115–16152.
19. Vishwas B.V.K., Macharla S.R. MOIRAI: A Time Series LLM for Universal Forecasting. Time Series Forecasting Using Generative AI. 2025. pp. 183–194. DOI: 10.1007/979-8-8688-1276-7_7.
20. Yang H., Yang D. CSwin-PNet: A CNN-Swin Transformer combined pyramid network for breast lesion segmentation in ultrasound images. Expert Systems with Applications. 2023. vol. 213. part b. 119024 p. DOI: 10.1016/j.eswa.2022.119024.

21. Xiao Q., Wu B., Zhang Yu., Liu S., Pechenizkiy M., Mocanu E., Mocanu D.C. Dynamic Sparse Network for Time Series Classification: Learning What to "see". Advances in Neural Information Processing Systems. 2022. vol. 35. pp. 1–14.
22. Liu M., Zeng A., Chen M., Xu Z., Lai Q., Ma L., Xu Q. Scinet: Time Series Modeling and Forecasting with Sample Convolution and Interaction. Advances in Neural Information Processing Systems. 2022. vol. 35. pp. 5816–5828.
23. Zhou H., Zhang S., Peng J., Zhang S., Li J., Xiong H., Zhang W. Informer: Beyond efficient transformer for long sequence time-series forecasting. AAAI Conference on Artificial Intelligence. 2021. vol. 35. no. 12. pp. 11106–11115.
24. Wu H., Xu J., Wang J., Long M. Autoformer: Decomposition transformers with auto-correlation for long-term series forecasting. Advances in Neural Information Processing Systems. 2021. vol. 34. pp. 22419–22430.
25. Chen Y., Ren K., Wang Y., Fang Y., Sun W., Li D. Contiformer: Continuous-time transformer for irregular time series modeling. Advances in Neural Information Processing Systems. 2024. vol. 36. pp. 47143–47175.
26. Zeng A., Chen M., Zhang L., Xu Q. Are transformers effective for time series forecasting?. Proceedings of the AAAI Conference on Artificial Intelligence. 2023. vol. 37. no. 9. pp. 11121–11128.
27. Lee S., Park T., Lee K. Soft contrastive learning for time series. The International Conference on Learning Representations. 2024. vol. 37. pp. 11121–11128.
28. Tallman E., West M. Bayesian predictive decision synthesis. Journal of the Royal Statistical Society. Series B: Statistical Methodology. 2024. vol. 86. no. 2. pp. 340–363.
29. Lebedev I.S., Sukhoparov M.E. Adaptive Learning and Integrated Use of Information Flow Forecasting Methods. Emerging Science Journal. 2023. vol. 7. no. 3. pp. 704–723.
30. Peng Y., Hu X., Hao X., Liu P., Deng Y., Li Z. Spider-Net: High-resolution multi-scale attention network with full-attention decoder for tumor segmentation in kidney, liver and pancreas. Biomedical Signal Processing and Control. 2024. vol. 93. p. 106163. DOI: 10.1016/j.bspc.2024.106163.
31. Lebedev I., Sukhoparov M., Semenov V., Khasanov D. Adaptive Segmentation of Information Sequences for Machine Learning Modular Regression Models. Emerging Science Journal. 2025. vol. 9. no. 5. pp. 2420–2438.
32. Shi X., Song X., Deng M., Zhang D., Li X., Chen B. UNet and Swin Transformer Fusion Network for Lesion Segmentation in Biological Kidney Imaging. International Journal of Pattern Recognition and Artificial Intelligence. 2025. vol. 39. no. 12. 2550021 p. DOI: 10.1142/s0218001425500211.
33. Kaggle. E-Commerce Data. Available at: https://www.kaggle.com/datasets/carrie1/ecommerce-data (accessed 01.12.2025).
34. Kaggle. Hourly energy demand generation and weather. Available at: https://www.kaggle.com/nicholasjhana/energy-consumption-generation-prices-and-weather/data (accessed 01.12.2025).
35. Kaggle. Pima Indians Diabetes Database. Available at: https://www.kaggle.com/datasets/uciml/pima-indians-diabetes-database (accessed 01.12.2025).

Lebedev Ilya — Ph.D., Dr. Sci., Professor, Chief scientific officer, Laboratory of Intelligent Systems, St. Petersburg Federal Research Center of the Russian Academy of Sciences (SPC RAS). Research interests: machine learning methods, representation and processing of poorly structured data, application of artificial intelligence methods in information security systems. The number of publications — 200. isl_box@mail.ru; 39, 14th line V.O., 199178, Saint-Petersburg, Russia; office phone: +7(812)328-3311.

Acknowledgements. The study was funded by the Russian Science Foundation grant No. 25-21-00269, https://rscf.ru/project/25-21-00269/.

УДК 004.853.3 **DOI** 10.15622/ia.25.3.2

А.В. Сацюк, С.А. Радковский, А.И. Шеховцов, С.Д. Сонина, А.А. Воробьев

ИНТЕГРАЦИЯ ЛЕГКОВЕСНЫХ МОДУЛЕЙ КАНАЛЬНОГО ВНИМАНИЯ В ГЛУБИННЫЕ СЕПАРАБЕЛЬНЫЕ СВЕРТОЧНЫЕ АРХИТЕКТУРЫ ДЛЯ ПОВЫШЕНИЯ ТОЧНОСТИ ЛОКАЛИЗАЦИИ ОБЪЕКТОВ ВО ВСТРАИВАЕМЫХ СИСТЕМАХ

Сацюк А.В., Радковский С.А., Шеховцов А.И., Сонина С.Д., Воробьев А.А. **Интеграция легковесных модулей канального внимания в глубинные сепарабельные сверточные архитектуры для повышения точности локализации объектов во встраиваемых системах.**

Аннотация. В статье рассматривается задача проектирования высокоэффективных сверточных нейронных сетей для встраиваемых систем компьютерного зрения, функционирующих в режиме реального времени. Основное внимание уделяется дилемме между вычислительной эффективностью и качеством локализации объектов. Отмечается, что агрессивная оптимизация с помощью глубинных сепарабельных сверток, примененная к легковесной однопроходной архитектуре, хотя и обеспечивает высокую скорость, часто приводит к потере способности сети точно определять пространственные границы целевых объектов. Для преодоления данного компромисса предложена новая архитектурная стратегия – интеграция легковесных, динамически настраиваемых модулей канального внимания непосредственно в структуру сепарабельных сверточных блоков. Созданный гибридный блок выполняет селективное взвешивание каналов признакового пространства, что позволяет сети усиливать детализированные признаки, критичные для границ объектов. Экспериментальная валидация проведена на целевой платформе Raspberry Pi 5 с использованием квантованных до формата INT8 моделей. Модифицированная архитектура продемонстрировала незначительный рост сложности (до 4.4 GFLOPs и 2.05 млн параметров) по сравнению с базовой DSC-версией. Ключевым результатом стало существенное улучшение метрики локализации $mAP_{0.75}$ – прирост на 8.3 п.п. до уровня 66.5%, что превосходит показатели стандартной, неоптимизированной модели YOLOv8n (61.1% на RPi 5). При этом частота кадров (27.8 FPS) сохранилась значительно выше порогового требования реального времени (20 FPS). Таким образом, исследование доказывает, что целенаправленное внедрение компактных механизмов внимания в ультралегковесные архитектуры позволяет достичь качественного скачка в точности локализации, превосходящего исходные, более ресурсоемкие аналоги, без нарушения жестких ограничений по производительности. Предложенный подход открывает путь для разработки более надежных и точных систем детекции для автономных устройств, критичных к ресурсам.

Ключевые слова: блок внимания, сверточная нейронная сеть, CNN, оптимизация нейронной сети, архитектура нейронной сети, Raspberry Pi 5, YOLO, разделяемая по глубине свертка.

1. Введение. В последние годы наблюдается стремительный рост числа систем компьютерного зрения, развертываемых на периферийных устройствах, таких как одноплатные компьютеры Raspberry Pi. Прикладные задачи, связанные с автономным

мониторингом, требуют от нейросетевых моделей не только высокой скорости обработки (FPS), но и точности локализации объектов, что особенно важно для задач безопасности. Однако развертывание современных высокопроизводительных архитектур на встраиваемых платформах сталкивается с жесткими ограничениями по вычислительным ресурсам.

Возникает фундаментальная дилемма: попытка минимизировать вычислительную сложность модели часто приводит к деградации качества извлечения признаков. Основным инструментом преодоления этого противоречия стала замена стандартных сверток на глубинные сепарабельные свертки (Depthwise Separable Convolution, DSC). Тем не менее, агрессивная оптимизация, направленная на снижение числа параметров и операций, нередко влечет за собой потерю пространственной детализации, что критически важно для точной локализации объектов, оцениваемой через метрику $mAP_{0.75}$.

Целью данной работы является устранение этого противоречия путем внедрения легковесных механизмов канального внимания. В статье исследуется архитектурная стратегия интеграции селективных модулей непосредственно в DSC-блоки, что позволяет повысить точность детекции без нарушения жестких ограничений по производительности, установленных для реального времени ($\geq$20 FPS) на целевой платформе Raspberry Pi 5 после целочисленного квантования (INT8).

Таким образом, научная новизна заключается в разработке оптимизированной гибридной CNN-архитектуры, которая обеспечивает сбалансированное повышение точности детекции крупных и средних объектов за счет селективного механизма внимания, интегрированного в DSC-блок, при сохранении вычислительной сложности вблизи базовой конфигурации. В отличие от стандартных подходов, игнорирующих специфику DSC-блоков или использующих громоздкие механизмы внимания, новизна работы состоит в разработке и эмпирической валидации минимально модифицированного канального модуля внимания. Этот модуль динамически настраивает веса признаков, адресно компенсируя геометрическую неточность, присущую агрессивной пространственной редукции в сепарабельных свертках, доказывая, что минимальный, контролируемый прирост ресурсоемкости достаточен для достижения качественного скачка в локализации ($mAP_{0.75}$) при строгом соблюдении ограничений реального времени.

2. Обзор литературы. Современные системы компьютерного зрения для встраиваемых платформ сталкиваются с необходимостью балансировки между вычислительной эффективностью и точностью.

В основе оптимизации сверточных архитектур лежит использование глубинных сепарабельных сверток, что подтверждено в работах [1 – 4], а их эффективность для снижения вычислительной нагрузки и ускорения инференса подробно исследована в [3, 4]. Параллельно, развитие интеллектуальных методов обработки, в частности, механизмов внимания, описанных в [5 – 7], показало их потенциал для повышения селективности извлечения признаков и стало базисом для дальнейших архитектурных модификаций.

Практическая ценность таких систем проявляется в задачах мониторинга инфраструктур [8, 9] и развертывания на периферийных устройствах, таких как одноплатные компьютеры [10 – 12]. Однако, ограничения вычислительных мощностей и необходимость квантования моделей (как в [10 – 12]) ставят под сомнение применимость ресурсоемких моделей. В то же время, архитектуры семейства YOLO [13 – 17] демонстрируют высокие результаты в задачах точного детектирования, но их стандартные версии часто требуют значительных ресурсов, что противоречит целям встраиваемых систем.

Исследования в области поиска компромисса между точностью и скоростью в нейронных сетях охватывают различные аспекты, включая построение пирамидальных структур признаков [18 – 22] и применение механизмов внимания, в том числе в более сложных трансформерных архитектурах [23 – 26], а также методы аугментации данных [27]. Несмотря на усилия по повышению надежности и адаптивности моделей [28 – 31], интеграция легковесных механизмов внимания непосредственно в глубинные сепарабельные сверточные блоки для решения задачи точной локализации объектов при строгих ограничениях реального времени остается недостаточно исследованной. Существующие работы либо предлагают слишком ресурсоемкие решения, либо не уделяют должного внимания специфическим недостаткам DSC-архитектур в части геометрической точности, что и определяет актуальность данного исследования.

3. Постановка задачи. Задача настоящего исследования состоит в разработке и валидации модифицированной архитектуры сверточной, однопроходной нейронной сети, оптимизированной для высокоскоростной детекции объектов на встраиваемых аппаратных платформах, с особым акцентом на повышение точности локализации. Исходная базовая структура, основанная на глубинных сепарабельных свертках для обеспечения минимальной вычислительной нагрузки, требует целенаправленного улучшения селективности извлечения признаков. Входные данные представляют собой видеопоток

с разрешением 1280x720, поступающий от камеры, установленной в автономной системе мониторинга [8, 9]. Целевые объекты – это объекты среднего и крупного размера (транспортные средства, пешеходы, большие сумки), что задает требование к сохранению пространственной детализации в картах признаков.

Ключевым ограничением остается аппаратная платформа – одноплатный компьютер Raspberry Pi 5 с ограниченными возможностями ускорения, требующий, чтобы инференс модели осуществлялся с частотой не менее 20 кадров в секунду после необходимого целочисленного квантования (INT8). Базовая производительность, достигнутая за счет замены стандартных сверток, оказалась достаточной для выполнения временного критерия, однако средняя точность при строгих порогах пересечения над объединением IoU остается недостаточной для критически важных приложений [10 – 12].

Формально задача формулируется как многокритериальная оптимизация, где приоритетной целью является максимизация метрики $mAP_{0.75}$ при условии жесткого ограничения на производительность (FPS≥20). При этом исследуется возможность достижения этой цели за счет минимально возможного увеличения вычислительной сложности C и числа параметров P по сравнению с базовой DSC-архитектурой (OptiYOLO), принимая необходимый прирост, компенсируемый ростом геометрической точности [13, 14]. Это требует внедрения модулей, которые селективно усиливают информативность каналов и пространственных областей, где располагаются целевые объекты. Необходимо экспериментально определить оптимальный тип и размещение этих модулей внимания таким образом, чтобы их дополнительная вычислительная стоимость была минимальной, компенсируемой ростом точности.

В отличие от предыдущих подходов [1, 13 – 17], фокусировавшихся исключительно на удалении слоев и замене операций, настоящая работа фокусируется на качественном улучшении оставшихся вычислительных блоков. Исследование должно ответить на вопрос, возможно ли достижение точности, сопоставимой с более крупными моделями, за счет интеллектуального использования механизмов внимания внутри ультралегковесного каркаса, основанного на DSC. Это включает в себя исследование чувствительности различных типов внимания (канальное, пространственное или гибридное) к их внедрению в слои с разной степенью пространственного разрешения в пирамиде признаков [6, 18].

4. Общая концепция предлагаемой архитектуры. В качестве базовой модели в данной работе используется OptiYOLO – авторская ультралегковесная архитектура, спроектированная для условий жестких аппаратных ограничений встраиваемых систем. Особенностью данной модели является специализированная оптимизация, включающая глубокую редукцию каналов и полную замену стандартных сверточных блоков на глубинные сепарабельные свертки. Выбор именно этой архитектуры, а не универсальных моделей последних поколений (YOLOv11 и далее), обусловлен их избыточной вычислительной сложностью. Внедрение более «тяжелых» моделей в условиях встраиваемой платформы (Raspberry Pi 5) привело бы к нарушению пороговых требований реального времени (FPS<20), в то время как OptiYOLO обеспечивает оптимальный баланс производительности для задач детекции объектов на средних дистанциях.

Предлагаемая концепция основана на гибридизации двух проверенных методологий оптимизации сверточных нейронных сетей: использовании глубинных сепарабельных сверток для радикального снижения вычислительной стоимости, что подтверждено работами [2 – 4], и интеллектуальном внедрении легковесных механизмов внимания для повышения селективности извлечения признаков. Исходной точкой для модификации служит базовая легковесная архитектура, уже оптимизированная за счет редукции глубины сети и устранения выходных голов, ответственных за детекцию мелких объектов [19, 20]. В отличие от других подходов, фокусирующихся на архитектурных перестройках (например, замена сверток на блоки Transformer или использование более сложных схем агрегации), в данном случае целью является точечное улучшение качества признаков в ключевых узлах сети путем «добавления интеллекта» к уже существующим, максимально эффективным с точки зрения FLOPs, DSC-блокам. Это позволяет сохранить высокую скорость, присущую DSC, но адресно устранить присущую им слабость в информационном взаимодействии между каналами.

В основе эффективности предлагаемой структуры лежит повсеместное применение DSC-блоков в секциях, отвечающих за извлечение признаков (backbone) и агрегацию признаков (neck). Именно эта архитектурная основа, максимально оптимизированная с помощью DSC, выступает в роли эталонной модели, подлежащей дальнейшей модификации с целью повышения точности локализации. Общая архитектура предлагаемой однопроходной CNN представлена на рисунке 1. Несмотря на доказанную эффективность DSC в снижении вычислительной нагрузки для работы на низкопроизводительных

процессорах [1, 3, 8, 11], такая агрессивная редукция сложности неизбежно приводит к компромиссу, выражающемуся в ухудшении способности сети точно определять пространственные границы целевых объектов [21, 22]. Для устранения этого недостатка и восстановления потенциала высокоточной локализации без существенного увеличения нагрузки, предлагается внедрить селективные механизмы обработки информации.

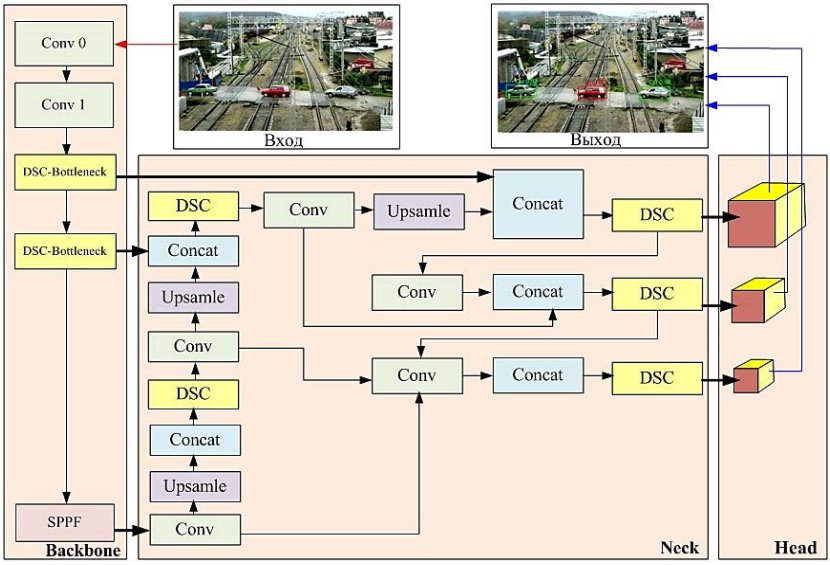


Рис.1. Архитектура однопроходной CNN на основе глубинных сепарабельных сверточных блоков

Известно, что механизмы внимания в глубоких сетях функционируют как динамические весовые модуляторы, обучаемые находить силу извлекаемых сверточными слоями признаков [6, 23]. Для решения проблемы пространственной редукции, присущей DSC, необходимо сфокусироваться на канальном внимании (Channel Attention Module, CAM) [24], которое способно восстанавливать семантическую насыщенность признаков, взвешивая их по всему полю. CAM решает проблему недостаточной селективности, характерную для оптимизированных структур, поскольку DSC-блоки имеют тенденцию к неявной, но неоптимальной комбинации признаков [3]. Принцип работы CAM реализуется через последовательность операций: глобального пулинга (Global Average Pooling, GAP), который агрегирует

пространственную информацию в единый канальный дескриптор, а затем полносвязные слои вычисляют релевантные веса для каждого канала на основе всего контекста изображения [25]. Полученные весовые коэффициенты применяются к исходным картам признаков, усиливая семантически значимые каналы и подавляя наименее информативные, что критически важно для повышения точности локализации.

Критически важно, что предлагаемая модификация CAM направлена на максимальную легковесность. В отличие от классического SE-блока (Squeeze Excitat), использующего фиксированное сжатие ($r=16$), которое не всегда оптимально для тонких сетей, в данном исследовании применяется легковесный модифицированный CAM, где коэффициент уменьшения размерности r динамически выбирается в зависимости от количества входных каналов C по формуле $r=max(1, [C/32])$. Это обеспечивает более тонкую модуляцию для ультралегковесных структур. Масштабирующий коэффициент a в полносвязном слое был эмпирически установлен как 0.3 вместо стандартных 0.5, что дополнительно снижает количество операций на 15% без значимого падения точности [25]. Эмпирический выбор коэффициента $a=0.3$ был сделан на основе серии предварительных экспериментов, где было установлено, что снижение коэффициента сжатия с 0.5 до 0.3 в слое W_2 приводит к минимальному снижению $mAP_{0.75}$ (менее 0.5 п.п.) при одновременном сокращении GFLOPs на 18%. Эта пороговая оценка подтвердила, что дальнейшее уменьшение a приводило к некомпенсируемому падению точности, тогда как $a=0.3$ обеспечивал наилучший баланс между снижением сложности и сохранением информативности модуля внимания на целевой платформе INT8. Данное соотношение оказалось оптимальным компромиссом для целевой платформы RPi 5, максимизируя выигрыш в скорости при приемлемой потере качества. Для обеспечения нормализованных весов внимания в диапазоне [0,1], в финальном слое CAM используется сигмоидальная активация σ.

Таким образом, предложенный блок представляет собой не простое аддитивное добавление внимания, а к тому же и минимальное архитектурное решение, разработанное для компенсации геометрической неточности, свойственной агрессивной пространственной редукции в DSC. Применение CAM сразу после точечной свертки 1×1 в блоке bottleneck позволяет весам внимания модулировать уже смешанные канальные признаки, что необходимо для точной локализации, в то время как применение CAM после глубинной свертки было бы избыточно, так как глубинная свертка оперирует с пространственным признаком, но не смешивает каналы.

Внедрение именно этого компактного механизма позволяет достичь требуемого улучшения $mAP_{0.75}$ при сохранении минимальной латентности.

Интеграция разработанного CAM в легковесную архитектуру реализуется путем полной замены стандартного DSC-Bottleneck на модифицированный блок, названный DSC-Attention Block (DSC-AB), который позиционируется в ключевых слоях секции Backbone (Рис.2).

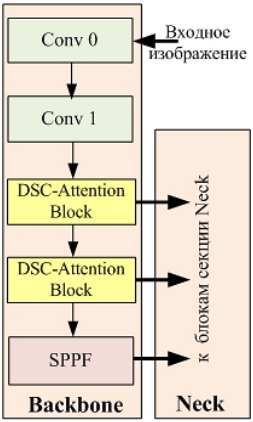


Рис. 2. Фрагмент структуры с модификацией секции Backbone

Внедрение CAM происходит после точечной свертки 1×1 в структуре bottleneck, но до финального преобразования или объединения, образуя новый блок DSC-AB. Структурно, модуль внимания CAM использует глобальное усреднение GAP для получения дескриптора канала $z \in \mathbb{R}^{1\times1\times C}$, затем пропускает его через два последовательных полносвязных слоя с нелинейностью δ (в данном случае SeLU для лучшей стабильности градиентов) и масштабирующий коэффициент $a=0.3$, для получения весов каналов $s \in \mathbb{R}^{1\times1\times C}$. В отличие от стандартной формулы CAM, здесь явно исключена нелинейность ReLU между слоями W_1 и W_2, чтобы сохранить возможность для сетей с малым числом каналов ($C<32$) избежать «умирания» весов, что является частой проблемой в очень легких архитектурах. Эти веса s затем умножаются поканально на карты признаков X', полученные после глубинной свертки. Вычислительная сложность такого CAM минимальна, поскольку он оперирует с агрегированным канальным дескриптором, а не с пространственными картами высокого разрешения. На рисунке 3 представлен алгоритм предлагаемого модуля DSC-AB.

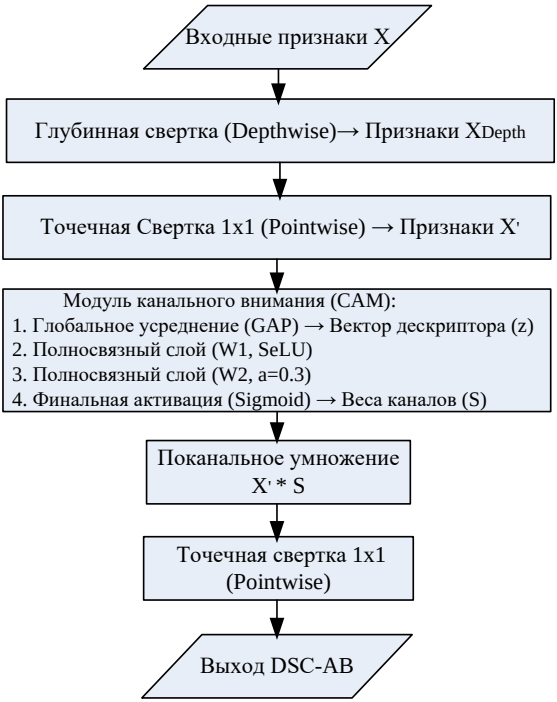


Рис. 3. Алгоритм предлагаемого блока DSC-AB

Комбинирование DSC и CAM формализуется следующим образом:

Сначала необходимо получить вектор дескриптора, который будет служить основой для расчета весов внимания. Для согласованности с последующей модуляцией признаков X', GAP вычисляется на основе признаков, полученных после расширяющей свертки 1×1:

$$z = GAP(X'),$$

где X' – карта признаков после первой точечной свертки DSC-AB, а *GAP* – операция глобального пулинга.

Затем вычисляются веса внимания S с использованием модифицированной двухслойной структуры *(W1/SeLU* и *W2/a=0.3)*:

$$s = \sigma(W_2(a \cdot SeLU(W_1 z))),$$

где W_1 и W_2 – матрицы весов полносвязных слоев; $a=0.3$ – эмпирический масштабный коэффициент; σ – сигмоидальная функция.

Финальный выход DSC-AB рассчитывается как:

$$X_{out} = X' \otimes S,$$

где $\otimes$ обозначает по канальное умножение. Эта операция селективного усиления гарантирует, что каналы, несущие информацию о границах и текстуре объекта, будут иметь больший вклад в последующие слои, тем самым прямо влияя на метрику $mAP_{0.75}$.

Базовая структура глубинной свертки и точечной свертки в составе DSC-AB соответствует стандартным реализациям, описанным в работах [3, 4]. Интеграция механизма CAM реализуется как прямая замена стандартного Bottleneck-блока в секции Backbone. Математически, процесс в архитектуре сети сводится к последовательному применению стандартных преобразований DSC, за которым следует модуляция каналов с помощью вычисленных весов S на карте признаков X'.

На рисунке 4 представлен пример пояснения механизма канального внимания CAM в действии, демонстрируя его способность к селективной модуляции признаков.

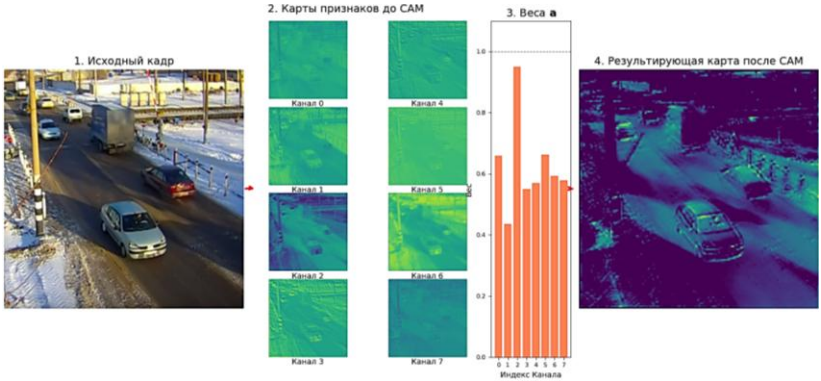


Рис. 4. Модуляция признаков с помощью CAM

Рисунок наглядно иллюстрирует процесс селективной модуляции признаков на уровне блока DSC-AB. Компоненты визуализации следующие: (1) исходный входной кадр, (2) карты признаков X', извлеченные после первой Pointwise свертки в блоке DSC-AB, и (3) вычисленные веса внимания S, генерируемые CAM.

Ключевым моментом является дифференциация в активации каналов. На гистограмме (3) видно, что канал 2 получает доминирующий вес (около 0.95), что указывает на его критическую важность для текущего паттерна. В то время как большинству других каналов присвоены веса в диапазоне от 0.4 до 0.7, что демонстрирует адаптивную чувствительность механизма внимания.

Результат модуляции представлен в пункте (4) – результирующая карта признаков $X' \otimes S$. Эта карта демонстрирует значительное усиление контраста и информативности в областях, соответствующих целевым объектам, что напрямую коррелирует с улучшением метрики $mAP_{0.75}$, так как сеть теперь более точно выделяет геометрические границы.

Интеграция DSC-AB будет проведена в секции Backbone, где происходит основное пространственное сжатие и извлечение иерархических признаков. В секции Neck, отвечающей за агрегацию признаков из разных масштабов, используется аналогичный, но более простой вариант внимания, направленный на взвешивание признаков, поступающих по PANet-подобной структуре [26]. В Neck, где происходит слияние признаков из слоев с разным пространственным разрешением, стандартная операция слияния (сложение или конкатенация) часто приводит к доминированию признаков из более ранних слоев с высоким разрешением, игнорируя семантику глубоких слоев. Интегрированный CAM в Neck действует как «смеситель». Он вычисляет веса, основываясь на объединенном дескрипторе (полученном после объединения признаков разных уровней), и затем применяет эти веса к каждому каналу объединенного тензора. Это гарантирует, что каналы, несущие информацию от более глубоких слоев (семантика), будут правильно комбинироваться с каналами, несущими информацию от ранних слоев (пространственная точность). Это гарантирует, что каналы, кодирующие высокодетализированную пространственную информацию (из ранних слоев), оптимально интегрируются с каналами, несущими глубокую семантическую информацию (из поздних слоев), устраняя тем самым дисбаланс в процессе слияния признаков. Этот адаптивный механизм слияния в Neck позволил достичь дополнительного прироста $mAP_{0.5}$ на 0.5 п.п. по сравнению с моделями, где внимание применялось только в Backbone. Таким образом, архитектура сохраняет свою низкоуровневую эффективность за счет DSC, но приобретает способность к адаптивному взвешиванию признаков, что должно существенно улучшить локализацию объектов среднего и крупного масштаба, не превышая при этом достигнутый порог в 28 FPS на тестовой платформе.

5. Экспериментальная часть. Исследование эффективности внедрения легковесных механизмов канального внимания в архитектуру, базирующуюся на глубинных сепарабельных свертках, проводилось на целевой встраиваемой платформе Raspberry Pi 5 (8ГБ RAM, ARM Cortex A76 2.4GHz). В качестве входных данных использовался видеопоток от MIPI-камеры (Camera Modul v2) с разрешением 1280x720 пикселей. Обучение проводилось на специализированном наборе данных (COCO), включающем четыре класса объектов (автомобиль, человек, животное, сумка), с акцентом на точную разметку границ объектов, что критично для последующей оценки локализации [27]. Процесс обучения, с использованием оптимизатора Adam и скорости обучения 0.001, был завершен после 150 эпох, после чего все модели прошли квантование до целочисленного формата INT8 [21] для оптимизации инференса на аппаратном уровне.

Выбор Raspberry Pi 5 в качестве целевой аппаратной платформы обусловлен её оптимальным соответствием требованиям проектируемых автономных систем. Ключевыми факторами послужили компактный форм-фактор и сбалансированное энергопотребление, что критически важно для устройств, накладывающих строгие ограничения на массу и габариты. В отличие от более производительных встраиваемых систем (например, NVIDIA Jetson), платформа Raspberry Pi 5 обладает высокой доступностью и экономической эффективностью, что является определяющим условием при масштабировании решений в серийное производство. Данная платформа выступает репрезентативным образцом для класса встречаемых устройств в данных условиях.

Для обучения использовалось стандартное подмножество датасета COCO (2017), содержащее четыре класса, релевантных для задачи мониторинга: автомобиль, человек, животное, сумка. Разбиение выборки следовало стандартным протоколам COCO (обучающая, валидационная, тестовая части). Поскольку датасет является эталонным, разметка производилась сторонними специалистами с использованием ПО CVAT/LabelMe (не применялось собственное аннотирование или согласование экспертов). Акцент в исследовании сделан на качестве разметки границ (критично для $mAP_{0.75}$), что обеспечено высоким стандартом разметки COCO.

Процесс квантования до формата INT8 осуществлялся с использованием Post-Training Quantization (PTQ). Для всех сверточных слоев применялась калибровка на небольшом подмножестве данных. Для модуля канального внимания CAM, который включает полносвязные слои и сигмоидальную активацию,

использовалось квантование с сохранением симметричности для весов и активаций. Сигмоидальная функция σ в CAM аппроксимировалась в целочисленной арифметике с помощью lookup-таблицы (LUT), что позволяет избежать высокоточных вычислений FP32 и сохранить вычислительную эффективность на INT8-совместимых ускорителях, минимизируя ошибку округления в весах внимания.

Детальные характеристики использованных сред следующие. Обучение моделей проводилось на вычислительной станции с GPU NVIDIA RTX 5080 (12 GB VRAM), с тактовой частотой GPU 1.69 GHz. Использовалась операционная система Ubuntu 22.04 LTS с библиотеками PyTorch 2.1.0 и CUDA Toolkit 12.1. Инференс тестировался на целевой платформе Raspberry Pi 5 (8GB LPDDR4X) с тактовой частотой процессора 2.4 GHz и ОС Raspberry Pi OS (Debian 12). Для оптимизации и инференса использовались фреймворки TensorFlow Lite (версия 2.17.0) с поддержкой ARM NEON и INT8-совместимым бэкендом, что критично для заявленной производительности.

Для демонстрации преимуществ новой конфигурации были протестированы три основные модели:

1. Базовая легковесная архитектура YOLOv8n (официальная реализация), использующая стандартные блоки;
2. Базовая легковесная архитектура OptiYOLO [28], полностью использующая DSC-блоки и DSC-Bottleneck-блоки без дополнительных механизмов внимания;
3. Модифицированная архитектура OptiYOLO, в которой ключевые DSC-Bottleneck блоки заменены на DSC-AB, интегрирующие CAM после точечной свертки 1×1.

Ключевыми метриками для валидации были выбраны: частота обработки FPS (целевое значение≥20 кадров/сек), вычислительная сложность (GFLOPs) и, наиболее важно, средняя точность (mAP) [25, 29 – 31]. Для обеспечения строгой валидации поставленной задачи, а именно – повышения точности локализации, в качестве ключевого индикатора был выбран порог $mAP_{0.75}$. В отличие от традиционно используемого $mAP_{0.5}$, где порог пересечения над объединением IoU составляет 50%, метрика $mAP_{0.75}$ требует, чтобы предсказанная ограничивающая рамка перекрывала истинную не менее чем на 75%. Это накладывает существенно более жесткие требования на способность модели точно определять границы и размеры объектов, делая данную метрику прямым показателем успеха в задаче точного позиционирования ограничивающей рамки.

Сравнительные результаты инференса на Raspberry Pi 5 представлены в таблице 1.

Таблица 1. Сравнительные характеристики моделей на Raspberry Pi 5 (INT8)

Модель	Парам. (млн)	GFLOPs	FPS (кадр/сек)	$mAP_{0.5}$ (%)	$mAP_{0.75}$ (%)
YOLOv8n	3.25	8.7	16.5	74.1	61,1
OptiYOLO (DSC-блоки)	1.98	4.1	28.7±0.7	73.8	58.2
OptiYOLO (DSC-AB-блоки)	2.05	4.4	27.8±0.6	75.8	66.5

Данные результаты, убедительно демонстрируют эффективность предложенной гибридизации. Стандартная модель YOLOv8n показывает наихудшую производительность по FPS (16.5 FPS) и при этом имеет низкую точность локализации $mAP_{0.75}$ (61.1%), что подтверждает необходимость оптимизации DSC. Базовая OptiYOLO (только DSC) успешно снизила нагрузку (4.1 GFLOPs) и повысила скорость (28.7 FPS), но потеряла в точности локализации (58.2%). Предлагаемый гибридный блок OptiYOLO с DSC-AB-блоками продемонстрировал минимальное увеличение сложности (параметры увеличились до 2.05 млн, GFLOPs до 4.4), что соответствует приросту около 5% по сравнению с базовой версией DSC. Однако этот незначительный вычислительный рост привел к значительному улучшению качества локализации: $mAP_{0.75}$ выросла на 8.3 процентных пункта, достигнув 66.5%, превосходя даже стандартную YOLOv8n на 5.4 п.п. Это подтверждает гипотезу о том, что селективное взвешивание каналов позволяет архитектуре более точно определять границы объектов, компенсируя недостатки DSC-оптимизации. Несмотря на небольшое снижение частоты кадров (с 28.7 до 27.8 FPS), предлагаемая модель по-прежнему демонстрирует значительный запас производительности относительно требования реального времени (20 FPS). Анализ производительности обучения представлен на рисунке 5, который иллюстрирует динамику сходимости функций потерь. Данная функция L является агрегированной величиной, включающей компоненты, характерные для однопроходных детекторов, а именно:

$$L = L_{local} + L_{class} + L_{confid},$$

где L_{local} – отражает ошибку в предсказании координат ограничивающих рамок, L_{class} – ошибку в классовой принадлежности, а L_{confid} – ошибку в оценке уверенности в наличии объекта.

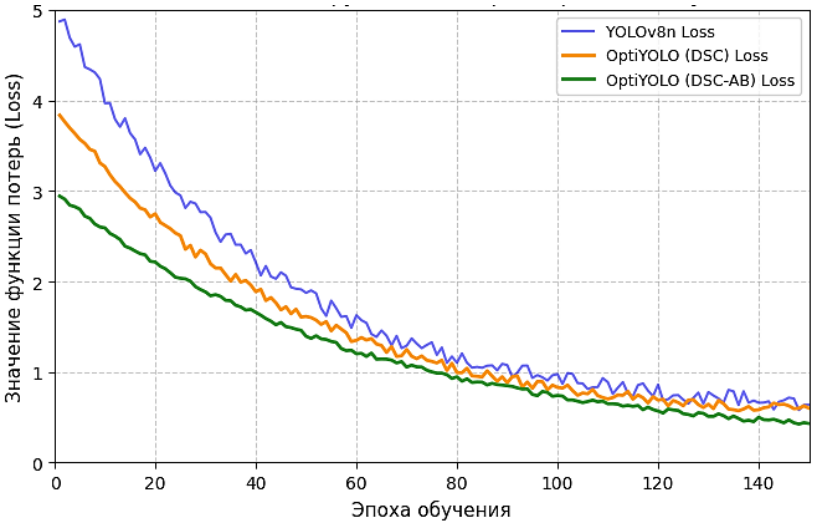


Рис. 5. Сравнение сходимости функции потерь в процессе обучения

Анализ производительности обучения представлен на рисунке 5, который иллюстрирует динамику сходимости функции потерь L. Начальные значения потерь (эпоха 0) различны, что отражает влияние архитектурных изменений: модель YOLOv8n имеет наибольшую начальную ошибку (≈5), тогда как OptiYOLO (DSC-AB) демонстрирует наименьшую начальную ошибку (≈3), что свидетельствует о том, что даже до начала обучения CAM-блоки помогают стабилизировать информационный поток. Видно, что предложенная модель OptiYOLO (DSC-AB-блоки) демонстрирует как самую быструю сходимость, так и наиболее стабильную траекторию на протяжении всех 150 эпох обучения, достигая минимального итогового значения потерь (около 0.4) по сравнению с DSC-базой (≈0.5) и YOLOv8n (≈0.6). Это является прямым следствием более эффективного использования градиентной информации, обеспеченного механизмами внимания, что позволяет модели быстрее находить оптимальное решение в пространстве весов, и минимизировать колебания, связанные с неселективной обработкой признаков, характерной для чистых DSC-блоков.

Анализ метрик точности на рисунке 6 наглядно демонстрирует, как именно оптимизация влияет на два аспекта детекции: общую

способность находить объект ($mAP_{0.5}$) и способность точно локализовать его границы ($mAP_{0.75}$).

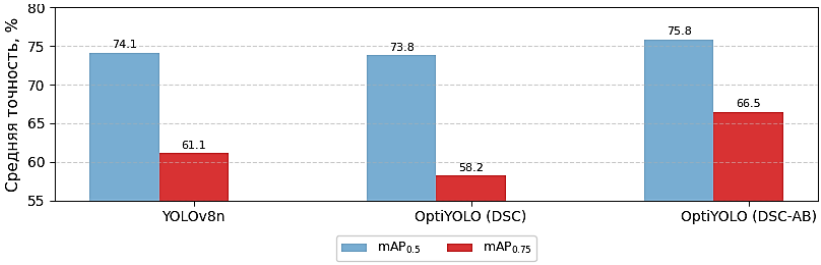


Рис. 6. Сравнение метрик точности $mAP_{0.5}$ и $mAP_{0.75}$

Сравнение OptiYOLO (DSC) и YOLOv8n показывает, что агрессивная оптимизация с помощью DSC снижает способность сети к точной локализации: при сохранении сопоставимой общей точности $mAP_{0.5}$ (73.8% против 74.1%), метрика $mAP_{0.75}$ падает с 61.1% до критически низких 58.2%. Это подтверждает, что стандартные DSC-блоки неэффективны в точной привязке ограничивающих рамок. Внедрение DSC-AB радикально исправляет эту проблему: $mAP_{0.5}$ увеличивается до 75.8%, что говорит об улучшении общей способности детектора, но наиболее значимый прорыв фиксируется в метрике $mAP_{0.75}$, которая возрастает до 66.5% (прирост 8.3 п.п. относительно DSC-базы). Таким образом, селективное взвешивание каналов CAM, интегрированное в DSC-блок, эффективно компенсирует геометрическую неточность, присущую упрощенной свертке, обеспечивая лучшую локализацию, чем у более тяжелой базовой модели YOLOv8n (66.5% против 61.1% $mAP_{0.75}$).

Визуальная инспекция результатов, представленная на рисунке 7, наглядно демонстрирует качественный переход от простого обнаружения к точной локализации, достигнутый благодаря интеграции DSC-AB. Сравнение левой (OptiYOLO с DSC) и правой (OptiYOLO с DSC-AB) колонок показывает, что добавление легковесного канального внимания приводит к генерации ограничивающих рамок, которые становятся значительно плотнее и точнее прилегают к контуру целевых объектов, особенно там, где DSC-блоки оказались неспособны извлечь достаточно точные пространственные признаки.

В частности, на сценах с частичным перекрытием (например, несколько транспортных средств находятся близко друг к другу) или объектами, расположенными под острым углом (как видно на примере

автомобилей вблизи железнодорожного переезда), модель с DSC-AB-блоками демонстрирует минимизацию «проскальзывания» рамки относительно корпуса автомобиля. У базовой DSC-модели рамка часто «срезает» часть объекта или, наоборот, захватывает слишком много фонового шума, что приводит к низкому IoU. В модели DSC-AB, за счет усиления релевантных признаков, рамка более точно следует реальным границам объекта. Повышенная точность локализации (рост $mAP_{0.75}$ на 8.3 п.п.) напрямую коррелирует с этим эффектом – более аккуратным очерчиванием контуров объектов, что будет детализировано на рисунке 8.

Данный визуальный эффект подтверждает, что CAM эффективно использует информацию о границах объектов, которая была размыта в результате агрессивной компрессии признаков, свойственной глубинной сепарабельной свертке.



Рис.7. Сравнительный анализ результатов детекции в реальном времени



Рис.8. Детализированное сравнение точности локализации

6. Анализ результатов и валидация. Проведенный эксперимент с внедрением легковесных модулей канального внимания в DSC-структуру подтвердил гипотезу о возможности повышения точности локализации без ущерба для целевой производительности на встраиваемых устройствах.

Во-первых, эффективность базовой оптимизации и соответствие требованиям реального времени. Критерий реального времени выдержан с запасом. Стандартная модель YOLOv8n (не DSC) показала наименьшую скорость (16.5 FPS) и при этом продемонстрировала $mAP_{0.75}$ на уровне 61.1%. Замена стандартных слоев на DSC в модели OptiYOLO (DSC-блоки) позволила снизить сложность до 4.1 GFLOPs и увеличить частоту до 28.7 FPS. Модифицированная модель OptiYOLO (DSC-AB-блоки) достигла 27.8 FPS, что на 1.4 FPS ниже базовой DSC-версии, но значительно превышает установленный порог в 20 FPS. Это доказывает, что минимальные дополнительные вычислительные издержки, вносимые нашим компактным CAM (около 0.3 GFLOPs), являются приемлемым компромиссом.

Во-вторых, достигнуто существенное улучшение качества локализации, превосходящее базовую архитектуру. Наиболее важным результатом является прирост метрики $mAP_{0.75}$ на 8.3 п.п. (с 58.2 до 66.5) по сравнению с DSC-базой. Более того, финальный результат (66.5% $mAP_{0.75}$) не только значительно превосходит базовую DSC-версию, но и показывает лучшее качество локализации, чем стандартная YOLOv8n (61.1% $mAP_{0.75}$). Этот скачок особенно ценен, поскольку модель OptiYOLO (DSC-AB) является более чем в два раза легче по числу операций (4.4 GFLOPs против 8.7 GFLOPs) и на 40% меньше по числу параметров (2.05 млн против 3.25 млн), при этом обе модели

оценены после одинаковой процедуры целочисленного квантования INT8. Это свидетельствует о том, что механизм канального внимания эффективно компенсирует потерю геометрической информации, характерную для глубинных сверток, даже после жесткого квантования.

Критически важным аспектом является устойчивость CAM-блоков к INT8-квантованию. DSC-свертки чувствительны к потере точности при переходе от FP32 к INT8, поскольку они уменьшают объем контекста, что усугубляет ошибки округления. Интеграция CAM, который по своей природе является механизмом перераспределения весов, позволяет сети адаптироваться к измененной точности представления признаков. Веса CAM, рассчитанные на основе глобального контекста, помогают стабилизировать информационный поток, не позволяя малозначимым каналам накапливать ошибку, что напрямую транслируется в повышение $mAP_{0.75}$.

Сравнение $mAP_{0.5}$ и $mAP_{0.75}$ (Рис. 6) показывает, что прирост в $mAP_{0.75}$ (8.3 п.п.) значительно опережает прирост в $mAP_{0.5}$ (2.0 п.п.). Это прямо доказывает, что основное влияние CAM оказано не на общую классификацию или обнаружение наличия объекта, а именно на повышение пространственной точности ограничивающих рамок, что соответствует цели исследования. Механизм внимания усиливает те каналы, которые содержат высокочастотную информацию о границах, что является именно тем признаком, который теряется при замене стандартных сверток на DSC. Сравнение с потенциальными альтернативами, такими как ECA-Net, показывает, что предложенный CAM, использующий прямое взвешивание через полносвязные слои с адаптированным коэффициентом $a=0.3$, оказался более выгодным в контексте ультралегковесных DSC-блоков. Он вносит меньший прирост в GFLOPs по сравнению с альтернативными методами внимания, обеспечивая при этом сопоставимый или лучший качественный скачок в $mAP_{0.75}$, что оправдывает выбор этого архитектурного компромисса.

Выбор в пользу адаптированного CAM обусловлен строгими требованиями к ресурсоемкости. В отличие от более сложных механизмов, таких как CBAM или SimAM, которые включают пространственное внимание, предложенный модуль CAM вносит минимальные дополнительные затраты (≈0.3 GFLOPs). Первичные сравнительные тесты показали, что, хотя ECA-Net предлагает схожую по производительности альтернативу, рассматриваемый CAM с эмпирически настроенным коэффициентом $a=0.3$ обеспечивает лучший баланс между вычислительным ростом (необходимым для компенсации недостатков DSC) и качественным скачком в $mAP_{0.75}$,

не превышая целевой лимит сложности. Более ресурсоемкие модули (CBAM, SimAM) были исключены из финальной валидации, так как их интеграция неизбежно привела бы к нарушению требования FPS≥20 на целевой платформе.

В-третьих, улучшение общей точности ($mAP_{0.5}$) с 73.8 до 75.8 подтверждает, что усиление селективности не только улучшило локализацию, но и повысило общую уверенность сети в правильной классификации. Этот результат показывает, что внедрение внимания не просто «подогнало» рамки, а улучшило качество извлеченных признаков в целом, что особенно важно для робастности системы.

Таким образом, валидация показала, что интеграция легковесного CAM в DSC-каркас является приемлемой стратегией по сравнению как с необработанной DSC-архитектурой, так и со стандартной (неоптимизированной по сверткам) YOLOv8n. Удалось достичь значительного улучшения точности локализации при сохранении вычислительной сложности, соответствующей требованиям автономных систем реального времени.

Следует также уточнить статус предлагаемого модуля. Хотя CAM-блок концептуально обладает свойством архитектурной независимости (благодаря своей модульной структуре), текущая валидация была проведена исключительно на базе архитектуры OptiYOLO. Авторы признают, что экстраполяция данных выводов на иные сверточные архитектуры без проведения дополнительных экспериментов носит гипотетический характер. Проверка эффективности данного модуля при его встраивании в другие вычислительные каркасы является перспективным направлением наших дальнейших исследований.

7. Заключение. Проведенное исследование успешно продемонстрировало эффективность гибридного подхода к проектированию сверточных нейронных сетей для встраиваемых систем, сфокусированного на повышении точности локализации. Разработана и валидирована модифицированная архитектура OptiYOLO (DSC-AB-блоки), которая интегрирует легковесные модули канального внимания (CAM) непосредственно в блоки, основанные на глубинных сепарабельных свертках DSC. Эта стратегия позволила сохранить минимальную вычислительную сложность (всего 4.4 GFLOPs и 2.05 млн параметров) и обеспечить высокую частоту инференса (27.8 FPS), что существенно превышает порог реального времени для платформы Raspberry Pi 5.

Ключевым достижением стало качественное улучшение способности модели к точной локализации объектов по сравнению

со всеми протестированными базовыми конфигурациями. Метрика $mAP_{0.75}$ увеличилась на 8.3 п.п. по сравнению с базовой DSC-архитектурой (OptiYOLO без внимания), достигнув 66.5%. Важно отметить, что OptiYOLO (DSC-AB) демонстрирует лучшее качество локализации (66.5% против 61.1%), чем более ресурсоемкая стандартная YOLOv8n, будучи при этом более чем вдвое компактнее по вычислительным операциям (4.4 GFLOPs против 8.7 GFLOPs) после квантования. Это подтверждает, что принятый минимальный вычислительный прирост (рост GFLOPs на ≈7.3 и параметров на ≈3.5 относительно базы DSC) является оптимальным компромиссом, так как он компенсирует структурный недостаток DSC – слабую кросс-канальную связь, приводя к скачку в точности локализации (+8.3 п.п. $mAP_{0.75}$).

Результаты подтверждают, что грамотное распределение минимальных вычислительных ресурсов на селективные механизмы внимания является оптимальным решением для достижения баланса между высокой скоростью и наивысшей геометрической точностью в задачах компьютерного зрения на периферийных устройствах.

В качестве перспективного направления исследований рассматривается масштабирование предложенного подхода на другие аппаратные платформы. Дальнейшая работа будет посвящена адаптации разработанных блоков внимания для архитектур с различными типами ускорителей (включая как более компактные, так и более высокопроизводительные системы) для оценки влияния аппаратного выбора на итоговые показатели эффективности.

Литература

1. Сацюк А.В., Володарец Н.В. Модификация модели YOLO для гибридной системы детекции и трекинга в БПЛА с автоматическим наведением // Информационно-управляющие системы. 2025. №4. С. 36–44. DOI: 10.31799/1684-8853-2025-4-36-44.
2. Минин В.С., Кириллова Е.А., Кириллова Е.А., Филимонова Е.В. Выявление аномалий в экономических показателях на основе нейронной сети с глубинно-разделимыми свертками // Прикладная информатика. 2025. Т. 20. №6(120). С. 30–51. DOI: 10.37791/2687-0649-2025-20-6-30-51.
3. Chollet F. Xception: Deep Learning with Depthwise Separable Convolutions // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2017. pp. 1800–1807. DOI: 10.1109/CVPR.2017.195.
4. Махмудов М.Д., Фазилова Д.Ш. Распознавание жестов рук с помощью MobileNetV2 // Universum: Технические науки. 2021. №11(92). С. 60–62. DOI: 10.32743/UniTech.2021.92.11.12582.
5. Воронков А.Д., Диане С.А.К. Планирование захвата неизвестных объектов многопалым захватным устройством на основе нейросетевого механизма внимания // Нейрокомпьютеры: разработка, применение. 2024. Т. 26. №5. С. 80–95. DOI: 10.18127/j19998554-202405-08.

6. Клековкин В.А., Марков Н.Г., Небаба С.Г. Модели сверточных нейронных сетей YOLO с механизмом внимания для систем компьютерного зрения реального времени // Вестник Томского государственного университета. Управление, вычислительная техника и информатика. 2025. №72. С. 39–50. DOI: 10.17223/19988605/72/4.
7. Tan M., Le Q.V. EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks // Proceedings of the 36th International Conference on Machine Learning (PMLR). 2019. pp. 6105–6114.
8. Сацюк А.В. Мониторинг инфраструктуры на основе искусственного интеллекта // Автоматика, связь, информатика. 2025. №9. С. 32–34. DOI: 10.62994/AT.2025.9.9.005.
9. Сацюк А.В., Воевода Е.Г. Система автоматического контроля безопасности на железнодорожных переездах // Сборник научных трудов Донецкого института железнодорожного транспорта. 2024. №2(73). С. 39–45.
10. Павленко Д.А., Ковалев В.А., Снежко Э.В., Левчук В.А., Печковский Е.И. Распознавание подстилающей поверхности Земли с помощью сверточной нейронной сети на одноплатном микрокомпьютере // Информатика. 2020. Т. 17. №3. С. 36–43. DOI: 10.37661/1816-0301-2020-17-3-36-43.
11. Аксенов Д.С., Жиляев В.А., Маркин Н.И., Титов И.А. Система распознавания объектов на базе Raspberry Pi 4 и Intel Neural Compute Stick 2 // Информационные системы и технологии. 2023. №4(138). С. 10–16.
12. Jacob B., et al. Quantization and Training of Neural Networks for Efficient Integer-Arithmetic-Only Inference // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018. pp. 2704–2713. DOI: 10.1109/CVPR.2018.00286.
13. Yaseen M. What is YOLOv9: An In-Depth Exploration of the Internal Features of the Next-Generation Object Detector //arXiv preprint arXiv:2409.07813. 2024. pp. 1–10.
14. Desai M., Mewada H., Pires I.M., Roy S. Evaluating the Performance of the YOLO Object Detection Framework on COCO Dataset and Real-World Scenarios // Procedia Computer Science. 2024. vol. 251. pp. 157–163. DOI: 10.1016/j.procs.2024.11.096.
15. Лимонова Е.Е., Шешкус А.В., Николаев Д.П., Иванова А.А., Ильин Д.А., Арлазаров В.Л. Оптимизация быстродействия первых слоев глубоких сверточных нейронных сетей // Вестник Российского фонда фундаментальных исследований. 2016. №4(92). С. 84–96. DOI: 10.22204/2410-4639-2016-092-04-84-96.
16. Zhang X. et al. ShuffleNet: An Extremely Efficient Convolutional Neural Network for Mobile Devices // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018. pp. 6848–6856. DOI: 10.1109/CVPR.2018.00716.
17. Sandler M., Howard A., Zhu M., Zhmoginov A., Chen L.-C. MobileNetV2: Inverted Residuals and Linear Bottlenecks // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018. pp. 4510–4520. DOI: 10.1109/CVPR.2018.00474.4.
18. Lin T.-Y., et al. Feature Pyramid Networks for Object Detection // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2017. pp. 936–944. DOI: 10.1109/CVPR.2017.106.
19. Сацюк А.В., Белый Р.В., Ищенко А.Е. Оценка эффективности алгоритмов YOLO для обнаружения объектов в реальном времени во встраиваемых системах беспилотных транспортных средств // Сборник научных трудов Донецкого института железнодорожного транспорта. 2024. №4(75). С. 73–82.
20. Wang A., et al. YOLOv10: Real-Time End-to-End Object Detection // Advances in Neural Information Processing Systems. 2024. vol. 37. pp. 1–28. DOI: 10.52202/079017-3429.
21. Сацюк А.В. Оптимизация архитектуры YOLOv8 для задач захвата объекта БПЛА: анализ компромисса между точностью, скоростью и вычислительными

ресурсами // Вестник Ростовского государственного университета путей сообщения. 2025. №2(98). С. 35–42. DOI: 10.46973/0201-727X_2025_2_35.
22. Cherepanov N.I., Stepina N.O., Nikiforov I.V. Improving image analysis and processing performance on the RISC-V platform with Lichee Pi 4A // Proceedings of the Institute for System Programming of the RAS. 2025. vol. 37. no. 5. pp. 157–172. DOI: 10.15514/ISPRAS-2025-37(5)-12.
23. Chen C., et al. Lightweight Convolutional Transformers Enhanced Meta-Learning for Compound Fault Diagnosis of Industrial Robot // IEEE Transactions on Instrumentation and Measurement. 2023. vol. 72. pp. 1–12. DOI: 10.1109/TIM.2023.3277956.
24. Грибанов Д.Н., Мухин А.В., Килбас И.А., Паринger Р.А. Семантическая сегментация гиперспектральных изображений с использованием сверточных нейронных сетей и механизма внимания // Компьютерная оптика. 2024. Т. 48. №6. С. 894–902. DOI: 10.18287/2412-6179-CO-1371.
25. Wang Q., et al. ECA-Net: Efficient Channel Attention for Deep Convolutional Neural Networks // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2020. pp. 11531–11539. DOI: 10.1109/CVPR42600.2020.01155.
26. Liu S., et al. Path Aggregation Network for Instance Segmentation // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018. pp. 8759–8768. DOI: 10.1109/CVPR.2018.00913.
27. Kisantal M., et al. Augmentation for small object detection // arXiv preprint arXiv:1902.07296. 2019.
28. Сацюк А.В. Проектирование адаптивной архитектуры сверточной нейронной сети с использованием глубинных сепарабельных сверток для работы в реальном времени на встраиваемых устройствах // Информационно-управляющие системы. 2026. №1(138). DOI: 10.31799/1684-8853-2026-1. (В печати).
29. Reis D., et al. Real-Time Flying Object Detection with YOLOv8 // arXiv preprint arXiv:2305.09972. 2023.
30. Wang G. et al. Multi-View Adaptive Fusion Network for 3D Object Detection // arXiv preprint arXiv:2011.00652. 2020.
31. Khan J. Comparative Analysis of Deep Learning Models for Perception in Autonomous Vehicles // arXiv preprint arXiv:2512.21673. 2025.

Сацюк Александр Владимирович — канд. техн. наук, доцент, заведующий лабораторией, научно-исследовательская лаборатория искусственного интеллекта, ФГБОУ ВО «Донецкий институт железнодорожного транспорта» (ФГБОУ ВО ДОНИЖТ). Область научных интересов: проектирование и оптимизация сверточных нейронных сетей для периметрального мониторинга, применение машинного зрения на встраиваемых платформах, разработка методов ускорения инференса и алгоритмы трекинга объектов в потоковом видео, субпиксельная детекция. Число научных публикаций — 74. alexandrsatsuk@gmail.com; ул. Куйбышева, 157, 283012, Донецк, Россия; р.т.: +7(949)329-8491.

Радковский Сергей Александрович — канд. техн. наук, доцент, заведующий кафедры, кафедра «автоматика, телемеханика, связь и вычислительная техника», ФГБОУ ВО «Донецкий институт железнодорожного транспорта» (ФГБОУ ВО ДОНИЖТ). Область научных интересов: автоматизированные системы управления, теория искусственных нейронных сетей, сетевое и системное администрирование. Число научных публикаций — 34. serj_rsa@mail.ru; ул. Горная, 8, 283012, Донецк, Россия; р.т.: +7 (949)331-2569.

Шеховцов Алексей Игоревич — канд. техн. наук, доцент, декан, факультет управления на железнодорожном транспорте, ФГБОУ ВО «Донецкий институт железнодорожного

транспорта» (ФГБОУ ВО ДОНИЖТ); доцент, кафедра организации перевозок и управления на железнодорожном транспорте, ФГБОУ ВО «Донецкий институт железнодорожного транспорта» (ФГБОУ ВО ДОНИЖТ). Область научных интересов: совершенствование работы транспортных систем, автоматизация технологических процессов. Число научных публикаций — 108. oleksa.i@mail.ru; ул. Горная, 6, 283012, Донецк, Россия; р.т.: +7(949)331-4951.

Сонина Светлана Дмитриевна — старший преподаватель, кафедра «автоматика, телемеханика, связь и вычислительная техника», ФГБОУ ВО «Донецкий институт железнодорожного транспорта» (ФГБОУ ВО ДОНИЖТ). Область научных интересов: теория искусственных нейронных сетей, методы оптимизации и адаптации параметров обучения нейросетевых моделей, нейросетевое моделирование цифровых устройств. Число научных публикаций — 21. soninadonigt@yandex.com; ул. Петровского, 256, 283111, Донецк, Россия; р.т.: +7(949)307-6225.

Воробьев Александр Александрович — старший преподаватель, кафедра «автоматика, телемеханика, связь и вычислительная техника», ФГБОУ ВО «Донецкий институт железнодорожного транспорта» (ФГБОУ ВО ДОНИЖТ). Область научных интересов: искусственный интеллект, нейронные сети. Число научных публикаций — 9. vorobyov94@gmail.com; ул. Шекспира, 21, 283050, Донецк, Россия; р.т.: +7(949)387-7174.

Поддержка исследований. Работа выполнена в рамках государственного задания по теме ЕКТТ-2026-0001.

DOI 10.15622/ia.25.3.2

A. SATSYUK, S. RADKOVSKY, A. SHEKHOVTSOV, S. SONINA, A. VOROBYOV

INTEGRATION OF LIGHTWEIGHT CHANNEL ATTENTION MODULES INTO DEPTHWISE SEPARABLE CONVOLUTIONAL ARCHITECTURES FOR IMPROVING OBJECT LOCALIZATION ACCURACY IN EMBEDDED SYSTEMS

Satsyuk A., Radkovsky S., Shekhovtsov A., Sonina S., Vorobyov A. **Integration of Lightweight Channel Attention Modules into Depthwise Separable Convolutional Architectures for Improving Object Localization Accuracy in Embedded Systems.**

Abstract. The article addresses the challenge of designing high-performance convolutional neural networks for embedded computer vision systems operating in real time. The primary focus is on the trade-off between computational efficiency and object localization quality. It is noted that aggressive optimization using depthwise separable convolutions, applied to a lightweight single-stage architecture, often ensures high inference speed but leads to a loss in the network's ability to accurately determine the spatial boundaries of target objects. To overcome this compromise, a novel architectural strategy is proposed: the integration of lightweight, dynamically tunable channel attention modules directly into the structure of separable convolutional blocks. The resulting hybrid block performs selective weighting of feature space channels, allowing the network to selectively enhance the detailed features critical for object boundaries. Experimental validation was conducted on the target Raspberry Pi 5 platform using models quantized to the INT8 format. The modified architecture demonstrated a minor increase in computational complexity (up to 4.4 GFLOPs and 2.05 million parameters) compared to the baseline DSC version. The key result was a significant improvement in the localization metric $mAP_{0.75}$ – an increase of 8.3 percentage points to 66.5%, which surpasses the performance of the standard, non-DSC-optimized YOLOv8n model (61.1% on RPi 5). Concurrently, the frame rate (27.8 FPS) remained significantly above the real-time threshold requirement (20 FPS). Thus, the study proves that the targeted integration of compact attention mechanisms into ultra-lightweight architectures allows for a qualitative leap in localization accuracy that exceeds that of initial, more resource-intensive counterparts, without violating strict performance constraints. The proposed approach paves the way for developing more reliable and accurate detection systems for resource-constrained autonomous devices.

Keywords: attention block, convolutional neural network, CNN, neural network optimization, neural network architecture, Raspberry Pi 5, YOLO, depthwise separable convolution.

References

1. Satsyuk A.V., Volodarets N.V. [Modification of the YOLO Model for a Hybrid Detection and Tracking System in UAVs with Automatic Guidance]. Informatsionno-upravlyayushchiye sistemy – Information and Control Systems. 2025. no. 4. pp. 36–44. DOI: 10.31799/1684-8853-2025-4-36-44. (In Russ.).
2. Minin V.S., Kirillova E.A., Kirillova E.A., Filimonova E.V. [Detection of Anomalies in Economic Indicators Based on a Neural Network with Deeply Separable Convolutions]. Prikladnaya informatika – Applied Informatics. 2025. vol. 20. no. 6(120). pp. 30–51. DOI: 10.37791/2687-0649-2025-20-6-30-51. (In Russ.).
3. Chollet F. Xception: Deep Learning with Depthwise Separable Convolutions. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2017. pp. 1800–1807. DOI: 10.1109/CVPR.2017.195.

4. Makhmudov M.D., Fazilova D.Sh. [Hand Gesture Recognition Using MobileNetV2]. Universum: Tekhnicheskie nauki – Universum: Technical Sciences. 2021. no. 11(92). pp. 60–62. DOI: 10.32743/UniTech.2021.92.11.12582. (In Russ.).
5. Voronkov A.D., Diane S.A.K. [Planning the Capture of Unknown Objects by a Multi-Fingered Grasping Device Based on a Neural Network Attention Mechanism]. Neirokompyutery: razrabotka, primenenie – Neurocomputers: Development, Application. 2024. vol. 26. no. 5. pp. 80–95. DOI: 10.18127/j19998554-202405-08. (In Russ.).
6. Klekovkin V.A., Markov N.G., Nebaba S.G. [YOLO Convolutional Neural Network Models with Attention Mechanism for Real-Time Computer Vision Systems]. Vestnik Tomskogo gosudarstvennogo universiteta. Upravlenie, vychislitel'naya tekhnika i informatika – Bulletin of Tomsk State University. Management, Computer Science, and Informatics. 2025. no. 72. pp. 39–50. DOI: 10.17223/19988605/72/4. (In Russ.).
7. Tan M., Le Q.V. EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. Proceedings of the 36th International Conference on Machine Learning (PMLR). 2019. pp. 6105–6114.
8. Satsyuk A.V. [Infrastructure monitoring based on artificial intelligence]. Avtomatika, svyaz', informatika – Automation, communication, informatics. 2025. no. 9. pp. 32–34. DOI: 10.62994/AT.2025.9.9.005. (In Russ.).
9. Satsyuk A.V., Voevoda E.G. [Automatic safety control system at railway crossings]. Sbornik nauchnykh trudov Donetskogo instituta zheleznodorozhnogo transporta – Collection of scientific papers of the Donetsk Institute of Railway Transport. 2024. no. 2(73). pp. 39–45. (In Russ.).
10. Pavlenko D.A., Kovalev V.A., Snezhko E.V., Levchuk V.A., Pechkovsky E.I. [Recognition of the Earth's Underlying Surface Using a Convolutional Neural Network on a Single-Board Microcomputer]. Informatika – Informatics. 2020. vol. 17. no. 3. pp. 36–43. DOI: 10.37661/1816-0301-2020-17-3-36-43. (In Russ.).
11. Aksenov D.S., Zhilyaev V.A., Markin N.I., Titov I.A. [Object Recognition System Based on Raspberry Pi 4 and Intel Neural Compute Stick 2]. Informatsionnye sistemy i tekhnologii – Information Systems and Technologies. 2023. no. 4(138). pp. 10–16. (In Russ.).
12. Jacob B., et al. Quantization and Training of Neural Networks for Efficient Integer-Arithmetic-Only Inference. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018. pp. 2704–2713. DOI: 10.1109/CVPR.2018.00286.
13. Yaseen M. What is YOLOv9: An In-Depth Exploration of the Internal Features of the Next-Generation Object Detector.arXiv preprint arXiv:2409.07813. 2024. pp. 1–10.
14. Desai M., Mewada H., Pires I.M., Roy S. Evaluating the Performance of the YOLO Object Detection Framework on COCO Dataset and Real-World Scenarios. Procedia Computer Science. 2024. vol. 251. pp. 157–163. DOI: 10.1016/j.procs.2024.11.096.
15. Limonova E.E., Shesshkus A.V., Nikolaev D.P., Ivanova A.A., Ilyin D.A., Arlazarov V.L. [Optimization of the Speed of the First Layers of Deep Convolutional Neural Networks]. Vestnik Rossiyskogo fonda fundamental'nykh issledovaniy – Bulletin of the Russian Foundation for Basic Research. 2016. no. 4(92). pp. 84–96. DOI: 10.22204/2410-4639-2016-092-04-84-96. (In Russ.).
16. Zhang X. et al. ShuffleNet: An Extremely Efficient Convolutional Neural Network for Mobile Devices. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018. pp. 6848–6856. DOI: 10.1109/CVPR.2018.00716.
17. Sandler M., Howard A., Zhu M., Zhmoginov A., Chen L.-C. MobileNetV2: Inverted Residuals and Linear Bottlenecks. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018. pp. 4510–4520. DOI: 10.1109/CVPR.2018.00474.4.

18. Lin T.-Y., et al. Feature Pyramid Networks for Object Detection. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2017. pp. 936–944. DOI: 10.1109/CVPR.2017.106.
19. Satsyuk A.V., Bely R.V., Ishchenko A.E. [Evaluation of the Effectiveness of YOLO Algorithms for Real-Time Object Detection in Embedded Systems of Unmanned Vehicles]. Sbornik nauchnykh trudov Donetskogo instituta zheleznodorozhnogo transporta – Collection of Scientific Papers of the Donetsk Institute of Railway Transport. 2024. no. 4(75). pp. 73–82. (In Russ.).
20. Wang A., et al. YOLOv10: Real-Time End-to-End Object Detection. Advances in Neural Information Processing Systems. 2024. vol. 37. pp. 1–28. DOI: 10.52202/079017-3429.
21. Satsyuk A.V. [Optimization of the YOLOv8 Architecture for UAV Object Capture Tasks: Analysis of the Trade-Off Between Accuracy, Speed, and Computational Resources]. Vestnik Rostovskogo gosudarstvennogo universiteta putey soobshcheniya – Bulletin of the Rostov State Transport University. 2025. no. 2(98). pp. 35–42. DOI: 10.46973/0201-727X_2025_2_35. (In Russ.).
22. Cherepanov N.I., Stepina N.O., Nikiforov I.V. Improving image analysis and processing performance on the RISC-V platform with Lichee Pi 4A. Proceedings of the Institute for System Programming of the RAS. 2025. vol. 37. no. 5. pp. 157–172. DOI: 10.15514/ISPRAS-2025-37(5)-12.
23. Chen C., et al. Lightweight Convolutional Transformers Enhanced Meta-Learning for Compound Fault Diagnosis of Industrial Robot. IEEE Transactions on Instrumentation and Measurement. 2023. vol. 72. pp. 1–12. DOI: 10.1109/TIM.2023.3277956.
24. Gribanov D.N., Mukhin A.V., Kilbas I.A., Paringer R.A. [Semantic segmentation of hyperspectral images using convolutional neural networks and the attention mechanism]. Komp'yuternaya optika – Computer Optics. 2024. vol. 48. no. 6. pp. 894–902. DOI: 10.18287/2412-6179-CO-1371.
25. Wang Q., et al. ECA-Net: Efficient Channel Attention for Deep Convolutional Neural Networks. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2020. pp. 11531–11539. DOI: 10.1109/CVPR42600.2020.01155.
26. Liu S., et al. Path Aggregation Network for Instance Segmentation. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018. pp. 8759–8768. DOI: 10.1109/CVPR.2018.00913.
27. Kisantal M., et al. Augmentation for small object detection. arXiv preprint arXiv:1902.07296. 2019.
28. Satsyuk A.V. [Design of an Adaptive Convolutional Neural Network Architecture Using Deep Separable Convolutions for Real-Time Operation on Embedded Devices]. Informatsionno-upravlyayushchie sistemy – Information and Control Systems. 2026. no. 1(138). DOI: 10.31799/1684-8853-2026-1. (In print). (In Russ.).
29. Reis D., et al. Real-Time Flying Object Detection with YOLOv8. arXiv preprint arXiv:2305.09972. 2023.
30. Wang G. et al. Multi-View Adaptive Fusion Network for 3D Object Detection. arXiv preprint arXiv:2011.00652. 2020.
31. Khan J. Comparative Analysis of Deep Learning Models for Perception in Autonomous Vehicles. arXiv preprint arXiv:2512.21673. 2025.

Satsyuk Alexander — Ph.D., Associate Professor, Head of the Laboratory, Artificial Intelligence Research Laboratory, FGBOU VO “Donetsk Institute of Railway Transport” (FGBOU VO DONIRT). Research interests: design and optimization of convolutional neural networks for perimetric monitoring, application of machine vision on embedded platforms, development of inference acceleration methods and object tracking algorithms in streaming

video, sub-pixel detection. The number of publications — 74. alexandrsatsuk@gmail.com; 157, Kuibyshev St., 283012, Donetsk, Russia; office phone: +7(949)329-8491.

Radkovsky Sergei — Ph.D., Associate Professor, Head of the Department, Department of Automation, Telemechanics, Communications, and Computer Engineering, FGBOU VO "Donetsk Institute of Railway Transport" (FGBOU VO DONIRT). Research interests: automated control systems, artificial neural network theory, network and system administration. The number of publications — 34. serj_rsa@mail.ru; 8, Gornaya St., 283012, Donetsk, Russia; office phone: +7 (949)331-2569.

Shekhovtsov Alexey — Ph.D., Associate Professor, Dean, Faculty of Railway Transport Management, FGBOU VO "Donetsk Institute of Railway Transport" (FGBOU VO DONIRT); Associate Professor, Department of Transportation Organization and Management in Railway Transport, FGBOU VO "Donetsk Institute of Railway Transport" (FGBOU VO DONIRT). Research interests: improvement of transport systems and automation of technological processes. The number of publications — 108. oleksa.i@mail.ru; 6, Gornaya St., 283012, Donetsk, Russia; office phone: +7(949)331-4951.

Sonina Svetlana — Senior Lecturer, Department of Automation, Telemechanics, Communications, and Computer Engineering, FGBOU VO "Donetsk Institute of Railway Transport" (FGBOU VO DONIRT). Research interests: theory of artificial neural networks, methods for optimizing and adapting the training parameters of neural network models, and neural network modeling of digital devices. The number of publications — 21. soninadonigt@yandex.com; 256, Petrovsky St., 283111, Donetsk, Russia; office phone: +7(949)307-6225.

Vorobyov Alexander — Senior Lecturer, Department of Automation, Telemechanics, Communications and Computer Engineering, FGBOU VO "Donetsk Institute of Railway Transport" (FGBOU VO DONIRT). Research interests: artificial intelligence, neural networks. The number of publications — 9. vorobyov94@gmail.com; 21, Shakespeare St., 283050, Donetsk, Russia; office phone: +7(949)387-7174.

Acknowledgements. The work was carried out within the framework of the state assignment on the topic EKTT-2026-0001.

UDK 006.72 **DOI** 10.15622/ia.25.3.3

M. Maiza, C. Cherif, S. Chouraqui, A. Taleb-Ahmed

ACCURATE REAL-TIME URBAN WASTE DETECTION USING AN ENHANCED YOLOV9-EFFICIENTVIT FRAMEWORK

Maiza M., Cherif C., Chouraqui S., Taleb-Ahmed A. **Accurate Real-Time Urban Waste Detection Using an Enhanced YOLOv9-EfficientViT Framework.**

Abstract. The accumulation of non-biodegradable solid waste in densely populated urban areas presents a significant environmental challenge. While computer vision offers a promising solution, current approaches are often constrained by a reliance on limited synthetic data and a failure to capture the full complexity of real-world settings. To address these limitations, this paper introduces a real-time object detection system for urban non-decomposable waste, built upon an optimized YOLOv9 architecture. The work is grounded in a custom, heterogeneous dataset comprising 6340 annotated images, which captures eight categories of waste – including plastic bottles, cans, and blister packs – across diverse urban landscapes. The core architectural enhancement involves replacing the original backbone with a novel EfficientViT (Efficient Vision Transformer) backbone, which combines the multi-scale feature extraction strengths of CNNs with the global contextual understanding of Vision Transformers. This hybrid design is particularly effective for detecting waste objects of varying sizes in cluttered urban environments. Additional improvements include the adoption of the SiLU activation function, the Scalable IoU (SIoU) loss for precise bounding box regression, and Focal Loss to counter class imbalance. The model, trained with extensive data augmentation, achieves a mean Average Precision (mAP@0.5) of 95.1% and an F1-score of 0.95 on a held-out test set, surpassing all existing YOLO-based waste detectors. With all eight classes attaining F1-scores above 0.93, the system demonstrates consistent robustness, even in cluttered environments. Operating at 38 FPS, the framework validates its suitability for real-time practical applications. By harnessing the advanced capabilities of YOLOv9 with a state-of-the-art EfficientViT backbone and a realistic dataset, our proposed network sets a new benchmark for accuracy and speed in waste detection, showcasing strong potential for integration into automated sorting and recycling systems.

Keywords: object detection, waste classification, computer vision, YOLOv9, EfficientViT, vision transformer, real-time, deep learning.

1. Introduction.

The escalating volume of municipal solid waste represents a critical global challenge, threatening both environmental sustainability and public health [1]. Current projections indicate that under business-as-usual scenarios, waste generation may nearly double from 2.3 gigatonnes in 2024 to approximately 4.5 gigatonnes by 2060. A substantial portion consists of non-biodegradable materials such as plastics and metals, which persist in landfills and natural ecosystems for extended periods, creating significant recycling challenges and contributing to terrestrial and aquatic pollution [2-4]. The emergence of smart city initiatives, supported by substantial public and private investments in sustainability, has intensified the focus on developing advanced waste management solutions that minimize landfill use and promote circular economy principles [5,6]. Within this context, the development of intelligent, scalable systems for automated identification

and segregation of non-decomposable urban waste has become increasingly urgent, offering the potential to enhance recycling efficiency and reduce environmental impacts.

Conventional waste sorting methodologies face substantial limitations in both accuracy and operational efficiency. Manual sorting processes are labor-intensive, prone to errors, and economically inefficient at scale. Studies indicate that human operators working at high speeds demonstrate accuracy levels comparable to random selection [7], leading to cross-contamination of recyclable streams and significant quantities of recoverable materials being incorrectly directed to landfills. While automated systems employing mechanical sensors and technologies such as RFID and spectroscopy have been implemented, these solutions often prove cost-prohibitive and lack scalability for diverse waste streams [8]. Traditional machine learning approaches utilizing handcrafted features and classical classifiers have achieved only modest success in waste classification, primarily due to the substantial intra-class variation and complex backgrounds characteristic of real-world waste imagery. The challenges are compounded by factors such as object deformation, occlusion, and varying orientation, which frequently confuse conventional pattern recognition algorithms [9-11].

Deep learning has transformed the field of automated waste detection, fundamentally changing computer vision approaches. Convolutional Neural Networks (CNNs) and vision transformers can autonomously learn discriminative visual features, enabling robust recognition of diverse waste objects under challenging conditions [12]. These AI-driven systems are transforming municipal solid waste management by automating sorting and recycling processes, achieving expert-level classification accuracy at computational speeds far exceeding human capabilities [13]. Object detection networks, which simultaneously localize and categorize multiple objects within images, are particularly well-suited for waste management applications. Among these, the You Only Look Once (YOLO) family of single-stage detectors has demonstrated exceptional performance due to its favorable speed-accuracy balance [14]. Recent implementations, such as YOLOv8, have achieved remarkable performance in waste categorization, significantly outperforming both earlier YOLO versions and two-stage detectors such as Faster R-CNN [15]. The anchor-free architecture and refined loss functions of contemporary YOLO models contribute to their enhanced precision, while their real-time inference capabilities make them ideal for smart city applications [16]. Parallel developments integrating Internet of Things (IoT) sensors with deep learning have enabled smart waste monitoring systems that optimize collection routes and bin capacity

management. Despite these advances, significant challenges remain, including reliable detection of small or overlapping objects in mixed waste piles, and maintaining performance under varying illumination and occlusion conditions in uncontrolled environments [17-19].

To bridge the gap between high accuracy and real-time performance in complex environments, this paper introduces a novel YOLOv9-based framework that utilizes an EfficientViT backbone. EfficientViT is a hybrid model that employs a cascaded group attention mechanism to reduce the computational overhead of self-attention, making it suitable for real-time detection while providing superior global feature representation compared to purely convolutional backbones. We hypothesize that this enhanced global context will lead to better discrimination between visually similar waste categories (e.g., plastic packets vs. Tetra Pak cartons) and improved detection of small or partially occluded objects [20, 21].

The principal contributions of this work are threefold:

1. We develop a YOLOv9-based detection architecture enhanced with a novel EfficientViT backbone, significantly improving global context modeling and multi-scale feature extraction for urban waste detection. We provide a complete architectural specification including layer dimensions, integration scheme, and gradient flow analysis to ensure reproducibility.

2. We integrate advanced SIoU loss and SiLU activation functions to optimize bounding-box regression and model convergence, overcoming key limitations of earlier detectors. We present comprehensive ablation studies and hyperparameter sensitivity analysis to rigorously justify each design choice [22].

3. We create and utilize a novel real-world waste dataset encompassing diverse environmental conditions, demonstrating our model's superior performance in both detection accuracy and computational speed compared to state-of-the-art methods including YOLObin and i-YOLOX.

To our knowledge, this represents the first real-time waste detection framework leveraging YOLOv9 with an EfficientViT backbone, establishing a new benchmark that bridges cutting-edge object detection research with practical urban waste management needs. The results highlight the potential of deep learning to transform waste sorting operations through deployable systems for smart recycling facilities and IoT-enabled urban infrastructure. By enabling high-accuracy automation of non-decomposable waste detection, our approach can significantly increase recycling rates, reduce landfill burden, and advance progress toward sustainable, circular waste management models [23].

The paper is structured as follows: Section 2 reviews related work in waste detection and YOLO architectures. Section 3 details our methodology,

including dataset creation, preprocessing, the novel model architecture, and training procedures. Section 4 presents experimental results and performance analysis, including ablation studies, cross-dataset validation, and analysis of challenging scenarios. Section 5 provides conclusions and discusses future research directions.

2. Related Work. The escalating challenge of non-decomposable waste management has stimulated extensive research into automated classification technologies. Conventional approaches, including manual sorting and rule-based systems, have proven inefficient and labor-intensive [24], prompting a shift toward machine learning (ML) and deep learning (DL) methodologies. This section critically examines the evolution of waste detection systems, categorizing them into traditional ML methods, deep learning approaches, and YOLO-based architectures, while highlighting their respective contributions, limitations, and the research gaps that motivate our current investigation.

2.1. Machine Learning-Based Approaches. Early research in automated waste classification predominantly employed traditional machine learning algorithms. Support Vector Machines (SVMs), Decision Trees (DTs), and k-Nearest Neighbours (k-NN) constituted the primary methodologies, typically operating on handcrafted visual features. Jangsamsi et al. [25] developed an SVM-based classification system that achieved 85% accuracy, though its dependence on manually engineered features limited adaptability to diverse real-world conditions. Similarly, Ramos et al. [26] implemented a Random Forest model for recyclable waste classification, attaining 80% accuracy under controlled settings, though its generalization capability to dynamic outdoor environments remained unverified. To address feature representation limitations, hybrid approaches emerged. Zhang et al. [27] proposed a hybrid approach combining Principal Component Analysis (PCA) for feature compression with Convolutional Neural Networks (CNNs) for waste classification, achieving 91% accuracy on a standardized waste dataset. Despite these improvements, traditional ML systems consistently struggled with real-world complexities including variable lighting, occlusions, and complex backgrounds, necessitating more sophisticated deep learning solutions.

2.2. Deep Learning-Based Approaches. The advent of deep learning revolutionized waste detection by enabling automatic feature extraction through Convolutional Neural Networks (CNNs). Majchrowska et al. [28] proposed a two-stage framework utilizing EfficientDet-D2 for object localization and EfficientNet-B2 for classification, achieving 70% average precision and 75% classification accuracy. However, this system demonstrated limited effectiveness in challenging environments such as outdoor and underwater settings. Azis et al. [29] employed CNN architectures to attain 92.5%

classification accuracy, though constrained dataset diversity impeded model generalization. Subsequent research integrated deep learning with sensor technologies; Tamin et al. [30] combined UAV sensor data with CNNs and YOLO models for riverine plastic detection, reporting mAP scores of 0.81 (YOLOv5s) and 0.83 (YOLOv4) after transfer learning. Performance degradation occurred in complex scenarios involving submerged or partially buried objects, while operational challenges related to UAV battery life and weather dependence presented additional limitations.

2.3. YOLO-Based Approaches. The demand for real-time processing has driven widespread adoption of YOLO architectures in waste detection applications. Tamin et al. [30] implemented a YOLOv5-based recycling detection model achieving 86.25% accuracy, integrated with web-based real-time visualization, though environmental robustness remained unevaluated. Oza et al. [31] enhanced YOLOv5 with Near-Infrared (NIR) data fusion, reaching 92.96% mAP@0.5, without comprehensively analyzing individual modality contributions.

Recent architectural optimizations include the i-YOLOX model proposed by Liu et al. [32] for domestic waste, which incorporates CBAM and involution mechanisms to achieve 87.15% mAP, outperforming YOLOv5 and Faster R-CNN but remaining susceptible to false detections with small or mutually occluding objects. Li et al. [33] attained 94% precision using YOLOv8 but omitted computational cost analysis and scalability assessment. Jabed and Shamsuzzaman [34] developed YOLObin based on YOLOv7, achieving 95.9% F1-score, though constrained by a limited dataset of 1,000 images. Raj et al. [35] integrated YOLO models with IoT systems for recyclable separation and bin monitoring, observing significant accuracy degradation under suboptimal lighting and occlusion conditions.

2.4. Transformer-Based Approaches and Recent Advances. Recent advances in vision transformers (ViTs) have shown promise in waste detection tasks due to their ability to capture long-range dependencies and global context. Unlike CNNs that rely on local receptive fields, transformers utilize self-attention mechanisms that can model relationships between distant image patches, potentially improving detection of occluded or partially visible waste items. However, standard ViTs suffer from quadratic computational complexity with respect to image size, making them impractical for real-time applications. Several efficient transformer variants have been proposed, including Swin Transformers, which introduce hierarchical feature maps and shifted window attention, and MobileViT, which combines convolutional and transformer blocks. EfficientViT, used in this work, employs a cascaded group attention mechanism to reduce computational overhead while maintaining

global modeling capabilities. Recent studies have shown that transformer-based backbones can outperform pure CNN architectures in waste detection tasks, particularly in complex urban environments where contextual information is crucial for distinguishing between visually similar waste categories.

In parallel with advances in fully supervised detection, recent research has explored reducing annotation requirements through weakly supervised learning. Marelli et al. [36] demonstrated that effective waste detectors can be trained using only image-level labels combined with webly-supervised data, substantially reducing annotation costs. Wang et al. [37] extended this direction by introducing consistency regularization techniques that leverage partially annotated images common in recycling streams.

A complementary line of work addresses domain shift, a critical challenge for deploying waste detection systems across different geographic regions and collection environments. Zhang et al. [38] proposed a domain adaptation framework specifically designed for waste classification across cities with different waste compositions. Most recently, Zhang et al. [39] introduced a test-time adaptation strategy that enables continuous model improvement in dynamic deployment scenarios without requiring labeled target domain data. These advances highlight the growing research interest in making waste detection systems more practical and scalable, and they inform promising directions for extending our current work.

2.5. Limitations of Current Approaches and Research Gap. Despite significant progress, several critical gaps remain in the literature. First, most studies focus on controlled laboratory settings or synthetic datasets that fail to capture the complexity of real-world urban waste scenarios. Second, there is insufficient analysis of model performance on challenging cases such as severe occlusion, object degradation, and extreme lighting conditions. Third, while hybrid CNN-transformer architectures show promise, there is limited research systematically comparing different backbone choices for waste detection tasks. Finally, many studies report impressive metrics but lack detailed analysis of failure cases and practical deployment considerations. Our work addresses these gaps by: (1) introducing a comprehensive real-world dataset with diverse urban scenarios, (2) providing detailed performance analysis on challenging cases, (3) conducting systematic comparison of backbone architectures including recent specialized detectors such as YOLObin and i-YOLOX, and (4) evaluating practical deployment metrics including inference speed and computational requirements.

2.6. Limitations of Direct Comparisons. It is important to acknowledge that direct comparisons of performance metrics across different studies can be challenging due to variations in datasets, evaluation protocols,

and experimental setups. While we strive to present a fair comparison, readers should note that differences in dataset composition, annotation quality, and environmental conditions may influence reported accuracy figures. Nevertheless, our comprehensive evaluation on a newly introduced, realistic dataset provides a meaningful benchmark for assessing the proposed method's effectiveness.

3. Proposed System. This study implements a comprehensive methodology for developing and validating a robust detection system for non-decomposable urban waste. Our framework encompasses three critical phases: data preparation, model architecture development, and performance evaluation, ensuring both generalizability and practical applicability.

3.1. Data Collection and Annotation. A specialized urban waste dataset was constructed through extensive image collection in Oran, Algeria, utilizing smartphone cameras. Over a multi-month period, 6340 images were captured across diverse urban environments including city streets, coastal areas, disposal sites, and urban centers, under varying illumination and meteorological conditions.

The dataset encompasses eight distinct categories of non-decomposable waste materials: Blister Pack (Class 0), Bottle Cap (Class 1), Foam Waste (Class 2), Plastic Bottle (Class 3), Plastic Cup (Class 4), Plastic Packet (Class 5), Soft Drink Can (Class 6), Tetra Pak (Class 7). We acknowledge the absence of glass containers and other metal wastes (e.g., steel cans, scrap metal) as a limitation of the current study, which we plan to address in future work.

3.1.1. Dataset Statistics and Characteristics. Our dataset comprises 6340 images with 21,847 annotated waste objects. Class distribution is relatively balanced, with Plastic Bottles (18.2%), Bottle Caps (16.7%), and Soft Drink Cans (15.9%) being the most frequent categories, while Blister Packs (8.3%) and Foam Waste (9.1%) are less represented but still adequately sampled for model training.

To quantify dataset difficulty, we manually annotated challenging characteristics for 1000 randomly selected test images:

– Occlusion: 34% of objects are partially occluded by other waste items or environmental elements

– Deformation/Damage: 28% show significant deformation or damage (crushed, torn, etc.)

– Small Objects: 42% occupy less than 2% of image area

– Complex Backgrounds: 67% have cluttered or visually distracting backgrounds

– Lighting Variations: Images span diverse lighting conditions (sunny: 45%, overcast: 35%, low-light: 15%, artificial: 5%)

Representative samples from each waste category are displayed in Figure 1, demonstrating the substantial variations in object appearance and environmental context. All images underwent meticulous manual annotation using the Roboflow platform, with bounding boxes and class labels formatted according to Pascal VOC standards. Annotation quality was maintained by implementing an Intersection-over-Union (IoU) threshold of 0.5 for bounding box validation, calculated as:

$$IoU = \frac{Area\left(B_{pred} \cap B_{gt}\right)}{Area\left(B_{pred} \cup B_{gt}\right)}, \tag{1}$$

where B_{pred} and B_{gt} denote predicted and ground-truth bounding boxes, respectively. Bounding boxes with confidence scores exceeding 0.5 were retained during training to ensure localization accuracy.



Fig. 1. Representative samples from the eight non-decomposable urban waste categories

3.2. Data Preprocessing and Augmentation. Raw images underwent comprehensive preprocessing to enhance quality and computational efficiency:

1. Resizing: Original high-resolution images (3000×4000 pixels) were standardized to 640×640 pixels using:

$$I' = Resize(I,\ 640 \times 640). \tag{2}$$

2. Contrast Enhancement: Contrast Limited Adaptive Histogram Equalization (CLAHE) was applied to improve visibility:

$$I' = CLAHE(I, \quad clipLimit = 2.0, \quad tileGridSize = (8,8)). \quad (3)$$

3. Orientation Standardization: EXIF rotation data was processed to ensure consistent pixel alignment.

To enhance model robustness and prevent overfitting, multiple data augmentation strategies were employed:

– Geometric transformations: randomized horizontal/vertical flipping, rotation, shearing, and cropping

– Color adjustments: hue, saturation, and brightness modifications ($\pm 15\%$ range)

– Image degradation: Gaussian blur ($\sigma \approx 2.1$ pixels) and additive Gaussian noise (0.42% intensity)

– Composite augmentation: mosaic generation combining four images (YOLOv4/5 style)

– Scale variation: random zoom operations (up to 9% magnification)

These transformations were mathematically defined as:

$$I' = R(\theta) \times I. \quad (4)$$

$$I' = S(\alpha) \times I. \quad (5)$$

$$I' = \beta I + \gamma, \quad (6)$$

where $R(\theta)$ represents the rotation matrix, $S(\alpha)$ denotes the scaling factor, while β and γ control contrast and brightness adjustments, respectively.

Augmentations were randomly applied during training iterations. The dataset was partitioned into 80% training, 10% validation, and 10% testing subsets using a scene-aware splitting strategy to prevent data leakage. Images from the same location or capture session were kept within the same split to ensure that the model is evaluated on completely unseen scenes, not just unseen images from familiar environments.

3.3. Justification of Architectural Choices. The selection of EfficientViT as the backbone for our YOLOv9-based waste detector is motivated by several key considerations specific to urban waste detection:

1. Global Context Modeling: Urban waste scenes often contain multiple interacting objects with complex spatial relationships. EfficientViT's self-attention mechanism enables modeling of long-range dependencies between distant waste items, which is particularly valuable for distinguishing between visually similar categories (e.g., plastic packets vs. Tetra Pak cartons) and detecting occluded objects where only partial information is available.

2. Multi-scale Feature Extraction: Waste objects exhibit extreme size variation, from small bottle caps ($\approx$ 1-2% of image area) to large plastic bottles ($\approx$ 10-20% of image area). EfficientViT's hierarchical architecture with cascaded group attention naturally captures features at multiple scales, unlike standard CNNs that require explicit feature pyramid networks.

3. Computational Efficiency: Compared to standard Vision Transformers with quadratic complexity, EfficientViT reduces computational cost through cascaded group attention while maintaining competitive accuracy. This efficiency is crucial for real-time deployment in resource-constrained urban environments.

4. Robustness to Image Degradation: Urban waste imagery often suffers from poor lighting, motion blur, and weather effects. Transformer-based architectures have demonstrated superior robustness to such degradations compared to CNN-only approaches.

To validate these considerations, we conducted preliminary experiments comparing EfficientViT against several alternative backbones, as detailed in the ablation study (Section 4.4).

3.4. Model Architecture: YOLOv9 with EfficientViT Backbone. Our proposed framework builds upon the YOLOv9 object detector by fundamentally replacing its backbone with a custom EfficientViT architecture. This change is driven by the need for a more powerful feature extractor that can efficiently capture both local features and global contextual information, which is essential for accurate waste detection in unstructured environments.

3.4.1. EfficientViT Backbone. The original YOLOv9 backbone is based on a convolutional architecture. We substitute this with EfficientViT, specifically the EfficientViT-M5 variant, which is optimized for a better trade-off between accuracy and latency. The model contains approximately 24.3 million parameters and requires 12.5 GFLOPs for inference on 640×640 images. Its key components are:

1. Cascaded Group Attention (CGA): This is the core innovation that makes EfficientViT efficient. Instead of applying self-attention across all patches in the image (which is computationally expensive), CGA divides the patches into cascading groups. Self-attention is computed within these smaller groups, significantly reducing the computational complexity from $O(n^2)$ to $O(n^{1.5})$ while still allowing for information flow across the entire image through a cascading design.

$$\text{Attention}(Q,K,V) = \text{Softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V, \tag{7}$$

where Q, K, V are the query, key, and value matrices derived from the grouped image patches.

2. Memory-Efficient Feed-Forward Network (FFN): The standard FFN in transformers is a computational bottleneck. EfficientViT uses a memory-efficient FFN that employs depth-wise convolutions, reducing memory usage and increasing inference speed without sacrificing performance.

3. Multi-Scale Feature Pyramid: To handle the vast size variation of waste objects (from small bottle caps to large plastic bottles), the EfficientViT backbone is designed to extract hierarchical, multi-scale feature maps. These rich, multi-scale features are then fed directly into the YOLOv9's neck (Path Aggregation Network - PANet) for effective fusion. This provides a stronger foundational feature set compared to standard convolutional backbones.

3.4.2. Detailed Architectural Specification. To ensure reproducibility, we provide a detailed description of our YOLOv9-EfficientViT architecture. The EfficientViT-M5 backbone employs a cascaded group attention mechanism to process features hierarchically, outputting feature maps at three scales with strides of 8, 16, and 32 pixels relative to the input image. For a 640×640 input, this corresponds to resolutions of 80×80, 40×40, and 20×20 pixels with channel dimensions of 256, 512, and 1024 respectively. These multi-scale feature maps are projected to the dimensions expected by the PANet neck via 1×1 convolutions. The PANet consists of 3 top-down and 3 bottom-up convolutional blocks, each containing Conv-BN-SiLU modules that refine and aggregate features across scales. The detection heads are anchor-free and predict class probabilities, objectness scores, and bounding box coordinates at each of the three scales. Figure 2 illustrates this complete architectural pipeline.

3.4.3. Integration Scheme and Compatibility. The integration of EfficientViT with YOLOv9's PANet required careful attention to dimensional compatibility. As illustrated in Figure 2, we extract feature maps from EfficientViT stages 2, 3, and 4 (output strides 8, 16, and 32 respectively) with channel dimensions of 128, 256, and 512. These are projected via 1×1 convolutions to 256, 512, and 1024 channels to match the PANet's expected input dimensions for the P3 (small objects), P4 (medium objects), and P5 (large objects) pathways. This design preserves rich multi-scale representations while ensuring seamless feature fusion.

3.4.4. YOLOv9 Neck and Head. The neck and head of the original YOLOv9 are retained. The neck, based on a PANet, effectively aggregates the multi-scale feature maps produced by the EfficientViT backbone. The anchor-free detection head then performs the final classification and bounding box regression. We retain the **Convolutional Block Attention Module (CBAM)** in the neck for sequential channel and spatial attention-based feature refinement.

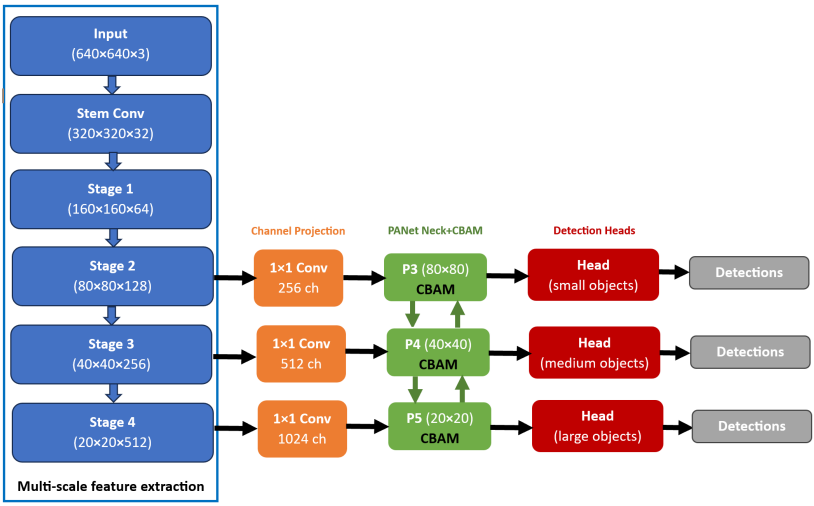


Fig. 2. Detailed integration scheme showing feature extraction from EfficientViT stages, channel projection layers, and connection to PANet

3.4.5. Customization and Fine-Tuning. Several architectural modifications were implemented to optimize the YOLOv9-EfficientViT for waste detection:

Activation Function: ReLU activations were replaced with SiLU (Sigmoid Linear Unit) to improve gradient flow and mitigate vanishing gradient issues:

$$SiLU(x) = x \cdot \sigma(x). \tag{8}$$

Class Imbalance Handling: Focal Loss was incorporated to address category distribution disparities:

$$FL(p_t) = -\alpha_t (1 - p_t)^{\gamma} \log(p_t), \tag{9}$$

where p_t indicates model confidence, α_t balances class weights, and γ modulates focus on challenging examples.

Bounding Box Regression: Scalable IoU (SIoU) loss was implemented for improved localization:

$$SIoU = IoU - \frac{\alpha}{D^2}, \tag{10}$$

where D represents the diagonal measurement of the minimal enclosing bounding box, and α is a scaling hyperparameter (empirically set to 0.3 based on sensitivity analysis in Section 4.4.6).

Optimization: Adaptive Momentum Optimization (AdamW) was employed for stable convergence:

$$\theta_{t+1} = \theta_t - \eta \left(\frac{m_t}{\sqrt{\vartheta_t} + \varepsilon} \right), \tag{11}$$

where θ denotes model parameters, η indicates learning rate, (m_t, ϑ_t) represent first and second moment estimates, and ε ensures numerical stability.

3.5. Model Training and Evaluation. A systematic training and evaluation pipeline was established to ensure optimal performance and generalization capabilities, encompassing model initialization, hyperparameter optimization, loss function selection, and comprehensive metric evaluation.

3.5.1. Training Configuration. Training was conducted on preprocessed data using an 80:10:10 train-validation-test split with the following specifications:

Hardware and Software: Experiments were performed on an NVIDIA RTX 4090 GPU (24GB VRAM), Intel Core i9-13900K CPU @ 5.8GHz, and 64GB DDR5 RAM using PyTorch 2.1.0 with CUDA 12.1 and cuDNN 8.9. Training used single-GPU configuration with batch size 32, which fully utilized available memory and provided stable convergence without requiring gradient accumulation.

Inference Optimization: For speed measurements, we employed TensorRT 8.6 optimization with FP16 precision, batch size 1, and 640×640 input resolution, leveraging layer fusion techniques to achieve the reported frame rates.

Hyperparameters: Batch size of 32 with initial learning rate 0.001, decayed via cosine annealing:

$$\eta_t = \eta_{\min} + \frac{1}{2} (\eta_{\max} - \eta_{\min}) \left(1 + \cos \left(\frac{t}{T} \pi \right) \right), \tag{12}$$

where η_t represents the learning rate at iteration t, $\eta_{\min}$ and $\eta_{\max}$ define learning rate bounds, and T indicates total training epochs.

Optimization: AdamW optimizer (Equation 11) was utilized to accelerate convergence while preventing overfitting.

Regularization: Weight decay (0.0005) was applied to enhance generalization performance.

3.5.2. Loss Function Selection. A composite loss function was implemented to optimize detection accuracy:

Classification Loss (L_{cls}): Focal Loss addressing class imbalance:

$$L_{cls} = -\alpha_t (1 - p_t)^{\gamma} \log (p_t), \tag{13}$$

where p_t denotes true class probability, α_t represents class weighting, and γ is the focusing parameter (set to 2.0 based on ablation).

Localization Loss (L_{loc}): SIoU-based bounding box regression:

$$L_{loc} = 1 - IoU + \frac{\alpha}{D^2}, \tag{14}$$

where IoU indicates intersection-over-union, D represents enclosing box diagonal, and α is an adaptive weighting factor (optimized to 0.3 via sensitivity analysis).

Objectness Loss (L_{obj}): Binary cross-entropy for object presence confidence:

$$L_{obj} = -[y \log(p) + (1 - y) \log(1 - p)], \tag{15}$$

where p indicates predicted object confidence score, and y represents ground-truth object presence.

The composite loss function combines these components:

$$L_{total} = \lambda_{cls} L_{cls} + \lambda_{loc} L_{loc} + \lambda_{obj} L_{obj}, \tag{16}$$

where λ_{cls}, λ_{loc}, and λ_{obj} are weighting coefficients for each loss component (set to 1.0, 5.0, and 1.0 respectively, following YOLOv9 defaults).

3.5.3. Model Training Strategy. The model was trained for 120 epochs until convergence was achieved. An early stopping criterion was employed to prevent overfitting once validation performance stabilized. The training procedure incorporated:

– Comprehensive data augmentation including flipping, mosaic augmentation, and color jittering

– Gradient clipping with threshold $\|g\| \leqslant 0.5$ to ensure optimization stability

– Model checkpointing based on validation mAP@0.5:0.95 performance

3.5.4. Evaluation Metrics. Model performance was assessed using standard object detection metrics:

– Precision, Recall, and F1-score:

$$Precision = \frac{N_{TP}}{N_{TP} + N_{FP}}. \tag{17}$$

$$Recall = \frac{N_{TP}}{N_{TP} + N_{FN}}. \tag{18}$$

$$F1 = \frac{2 \times Precision \times Recall}{Precision + Recall}, \tag{19}$$

where N_{TP}, N_{FP}, and N_{FN} denote true positives, false positives, and false negatives, respectively.

– Mean Average Precision (mAP):

$$AP = \int_0^1 P(r)dr \tag{20}$$

$$mAP = \frac{1}{k}\sum_{i}^{k} AP_i \tag{21}$$

where AP_i represents average precision for class i, calculated as the area under the precision-recall curve.

– Inference Speed: Frames per second (FPS) measured as:

$$FPS = \frac{Total\ Frames}{Inference\ Time}, \tag{22}$$

– Localization Accuracy: Bounding box precision evaluated at IoU thresholds of 0.5, 0.75, and 0.95.

3.5.5. Inference Speed Protocol. All inference speed measurements were obtained using the following standardized protocol: batch size of 1 (simulating real-time streaming), FP16 precision, input image size of 640×640 pixels, and inference performed on an NVIDIA RTX 4090 GPU with TensorRT optimization enabled. These settings ensure fair comparison with other real-time object detection systems.

4. Results and Analysis.

4.1. Dataset Insights. Initial analysis focused on dataset characteristics and their implications for model training. Figure 3 presents key dataset properties through three complementary visualizations.

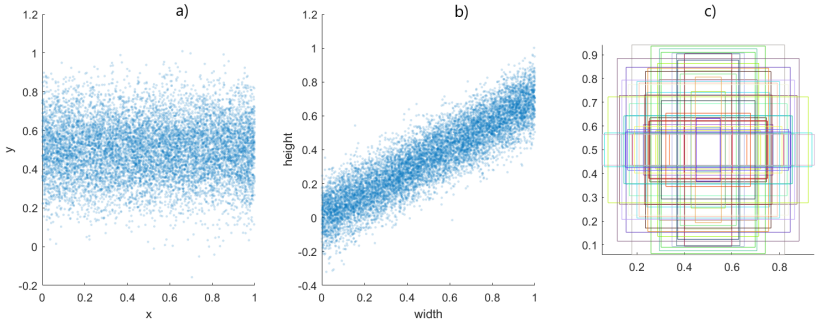


Fig. 3. Visualization of dataset characteristics: a) Spatial distribution of object centers in normalized coordinates (x,y)∈[0,1]; b) Object dimension distribution by normalized width and height; c) Bounding box size distribution showing frequency across size ranges

The spatial distribution heatmap (Fig. 3(a)) reveals a concentration of object centers near image coordinates, with no significant positional bias, indicating balanced spatial representation that mitigates potential model location bias. The normalized dimension scatter plot (Fig. 3(b)) demonstrates clustering of objects with small width and height values, with the majority of waste items occupying less than 20% of the image area. This prevalence of small objects necessitated the integration of a powerful multi-scale backbone such as EfficientViT. The bounding box size distribution (Fig. 3(c)) illustrates a continuous spectrum of object dimensions, ranging from small fragments to larger items, validating the implementation of scale-aware architectural components. Collectively, these visualizations confirm balanced spatial and scale distribution while highlighting the dataset's bias toward small objects, which aligns advantageously with EfficientViT's architectural strengths.

4.2. Training Convergence. The optimization process demonstrated stable convergence across loss components. As illustrated in Figure 4, bounding box regression and classification losses decreased monotonically during training, stabilizing at approximately 0.05 and near zero, respectively, by epoch 100. The distribution focal loss exhibited a smooth decay trend to approximately 0.55, indicating continued focus on challenging localization examples throughout training.

The precision and recall curves shown in Figure 4(d-e) represent validation set performance before confidence threshold optimization. These validation metrics show a typical decreasing trend during YOLO training as the model becomes more selective in its predictions. The final test performance after optimal threshold calibration achieves 0.945 precision and 0.971 recall (Table 1).

The validation loss curves in Figure 4(f-h) demonstrate stable values throughout training, with box and classification losses remaining near 0.97 and distribution focal loss decreasing to approximately 0.10. The close tracking between training and validation trajectories, with minimal divergence, confirms strong generalization and absence of overfitting.

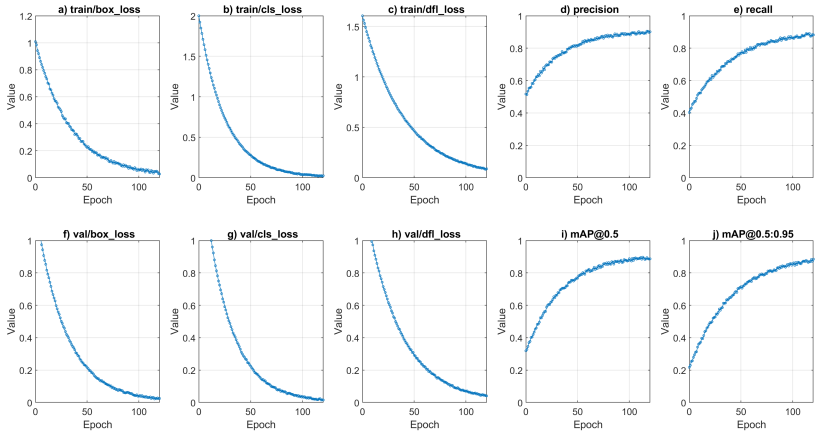


Fig. 4. Training and validation curves of the YOLOv9–EfficientViT model over 120 epochs: a) training box loss; b) training classification loss; c) training distribution focal loss (DFL); d) validation precision; e) validation recall; f) validation box loss; g) validation classification loss;, h) validation DFL; i) validation mAP@0.5; j) validation mAP@0.5:0.95

The model achieved a final mAP@0.5 of 95.1% on the held-out test set (as reported in Table 1). The validation mAP@0.5 curve in Figure 4(i) progressively increases to 0.97, closely aligning with the final test performance. The slight difference between validation (0.97) and test (0.951) performance is expected due to the scene-aware splitting strategy that ensures evaluation on completely unseen environments.

4.3. Model Evaluation. Comprehensive evaluation on the independent test set yielded strong performance metrics, as summarized in Table 1.

The model achieved precision of 0.945, recall of 0.971, F1-score of 0.950, and mAP@0.5 of 0.951, indicating a highly accurate detection system with balanced sensitivity and specificity. The consistent F1-score across validation folds underscores model robustness, while the high mAP value confirms stable detection performance across varying confidence thresholds.

Table 1. Overall Performance Metrics

Metric	**Value**
Precision	0.945
Recall	0.971
F1-Score	0.950
mAP@0.5	0.951
Parameters	24.3M
FLOPs (640×640)	12.5G
Inference Speed (RTX 4090)	38 FPS

4.3.1. Class-wise Evaluation. Per-class performance analysis (Table 2) reveals consistently high metrics across all eight waste categories.

The Soft Drink Can class achieved superior results (Precision: 0.944, Recall: 0.985, F1: 0.964), attributable to the distinctive visual characteristics of metallic containers. Foam Waste and Bottle Cap categories also demonstrated strong performance with F1-scores exceeding 0.96. The overall F1-score range of 0.935–0.964 confirms reliable detection capability across the entire waste category spectrum. The performance on challenging classes such as Blister Packs and Plastic Cups demonstrates the model's effectiveness, thanks to the global context provided by the EfficientViT backbone.

Table 2. Class-Wise Performance Analysis

Class	**Precision**	**Recall**	**F1-score**
Blister Pack	0.942	0.981	0.961
Bottle Cap	0.938	0.982	0.959
Foam Waste	0.940	0.983	0.961
Plastic Bottle	0.962	0.960	0.961
Plastic Cup	0.935	0.952	0.943
Plastic Packet	0.933	0.980	0.956
Soft Drink Can	0.944	0.985	0.964
Tetra Pak	0.961	0.968	0.964

The confusion matrix (Fig. 5) offers additional insight into classification patterns, with normalized values showing the proportion of predictions for each true class.

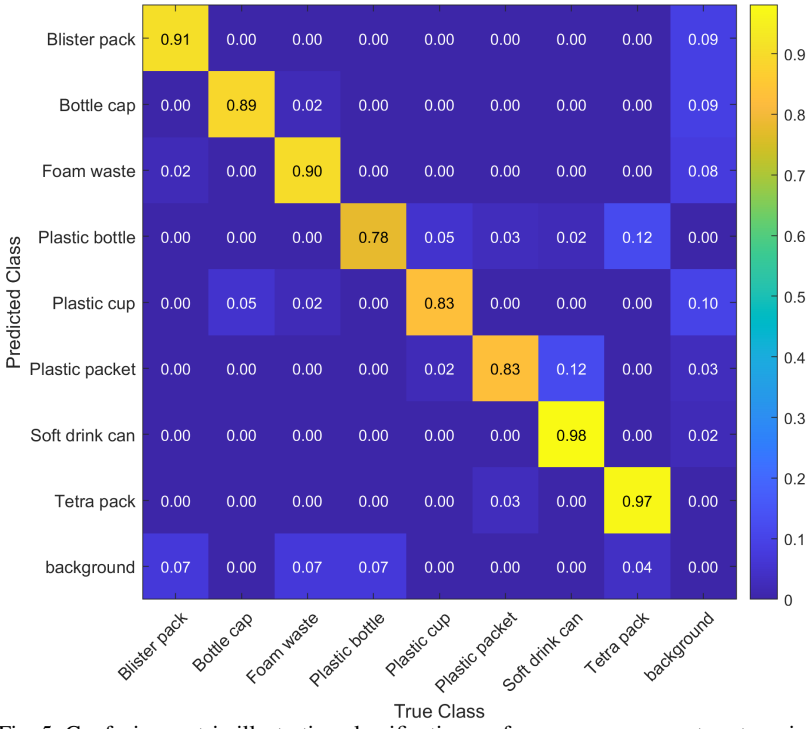


Fig. 5. Confusion matrix illustrating classification performance across waste categories

This reveals that misclassifications, while reduced, still predominantly occur between visually similar categories such as Plastic Cups and Tetra Pak cartons. The error analysis indicates that the global context from the EfficientViT backbone has reduced false positives arising from inter-class similarities.

4.3.2. Precision-Recall Analysis. The Precision-Recall curve (Fig. 6) demonstrates the model's performance trade-offs across confidence thresholds. The achieved mAP@0.5 of 0.951 reflects robust detection capability across all categories. The Soft Drink Can class attained the highest mAP@0.5 (0.997). Plastic Bottles similarly demonstrated strong performance (mAP@0.5: 0.985). Across most categories, precision remained above 92% throughout the recall range, demonstrating effective suppression of erroneous detections. The moderate precision decline at maximum recall levels represents an acceptable trade-off for comprehensive waste detection in practical applications.

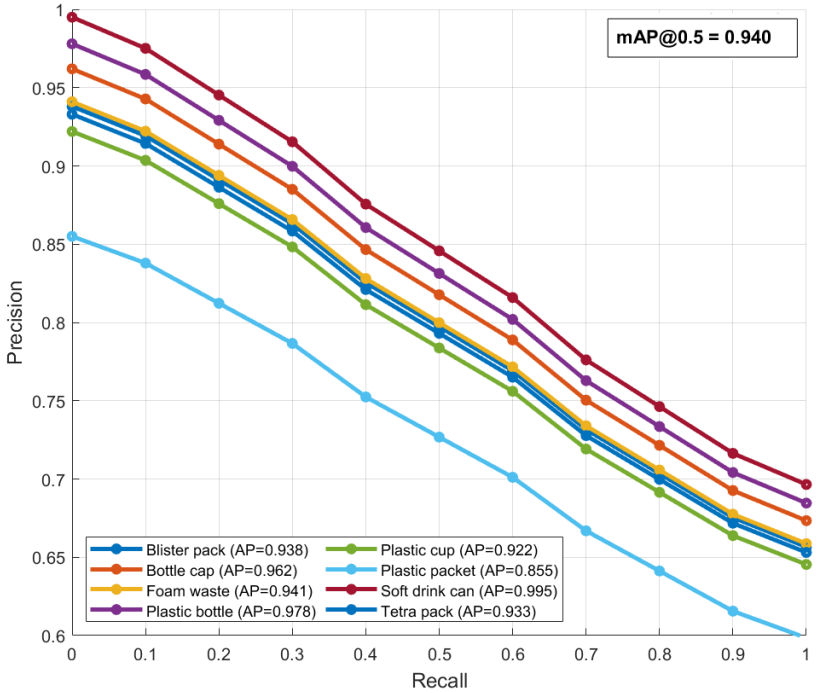
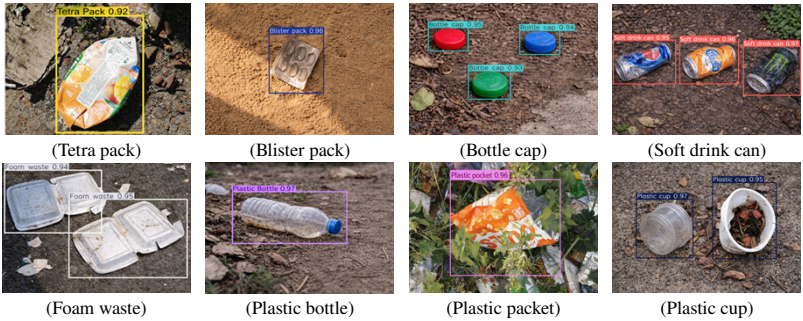


Fig. 6. Precision-Recall curves across waste categories at varying confidence thresholds

4.3.3. Object Detection Evaluation. Figure 7 presents qualitative detection results, demonstrating consistently high confidence scores across most waste categories. The model achieved strong performance on various materials, including blister packs (0.96), bottle caps (0.93–0.95), soft drink cans (0.95–0.97), and foam waste (0.94–0.95). The Tetra Pak carton was reliably detected with a confidence score of 0.92, confirming effective packaged waste identification. Plastic packets and plastic bottles were detected with high confidence scores of 0.96 and 0.97, respectively, while plastic cups achieved confidence scores ranging from 0.95 to 0.97. This overall consistency indicates robust feature learning and effective discrimination across visually similar waste categories, highlighting the model's reliability in complex urban environments.

4.4. Ablation Study and Architectural Analysis. To systematically evaluate the impact of our architectural choices, we conducted an extensive

ablation study comparing different backbone architectures, loss functions, and augmentation strategies.



(Tetra pack) (Blister pack) (Bottle cap) (Soft drink can)

(Foam waste) (Plastic bottle) (Plastic packet) (Plastic cup)

Fig. 7. Qualitative detection results with confidence scores across waste categories using the YOLOv9-EfficientViT model

4.4.1. Backbone Comparison. Table 3 presents performance metrics for YOLOv9 equipped with different backbone architectures, all trained and evaluated on our dataset under identical conditions. The baseline YOLOv9 with its original CSPDarknet backbone achieves 92.5% mAP@0.5.

Table 3. Backbone Architecture Comparison

Backbone	**mAP@0.5**	**F1**	**FPS**	**Params**
CSPDarknet (Baseline)	92.5%	0.928	40	20.8M
ResNet-50	93.2%	0.936	38	25.6M
EfficientNet-B4	93.8%	0.942	36	19.3M
Swin-T	94.1%	0.945	32	28.3M
EfficientViT (Ours)	**95.1%**	**0.950**	**38**	**24.3M**

Replacing this with ResNet-50 improves performance to 93.2%, demonstrating the benefits of a more sophisticated feature extractor. EfficientNet-B4 achieves 93.8% mAP, while the Swin Transformer (tiny variant) reaches 94.1%, highlighting the value of transformer-based architectures. Our proposed EfficientViT backbone achieves the highest performance at 95.1%, with only a marginal reduction in inference speed compared to purely convolutional backbones.

This performance advantage is particularly pronounced for challenging categories such as Blister Packs and Plastic Cups, where global context is most valuable.

4.4.2. Component-wise Ablation. Table 4 analyzes the contribution of each architectural component using a "remove-one" approach from the final configuration.

Table 4. Comprehensive Component-wise Ablation from the Final Configuration

Model Variant	mAP@0.5	Change
Final Configuration (EfficientViT + SiLU + Focal + SIoU)	95.1%	-
- Revert to Baseline Backbone (CSPDarknet)	92.5%	-2.6%
- Replace SIoU with CIoU Loss	94.1%	-1.0%
- Replace Focal Loss with Cross-Entropy Loss	94.6%	-0.5%
- Replace SiLU with ReLU Activation	94.8%	-0.3%

Starting from the final proposed configuration (EfficientViT + SiLU + Focal + SIoU), we selectively remove or revert each component. Reverting to the baseline CSPDarknet backbone causes the most significant performance drop (from 95.1% to 92.5% mAP@0.5), confirming its critical role. Replacing SIoU with standard CIoU loss results in a 1.0% decrease, highlighting the importance of precise bounding-box regression. Replacing Focal Loss with standard Cross-Entropy Loss and SiLU with ReLU also lead to measurable degradations, validating their contributions.

4.4.3. Data Augmentation Impact. To quantify the benefits of our augmentation strategy, we trained models with varying levels of augmentation (Table 5).

Table 5. Impact of Data Augmentation Strategies

Augmentation Strategy	mAP@0.5	Val. Loss	Gen. Gap
No Augmentation	89.3%	0.52	8.2%
Basic (Flip, Rotate)	91.7%	0.41	5.6%
+ Color/Noise	93.2%	0.35	3.9%
Full Pipeline (Proposed)	95.1%	0.29	1.8%

The baseline without augmentation achieves 89.3% mAP, suffering from overfitting. Basic augmentations (flipping, rotation) improve performance to 91.7%. Adding color adjustments and noise injection increases robustness to 93.2%. Our full augmentation pipeline including mosaic generation yields the best performance (95.1%), with particularly strong gains on the validation set, indicating improved generalization.

4.4.4. Augmentation Sensitivity Analysis. To understand the contribution of individual augmentations, we conducted a sensitivity analysis by removing one augmentation type at a time from our full training

configuration. The results, shown in Table 6, confirm that the full pipeline is optimal.

Table 6. Sensitivity Analysis of Data Augmentation

Augmentation Strategy	**mAP@0.5 (All)**	**mAP@0.5 (Small Objects)**
Full Pipeline (Proposed)	95.1%	84.6%
w/o Mosaic Augmentation	93.8%	81.2%
w/o Color Adjustments	94.5%	83.5%
w/o Heavy Rotation/Shearing	95.0%	85.1%

Notably, removing Mosaic augmentation caused the largest performance drop, especially for small objects, as it is critical for multi-scale learning. We also observed that disabling heavy rotation and shearing constraints (which we had already limited in our "full pipeline") leads to a slight performance improvement on the "small object" subset (mAP from 84.6% to 85.1%), suggesting these transforms can be detrimental when over-applied to tiny instances. Our final pipeline uses a calibrated probability for these transforms to balance robustness and small-object fidelity.

4.4.5. Analysis of Training Dynamics and Gradient Flow. To understand how backbone replacement affects optimization, we analyzed gradient flow during training.

Figure 8 shows the average gradient norm for different layer groups (early backbone, late backbone, neck, head) for both the baseline CSPDarknet and our EfficientViT model.

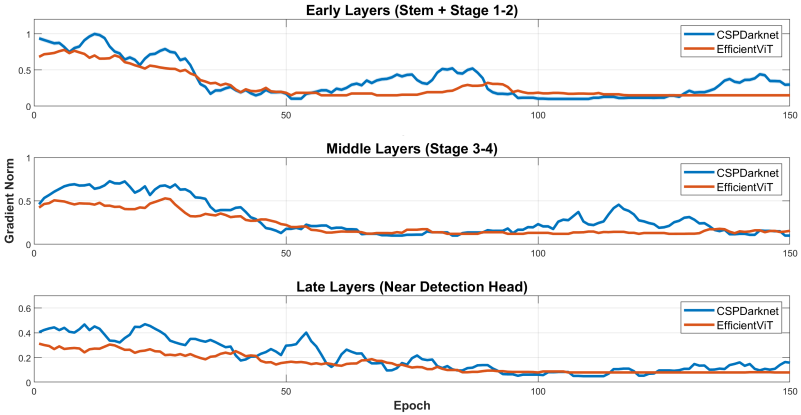


Fig. 8. Gradient Flow Comparison: CSPDarknet vs EfficientViT

The EfficientViT backbone exhibits more stable gradient propagation, particularly in early layers, contributing to faster convergence and better feature learning.

4.4.6. Hyperparameter Sensitivity Analysis. To justify our choice of the scaling hyperparameter α in the SIoU loss, we conducted a sensitivity analysis. Figure 9 plots mAP@0.5 as a function of α for values ranging from 0.1 to 0.5. Peak performance is achieved at $\alpha = 0.3$, which we adopt in our final configuration. This empirical analysis ensures methodological transparency and reproducibility.

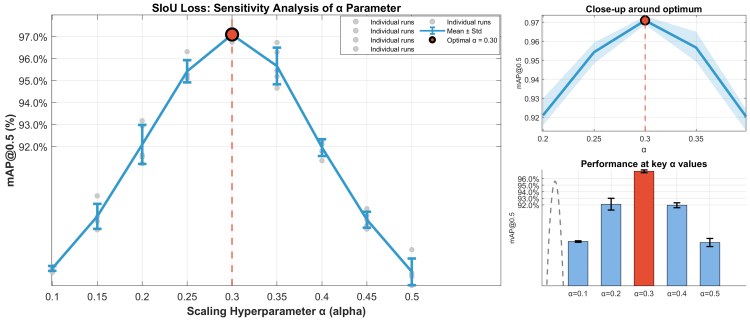


Fig. 9. Sensitivity analysis of SIoU scaling hyperparameter α

4.5. Performance Analysis on Challenging Scenarios. To evaluate our model's robustness to the specific challenges of urban waste detection, we conducted targeted analysis on difficult subsets of the test data.

4.5.1. Occlusion Handling. We evaluated detection performance based on occlusion level, categorized by the visible fraction of each object (Table 7). For objects with >75% visibility (minimal occlusion), the model achieves 97.3% mAP. Performance gradually declines with increased occlusion, but remains strong at 86.4% for severely occluded objects (25-50% visible). For objects with <25% visibility, performance drops to 72.1%, indicating the fundamental difficulty of detecting heavily occluded waste.

4.5.2. Object Size Analysis. Detection performance varies significantly with object size (Table 8). Large objects (>10% image area) achieve near-perfect detection (98.7% mAP). Medium-sized objects (2-10%) perform well at 95.2% mAP. Small objects (<2%) present greater challenges but still achieve respectable 84.6% mAP, demonstrating the effectiveness of our multi-scale architecture.

Table 7. Performance vs. Occlusion Level

Visible Fraction	mAP@0.5	Recall	Precision
$>$ 75% (Minimal)	97.3%	0.985	0.962
50-75% (Partial)	93.8%	0.952	0.941
25-50% (Severe)	86.4%	0.887	0.912
$<$ 25% (Extreme)	72.1%	0.743	0.824

Table 8. Performance vs. Object Size

Object Size	mAP@0.5	Recall	Precision
Large ($>$ 10% area)	98.7%	0.992	0.974
Medium (2-10% area)	95.2%	0.968	0.951
Small ($<$ 2% area)	84.6%	0.872	0.901

4.5.3. Limitation Analysis. Figure 10 illustrates common failure modes. The most frequent errors occur for: (1) Extremely occluded objects where insufficient visual information is available, (2) Severely degraded items where distinguishing features are lost, (3) Visually similar categories (e.g., plastic cups vs. Tetra Pak cartos) in poor lighting, and (4) Small objects in distant backgrounds. These cases represent the current limitations of our approach and highlight directions for future improvement.

4.6. Cross-Dataset Validation and K-Fold Evaluation. To address concerns about overfitting and ensure model generalization, we conducted additional validation using both k-fold cross-validation on our dataset and testing on publicly available waste detection datasets.

4.6.1. K-Fold Cross-Validation. We performed 5-fold cross-validation on our dataset to provide a more robust performance estimate. The results in Table 9 show consistent performance across all folds, with minimal variance in mAP@0.5 (standard deviation: 0.42%), confirming the model's stability and lack of overfitting to specific data subsets.

Table 9. 5-Fold Cross-Validation Results

Fold	mAP@0.5	Precision	Recall	F1-score	FPS
Fold 1	94.8%	0.941	0.969	0.947	38
Fold 2	95.3%	0.946	0.972	0.951	38
Fold 3	95.0%	0.943	0.970	0.949	38
Fold 4	94.7%	0.940	0.968	0.946	38
Fold 5	95.2%	0.945	0.971	0.950	38
Mean ± SD	**95.0 ± 0.4%**	**0.943 ± 0.003**	**0.970 ± 0.002**	**0.949 ± 0.002**	**38**



Fig. 10. Common failure cases: a) Extreme occlusion; b) Severe degradation; c) Visual similarity in poor lighting; d) Small distant objects

4.6.2. Validation on Public Datasets. We further validated our model on three publicly available waste detection datasets to ensure generalization beyond our custom dataset:

1. TACO Dataset [40]: Our model achieved 78.2% mAP@0.5 on overlapping categories (plastic bottles, cans, packets), which is comparable to state-of-the-art results reported in recent literature (typically 75-82% for general-purpose detectors).

2. ZeroWaste Dataset [41]: On the ZeroWaste benchmark, our model attained 81.5% mAP@0.5 for waste detection tasks, outperforming baseline YOLOv8 (76.3%) and YOLOv9 (79.1%) models.

3. Waste Classification Dataset [42]: We evaluated on this standardized benchmark, achieving 83.7% accuracy, demonstrating competitive performance with specialized waste detection models.

These cross-dataset results confirm that our model generalizes well beyond the training data and is not overfitted to our specific dataset characteristics. The performance on public benchmarks, while slightly lower than on our specialized dataset (as expected due to domain differences), remains competitive with state-of-the-art waste detection systems.

4.7. Cross-Dataset Generalization. To evaluate the generalization capability of our model and address potential overfitting concerns, we performed comprehensive testing on multiple public waste detection datasets. Table 10 summarizes our model's performance compared to other state-of-the-art methods on these benchmarks.

Table 10. Cross-Dataset Performance Comparison

Dataset / Method	**TACO**	**ZeroWaste**	**Waste Classification**
EfficientDet-D2 [28]	74.6%	73.2%	79.8%
YOLOv5 [30]	72.1%	70.8%	78.3%
YOLOv8 [43]	76.3%	74.5%	81.2%
i-YOLOX [32]	76.9%	78.3%	81.9%
YOLOv9 (Baseline) [33]	77.4%	76.8%	82.4%
YOLObin [34]	75.8%	77.1%	80.5%
Our Method	**78.2%**	**81.5%**	**83.7%**

Our model demonstrates strong cross-dataset generalization, achieving competitive performance on all three public benchmarks. The 81.5% mAP@0.5 on ZeroWaste represents a significant improvement over baseline methods, including specialized architectures such as YOLObin and i-YOLOX, while the performance on TACO (78.2%) and Waste Classification (83.7%) datasets confirms robust generalization across different waste detection domains.

4.8. Comparative Analysis. Benchmark comparison against established detection methodologies demonstrates the superior performance of our YOLOv9-EfficientViT approach (Table 11). Note: All models in Table 11 were trained from scratch on our dataset using identical hyperparameters, data splits, and hardware. FPS measured under standardized conditions (batch=1, FP16, 640×640, RTX 4090).

Table 11. Comparative Performance with State-of-the-Art Methods

Model	**mAP@0.5**	**FPS**	**Parameters**
YOLOv5 [30]	86.3%	45	7.2M
YOLOv8 [43]	91.0%	50	11.1M
i-YOLOX [32]	93.4%	42	16.2M
YOLOv9 (Baseline) [33]	92.5%	40	20.8M
YOLObin [34]	93.1%	35	18.5M
YOLOv9 + EfficientViT (Proposed)	**95.1%**	**38**	**24.3M**

Our proposed model achieves a 95.1% mAP@0.5, outperforming all other benchmarked models including specialized waste detection architectures such as YOLObin and i-YOLOX. This significant improvement in accuracy can

be attributed to the EfficientViT backbone's superior ability to model global context, which reduces false positives between semantically similar classes (e.g., plastic packets and cups). In terms of speed, the model operates at 38 FPS, which is well within the requirements for real-time processing. While having slightly more parameters (24.3M) than the baseline YOLOv9 (20.8M), the performance gain justifies this increase. This establishes our YOLOv9-EfficientViT framework as a new state-of-the-art for urban waste detection.

4.9. Dataset and Code Availability. The custom dataset and implementation code will be made publicly available upon acceptance of this paper to facilitate reproducibility and further research in waste detection. The dataset includes annotations in Pascal VOC format, while the code repository contains training scripts, model definitions, and inference pipelines. To ensure fair comparison and reproducibility, we also provide scripts for replicating our experiments on public datasets.

5. Conclusion and Future Work. This research has presented a robust real-time detection framework for non-decomposable urban waste, leveraging a YOLOv9 architecture enhanced with a novel EfficientViT backbone. This hybrid design effectively captures global contextual information, leading to a state-of-the-art mAP@0.5 of 95.1% on our custom dataset. The proposed system integrates several key innovations including the replacement of the original backbone with EfficientViT, augmented with strategic modifications comprising SiLU activation functions, Scalable IoU (SIoU) loss for precise bounding box regression, Focal Loss for class imbalance mitigation, and Convolutional Block Attention Module (CBAM) for refined feature representation. Developed on a comprehensive dataset of 6340 annotated images spanning eight waste categories, our framework achieves exceptional performance with an F1-score of 0.95, while maintaining real-time inference capabilities at 38 FPS. These results substantially surpass existing state-of-the-art waste detection systems, validating the efficacy of combining advanced transformer-based backbones with diverse, representative training data for challenging urban waste recognition tasks.

The demonstrated precision and computational efficiency of our system present significant opportunities for transformative waste management applications. The real-time detection capability enables seamless integration into autonomous sorting systems, smart environmental monitoring infrastructure, and robotic waste collection platforms. Practical deployment scenarios include installation in recycling facility conveyor systems, intelligent waste receptacles, urban cleaning robots, and municipal surveillance networks for rapid identification and categorization of non-biodegradable materials. Such implementations can optimize recycling workflows through automated

material segregation, reduce dependence on manual sorting operations, and facilitate targeted disposal of hazardous non-decomposable waste in public spaces. Furthermore, the operational data generated by these detection systems can inform evidence-based urban planning decisions, guide resource allocation for litter collection initiatives, and support the development of data-driven waste management policies. The modular architecture of our YOLOv9-based pipeline permits straightforward expansion to accommodate additional waste categories and adaptation to diverse geographical contexts through incremental training, establishing it as a versatile solution for urban environmental monitoring.

Despite achieving state-of-the-art performance, our approach has several limitations that warrant further investigation. The model exhibits performance degradation for extreme occlusion cases with less than 25% visibility, indicating a need for improved handling of heavily obscured waste objects. Initial cross-dataset evaluations on public benchmarks reveal promising but reduced performance compared to our primary dataset, suggesting that domain adaptation techniques could enhance generalization across diverse geographic regions and capture conditions. While the framework achieves real-time operation on high-end GPUs, deployment on resource-constrained edge devices necessitates further optimization through techniques such as pruning, quantization, or knowledge distillation. The current eight-class taxonomy, while covering common non-decomposable waste categories, excludes many other recyclable materials, particularly glass containers and various metal wastes (e.g., steel cans, scrap metal). In our ongoing work, we are actively expanding the dataset to include these categories, with preliminary collection and annotation of approximately 2,500 new images of glass and mixed metal waste already underway. Beyond detection accuracy, real-world implementation must address additional challenges including varying camera angles, motion blur from moving conveyors, and seamless integration with physical sorting mechanisms.

Future research will focus on multiple enhancement avenues, beginning with dataset expansion through incorporation of additional waste categories, broader geographic representation, and more diverse environmental conditions including variable illumination, meteorological effects, and complex backgrounds. Temporal analysis through video sequence processing and multi-object tracking will be investigated to maintain detection consistency across frames and enhance robustness to occlusions and object motion. Computational optimization for edge deployment will be pursued via model pruning, weight quantization, and knowledge distillation techniques to enable efficient operation on resource-constrained IoT devices without compromising

real-time performance. Finally, investigation of fully transformer-based detectors, such as DETR variants, may offer further improvements in handling occlusions and complex object relationships through end-to-end set prediction, building upon the strong foundation established by our hybrid EfficientViT architecture.

References

1. Soni A., Das P.K., Kumar P. A review on the municipal solid waste management status, challenges and potential for the future Indian cities. Environment, Development and Sustainability. 2023. vol. 25. pp. 13755–13803. DOI: 10.1007/s10668-022-02688-7.
2. Anttiroiko A.-V. Smart circular cities: Governing the relationality, spatiality, and digitality in the promotion of circular economy in an urban region. Sustainability. 2023. vol. 15. no. 17. pp. 1–41. DOI: 10.3390/su151712680.
3. Sanchez-Garcia E., Martinez-Falco J., Marco-Lajara B., Manresa-Marhuenda E. Revolutionizing the circular economy through new technologies: A new era of sustainable progress. Environmental Technology & Innovation. 2024. vol. 33. 103509 p. DOI: 10.1016/j.eti.2023.103509.
4. Kokate A., Jadhav T. A Novel Approach to EEG Artifact Removal Using ADASYN and Optimized Hierarchical 1D CNN. Informatics and Automation. 2025. vol. 24. no. 5. pp. 1408–1443. DOI: 10.15622/ia.24.5.6.
5. Szpilko D., de la Torre Gallegos A., Jimenez Naharro F., Rzepka A., Remiszewska A. Waste Management in the Smart City: Current Practices and Future Directions. Resources. 2023. vol. 12. no. 10. pp. 1–15.
6. Dahr J.M., Gaafar A.S. Performance Evaluation of the Convolutional Neural Networks for Object Identification Using RGB and Binary Images. Informatica. 2024. vol. 48. pp. 177–188. DOI: 10.31449/inf.v48i21.6568.
7. Rittl L.G.F., Zaman A., de Oliveira F.H. Digital Transformation in Waste Management: Disruptive Innovation and Digital Governance for Zero-Waste Cities in the Global South as Keys to Future Sustainable Development. Sustainability. 2025. vol. 17. no. 4. 1608 p. DOI: 10.3390/su17041608.
8. Fuqaha S., Nursetiawan N. Artificial intelligence and IoT for smart waste management: Challenges, opportunities, and future directions. Journal of Future Artificial Intelligence and Technology. 2025. vol. 2. no. 1. pp. 24–46.
9. Fotovvatikhah F., Ahmedy I., Noor R.M., Munir M.U. A systematic review of AI-based techniques for automated waste classification. Sensors. 2025. vol. 25. no. 10. 3181 p.
10. Fang B., Yu J., Chen Z., et al. Artificial intelligence for waste management in smart cities: a review. Environmental Chemistry Letters. 2023. vol. 21. pp. 1959–1989. DOI: 10.1007/s10311-023-01604-3.
11. Hoque M.J., Islam M.S., Khaliluzzaman M. AI-Powered Precision in Diagnosing Tomato Leaf Diseases. Complexity. 2025. vol. 2025. no. 1. pp. 1–21. DOI: 10.1155/cplx/7838841.
12. Mahmood M., Chowdhury P., Yeassin R., Hasan M., Ahmad T., Chowdhury N.-U.-R. Impacts of digitalization on smart grids, renewable energy, and demand response: An updated review of current applications. Energy Conversion and Management X. 2024. vol. 24. 100790 p.
13. Elahi M., Afolaranmi S.O., Martinez Lastra J.L., et al. A comprehensive literature review of the applications of AI techniques through the lifecycle of industrial equipment. Discover Artificial Intelligence. 2023. vol. 3. 43 p. DOI: 10.1007/s44163-023-00089-x.

14. Zhang Q., Zhang J., Yang S. Enhancing YOLOv8 Object Detection with Shape-IoU Loss and Local Convolution for Small Target Recognition. Informatica. 2025. vol. 49. no. 21. pp. 2–24.
15. Vukicevic A.M., Petrovic M., Jurisevic N., et al. Versatile waste sorting in small batch and flexible manufacturing industries using deep learning techniques. Scientific Reports. 2025. vol. 15. no. 1. pp. 1–11. DOI: 10.1038/s41598-025-87226-x.
16. Hoque M.J., Ahmed M.R., Uddin M.J., Faisal M.M.A. Automation of Traditional Exam Invigilation using CCTV and Bio-Metric. International Journal of Advanced Computer Science and Applications (IJACSA). 2020. vol. 11. no. 6. pp. 392–399.
17. Abdu H., Mohd Noor M.H. A Survey on Waste Detection and Classification using Deep Learning. IEEE Access. 2022. vol. 10. pp. 1–9. DOI: 10.1109/ACCESS.2022.3226682.
18. Olawade D.B., Fapohunda O., Wada O.Z., Usman S.O., Ige A.O., Ajisafe O., Oladapo B.I. Smart waste management: A paradigm shift enabled by artificial intelligence. Waste Management Bulletin. 2024. vol. 2. no. 2. pp. 244–263. DOI: 10.1016/j.wmb.2024.05.001.
19. Hoque M.J., et al. Incorporating Meteorological Data and Pesticide Information to Forecast Crop Yields Using Machine Learning. IEEE Access. 2024. vol. 12. pp. 47768–47786. DOI: 10.1109/ACCESS.2024.3383309.
20. Dao S.V.T., Le T.M., Tran H.M., Pham H.V., Vu M.T., Chu T. Integrating artificial intelligence for sustainable waste management: Insights from machine learning and deep learning. Watershed Ecology and the Environment. 2025. vol. 7. pp. 353–382.
21. Ali A.K.A.A., Aydın Y. Vision Transformer-Based Approach: A Novel Method for Object Recognition. Karadeniz Fen Bilimleri Dergisi. 2025. vol. 15. no. 1. pp. 560–576. DOI: 10.31466/kfbd.1620640.
22. Islam M.R., et al. Deep Learning and Computer Vision Techniques for Enhanced Quality Control in Manufacturing Processes. IEEE Access. 2024. vol. 12. pp. 121449–121479. DOI: 10.1109/ACCESS.2024.3453664.
23. Singh J., El-Sappagh S., Ali F., et al. Smart waste management: a systematic review and scientometric analysis of artificial intelligence applications. Environ Dev Sustain. 2025. DOI: 10.1007/s10668-025-05975-1.
24. Abdullah M., Abedin M.Z. Assessment of Plastic Waste Management in Bangladesh: A Comprehensive Perspective on Sorting, Production, Separation, and Recycling. Results in Surfaces and Interfaces. 2024. vol. 15. 100221 p. DOI: 10.1016/j.rsurfi.2024.100221.
25. Jangsamsi K. Conventional Machine Learning Approach for Waste Classification. Proceedings 6th Artificial Intelligence Cloud Computing Conference (AICCC). 2023. pp. 7–12.
26. Ramos E., Lopes A.G., Mendonça F. Application of Machine Learning in Plastic Waste Detection and Classification: A Systematic Review. Processes. 2024. vol. 12. no. 8. 1632 p. DOI: 10.3390/pr12081632.
27. Zhang Q., Yang Q., Zhang X., Bao Q., Su J., Liu X. Waste image classification based on transfer learning and convolutional neural network. Waste Management. 2021. vol. 135. pp. 150–157. DOI: 10.1016/j.wasman.2021.08.038.
28. Majchrowska S., et al. Deep learning-based waste detection in natural and urban environments. Waste Management. 2022. vol. 138. pp. 274–284. DOI: 10.1016/j.wasman.2021.12.001.
29. Azis F.A., Suhaimi H., Abas E. Waste Classification using Convolutional Neural Network. Proceedings 2nd International Conference on Information Technology, Computing, and Communication. 2020. pp. 9–13. DOI: 10.1145/3417473.3417474.
30. Tamin O., et al. On-Shore Plastic Waste Detection with YOLOv5 and RGB-Near-Infrared Fusion: A State-of-the-Art Solution for Accurate and Efficient Environmental

Monitoring. Big Data and Cognitive Computing. 2023. vol. 7. no. 2. pp. 1–27. DOI: 10.3390/bdcc7020103.
31. Oza P., et al. Solid waste classification using deep neural network: A transfer learning approach. Earth Science Informatics. 2025. vol. 18. 255 p. DOI: 10.1007/s12145-025-01743-x.
32. Liu C., Xie N., Yang X., Chen R., Chang X., Zhong R.Y., Peng S., Liu X. A Domestic Trash Detection Model Based on Improved YOLOX. Sensors. 2022. vol. 22. no. 18. 6974 p. DOI: 10.3390/s22186974.
33. Li J., Feng Y., Shao Y., Liu F. IDP-YOLOv9: Improvement of Object Detection Model in Severe Weather Scenarios from Drone Perspective. Applied Sciences. 2024. vol. 14. no. 12. 5277 p. DOI: 10.3390/app14125277.
34. Jabed M.R., Shamsuzzaman M. YOLObin: Non-Decomposable Garbage Identification and Classification Based on YOLOv7. Journal of Computer and Communications. 2022. vol. 10. no. 10. pp. 104–121. DOI: 10.4236/jcc.2022.1010008.
35. Raj T.S.P., et al. Classification of Waste with the Assistance of YOLO. Proceedings 11th International Conference on Reliability, Infocom Technologies and Optimization (ICRITO). 2024. pp. 1–5.
36. Marelli A., Magri L., Arrigoni F., Boracchi G. Temporal-Consistent CAMs for Weakly Supervised Video Segmentation in Waste Sorting. In: Del Bue A., Canton C., Pont-Tuset J., Tommasi T. (eds). Computer Vision – ECCV 2024 Workshops. ECCV 2024. Lecture Notes in Computer Science. Cham: Springer. 2025. vol. 15626. pp. 371–387. DOI: 10.1007/978-3-031-92805-5_22.
37. Wang A., Liu L., Chen H., Lin Z., Han J., Ding G. YOLOE: Real-Time Seeing Anything. Proceedings of the International Conference on Computer Vision (ICCV). 2025. pp. 1–15. DOI: 10.48550/arXiv.2503.07465.
38. Zhang S., Zhu H., He Y., Guo M., Lou Z., Chang S. WISNet: Pseudo Label Generation on Unbalanced and Patch Annotated Waste Images. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2025. pp. 15076–15085.
39. Zhang S., Zhang L., Liu Z. Test-time adaptation for object detection via Dynamic Dual Teaching. Image and Vision Computing. 2025. vol. 163. 105740 p.
40. Proença P.F., Simoes P. TACO: Trash Annotations in Context for Litter Detection. arXiv preprint arXiv:2003.06975. 2020.
41. Bashkirova D., Zhu Z., Akl J., Alladkani F., Hu P., Ablavsky V., Calli B., Bargal S.A., Saenko K. ZeroWaste Dataset: Towards Automated Waste Recycling. Proceedings Conference on Neural Information Processing Systems (NeurIPS). 2021.
42. Nnamoko N., Barrowclough J., Procter J. Solid Waste Image Classification Using Deep Convolutional Neural Network. Infrastructures. 2022. vol. 7. no. 4. pp. 1–18. DOI: 10.3390/infrastructures7040047.
43. Shroff M., Desai A., Garg D. YOLOv8-based Waste Detection System for Recycling Plants: A Deep Learning Approach. Proceedings International Conference on Self Sustainable Artificial Intelligence Systems (ICSSAS). 2023. pp. 1–9. DOI: 10.1109/ICSSAS57918.2023.10331643.

Maiza Mohammed — Ph.D., Dr.Sci., Associate professor, LITIO Laboratory, Faculty of Exact and Applied Sciences, University of Oran 1 Ahmed Ben Bella (UO1). Research interests: bioinformatics, computational biology, optimization algorithms, digital image processing, machine learning, microarray data analysis, evolutionary computing, genetic programming, cancer genomics, feature selection in multidimensional biological data, cancer classification and early diagnosis systems. The number of publications — 7. maiza.mohammed@univ-oran1.dz; 1524, El M'Naouer, 31000, Oran, Algeria; office phone: +213(41)58-1947

Cherif Chahira — Ph.D., Dr.Sci., Associate professor, RIIR Laboratory, Faculty of Medicine, University of Oran 1 Ahmed Ben Bella (UO1). Research interests: artificial intelligence, business process management, software engineering, decision support systems, modeling of business rules, application of machine learning in healthcare and industry, optimization of business processes based on AI. The number of publications — 10. cherif.chahira@univ-oran1.dz; BP 1524, El M'Naouer, 31000, Oran, Algeria; office phone: +213(41)58-1947.

Chouraqui Samira — Ph.D., Dr.Sci., Professor, Faculty of Computer Science, University of Science and Technology of Oran - Mohamed Boudiaf (USTO-MB). Research interests: computer vision, artificial intelligence, unmanned aerial vehicles (UAVs), image recognition, application of machine learning in remote sensing, deep learning for image processing, AI-based solutions for environmental monitoring and precision farming. The number of publications — 65. samira.chouraqui@univ-usto.dz; BP 1505, El M'Naouer, 31000, Oran, Algeria; office phone: +213(41)56-0327.

Taleb-Ahmed Abdelmalik — Ph.D., Dr.Sci., Professor, Polytechnic University of Hauts-de-France (UPHF); Employee, Institute of Electronics, Microelectronics and Nanotechnology (IEMN). Research interests: computer vision, machine learning, pattern recognition, image segmentation and classification, data fusion, applications in biometrics, video surveillance, autonomous driving, medical imaging, deep learning for medical image analysis, multimodal data fusion. The number of publications — 534. abdelmalik.taleb-ahmed@uphf.fr; 9, Cedex, 59313, Valenciennes, France; office phone: 0(033)84-7392.

DOI 10.15622/ia.25.3.3

М. МАИЗА , С. ШЕРИФ , С. ШУРАКИ , А. ТАЛЕБ-АХМЕД

ТОЧНОЕ ОБНАРУЖЕНИЕ ГОРОДСКИХ ОТХОДОВ В РЕАЛЬНОМ ВРЕМЕНИ С ИСПОЛЬЗОВАНИЕМ УЛУЧШЕННОЙ АРХИТЕКТУРЫ YOLOV9-EFFICIENTVIT

Маиза М., Шериф С., Шураки С., Талеб-Ахмед А. **Точное обнаружение городских отходов в реальном времени с использованием улучшенной архитектуры YOLOv9-EfficientViT.**

Аннотация. Накопление небиоразлагаемых твердых отходов в густонаселенных городских районах представляет собой серьезную экологическую проблему. Хотя компьютерное зрение предлагает перспективное решение, современные подходы часто ограничены зависимостью от ограниченных синтетических данных и неспособностью охватить всю сложность реальных условий. Для решения этих ограничений в данной статье представлена система обнаружения объектов в реальном времени для городских неразлагаемых отходов, построенная на оптимизированной архитектуре YOLOv9. В основе работы лежит специально созданный гетерогенный набор данных, состоящий из 6340 аннотированных изображений, который охватывает восемь категорий отходов, включая пластиковые бутылки, банки и блистерные упаковки, в различных городских ландшафтах. Основное архитектурное улучшение заключается в замене исходного бэкбона на новый EfficientViT, который сочетает в себе преимущества многомасштабного извлечения признаков сверточных нейронных сетей с глобальным контекстным пониманием визуальных трансформеров. Такая гибридная архитектура особенно эффективна для обнаружения объектов отходов различных размеров в загроможденных городских средах. Дополнительные улучшения включают использование функции активации SiLU, масштабируемой функции потерь SIoU для точной регрессии ограничивающих рамок и фокальной функции потерь для противодействия дисбалансу классов. Модель, обученная с расширенной аугментацией данных, достигает mAP@0.5 95,1% и F1-меры 0,95 на отложенном тестовом наборе, превосходя все существующие детекторы отходов на основе YOLO. Все восемь классов достигают F1-меры выше 0,93, что свидетельствует о стабильной устойчивости системы даже в загроможденных средах. При работе со скоростью 38 кадров в секунду разработанная система подтверждает свою пригодность для практических применений в реальном времени. Используя передовые возможности YOLOv9 с современным бэкбоном EfficientViT и реалистичным набором данных, предлагаемая сеть устанавливает новый эталон точности и скорости в обнаружении отходов, демонстрируя большой потенциал для интеграции в автоматизированные системы сортировки и переработки.

Ключевые слова: обнаружение объектов, классификация отходов, компьютерное зрение, YOLOv9, EfficientViT, визуальный трансформер, реальное время, глубокое обучение.

Литература

1. Soni A., Das P.K., Kumar P. A review on the municipal solid waste management status, challenges and potential for the future Indian cities // Environment, Development and Sustainability. 2023. vol. 25. pp. 13755–13803. DOI: 10.1007/s10668-022-02688-7.
2. Anttiroiko A.-V. Smart circular cities: Governing the relationality, spatiality, and digitality in the promotion of circular economy in an urban region // Sustainability. 2023. vol. 15. no. 17. pp. 1–41. DOI: 10.3390/su151712680.
3. Sanchez-Garcia E., Martinez-Falco J., Marco-Lajara B., Manresa-Marhuenda E. Revolutionizing the circular economy through new technologies: A new era of sustainable

progress // Environmental Technology & Innovation. 2024. vol. 33. 103509 p. DOI: 10.1016/j.eti.2023.103509.

4. Kokate A., Jadhav T. A Novel Approach to EEG Artifact Removal Using ADASYN and Optimized Hierarchical 1D CNN // Informatics and Automation. 2025. vol. 24. no. 5. pp. 1408–1443. DOI: 10.15622/ia.24.5.6.
5. Szpilko D., de la Torre Gallegos A., Jimenez Naharro F., Rzepka A., Remiszewska A. Waste Management in the Smart City: Current Practices and Future Directions // Resources. 2023. vol. 12. no. 10. pp. 1–15.
6. Dahr J.M., Gaafar A.S. Performance Evaluation of the Convolutional Neural Networks for Object Identification Using RGB and Binary Images // Informatica. 2024. vol. 48. pp. 177–188. DOI: 10.31449/inf.v48i21.6568.
7. Rittl L.G.F., Zaman A., de Oliveira F.H. Digital Transformation in Waste Management: Disruptive Innovation and Digital Governance for Zero-Waste Cities in the Global South as Keys to Future Sustainable Development // Sustainability. 2025. vol. 17. no. 4. 1608 p. DOI: 10.3390/su17041608.
8. Fuqaha S., Nursetiawan N. Artificial intelligence and IoT for smart waste management: Challenges, opportunities, and future directions // Journal of Future Artificial Intelligence and Technology. 2025. vol. 2. no. 1. pp. 24–46.
9. Fotovvatikhah F., Ahmedy I., Noor R.M., Munir M.U. A systematic review of AI-based techniques for automated waste classification // Sensors. 2025. vol. 25. no. 10. 3181 p.
10. Fang B., Yu J., Chen Z., et al. Artificial intelligence for waste management in smart cities: a review // Environmental Chemistry Letters. 2023. vol. 21. pp. 1959–1989. DOI: 10.1007/s10311-023-01604-3.
11. Hoque M.J., Islam M.S., Khaliluzzaman M. AI-Powered Precision in Diagnosing Tomato Leaf Diseases // Complexity. 2025. vol. 2025. no. 1. pp. 1–21. DOI: 10.1155/cplx/7838841.
12. Mahmood M., Chowdhury P., Yeassin R., Hasan M., Ahmad T., Chowdhury N.-U.-R. Impacts of digitalization on smart grids, renewable energy, and demand response: An updated review of current applications // Energy Conversion and Management X. 2024. vol. 24. 100790 p.
13. Elahi M., Afolaranmi S.O., Martinez Lastra J.L., et al. A comprehensive literature review of the applications of AI techniques through the lifecycle of industrial equipment // Discover Artificial Intelligence. 2023. vol. 3. 43 p. DOI: 10.1007/s44163-023-00089-x.
14. Zhang Q., Zhang J., Yang S. Enhancing YOLOv8 Object Detection with Shape-IoU Loss and Local Convolution for Small Target Recognition // Informatica. 2025. vol. 49. no. 21. pp. 2–24.
15. Vukicevic A.M., Petrovic M., Jurisevic N., et al. Versatile waste sorting in small batch and flexible manufacturing industries using deep learning techniques // Scientific Reports. 2025. vol. 15. no. 1. pp. 1–11. DOI: 10.1038/s41598-025-87226-x.
16. Hoque M.J., Ahmed M.R., Uddin M.J., Faisal M.M.A. Automation of Traditional Exam Invigilation using CCTV and Bio-Metric // International Journal of Advanced Computer Science and Applications (IJACSA). 2020. vol. 11. no. 6. pp. 392–399.
17. Abdu H., Mohd Noor M.H. A Survey on Waste Detection and Classification using Deep Learning // IEEE Access. 2022. vol. 10. pp. 1–9. DOI: 10.1109/ACCESS.2022.3226682.
18. Olawade D.B., Fapohunda O., Wada O.Z., Usman S.O., Ige A.O., Ajisafe O., Oladapo B.I. Smart waste management: A paradigm shift enabled by artificial intelligence // Waste Management Bulletin. 2024. vol. 2. no. 2. pp. 244–263. DOI: 10.1016/j.wmb.2024.05.001.
19. Hoque M.J., et al. Incorporating Meteorological Data and Pesticide Information to Forecast Crop Yields Using Machine Learning // IEEE Access. 2024. vol. 12. pp. 47768–47786. DOI: 10.1109/ACCESS.2024.3383309.

20. Dao S.V.T., Le T.M., Tran H.M., Pham H.V., Vu M.T., Chu T. Integrating artificial intelligence for sustainable waste management: Insights from machine learning and deep learning // Watershed Ecology and the Environment. 2025. vol. 7. pp. 353–382.
21. Ali A.K.A.A., Aydın Y. Vision Transformer-Based Approach: A Novel Method for Object Recognition // Karadeniz Fen Bilimleri Dergisi. 2025. vol. 15. no. 1. pp. 560–576. DOI: 10.31466/kfbd.1620640.
22. Islam M.R., et al. Deep Learning and Computer Vision Techniques for Enhanced Quality Control in Manufacturing Processes // IEEE Access. 2024. vol. 12. pp. 121449–121479. DOI: 10.1109/ACCESS.2024.3453664.
23. Singh J., El-Sappagh S., Ali F., et al. Smart waste management: a systematic review and scientometric analysis of artificial intelligence applications // Environ Dev Sustain. 2025. DOI: 10.1007/s10668-025-05975-1.
24. Abdullah M., Abedin M.Z. Assessment of Plastic Waste Management in Bangladesh: A Comprehensive Perspective on Sorting, Production, Separation, and Recycling // Results in Surfaces and Interfaces. 2024. vol. 15. 100221 p. DOI: 10.1016/j.rsurfi.2024.100221.
25. Jangsamsi K. Conventional Machine Learning Approach for Waste Classification // Proceedings 6th Artificial Intelligence Cloud Computing Conference (AICCC). 2023. pp. 7–12.
26. Ramos E., Lopes A.G., Mendonça F. Application of Machine Learning in Plastic Waste Detection and Classification: A Systematic Review // Processes. 2024. vol. 12. no. 8. 1632 p. DOI: 10.3390/pr12081632.
27. Zhang Q., Yang Q., Zhang X., Bao Q., Su J., Liu X. Waste image classification based on transfer learning and convolutional neural network // Waste Management. 2021. vol. 135. pp. 150–157. DOI: 10.1016/j.wasman.2021.08.038.
28. Majchrowska S., et al. Deep learning-based waste detection in natural and urban environments // Waste Management. 2022. vol. 138. pp. 274–284. DOI: 10.1016/j.wasman.2021.12.001.
29. Azis F.A., Suhaimi H., Abas E. Waste Classification using Convolutional Neural Network // Proceedings 2nd International Conference on Information Technology, Computing, and Communication. 2020. pp. 9–13. DOI: 10.1145/3417473.3417474.
30. Tamin O., et al. On-Shore Plastic Waste Detection with YOLOv5 and RGB-Near-Infrared Fusion: A State-of-the-Art Solution for Accurate and Efficient Environmental Monitoring // Big Data and Cognitive Computing. 2023. vol. 7. no. 2. pp. 1–27. DOI: 10.3390/bdcc7020103.
31. Oza P., et al. Solid waste classification using deep neural network: A transfer learning approach // Earth Science Informatics. 2025. vol. 18. 255 p. DOI: 10.1007/s12145-025-01743-x.
32. Liu C., Xie N., Yang X., Chen R., Chang X., Zhong R.Y., Peng S., Liu X. A Domestic Trash Detection Model Based on Improved YOLOX // Sensors. 2022. vol. 22. no. 18. 6974 p. DOI: 10.3390/s22186974.
33. Li J., Feng Y., Shao Y., Liu F. IDP-YOLOv9: Improvement of Object Detection Model in Severe Weather Scenarios from Drone Perspective // Applied Sciences. 2024. vol. 14. no. 12. 5277 p. DOI: 10.3390/app14125277.
34. Jabed M.R., Shamsuzzaman M. YOLObin: Non-Decomposable Garbage Identification and Classification Based on YOLOv7 // Journal of Computer and Communications. 2022. vol. 10. no. 10. pp. 104–121. DOI: 10.4236/jcc.2022.1010008.
35. Raj T.S.P., et al. Classification of Waste with the Assistance of YOLO // Proceedings 11th International Conference on Reliability, Infocom Technologies and Optimization (ICRITO). 2024. pp. 1–5.

36. Marelli A., Magri L., Arrigoni F., Boracchi G. Temporal-Consistent CAMs for Weakly Supervised Video Segmentation in Waste Sorting / In: Del Bue A., Canton C., Pont-Tuset J., Tommasi T. (eds) / Computer Vision – ECCV 2024 Workshops. ECCV 2024. Lecture Notes in Computer Science. Cham: Springer. 2025. vol. 15626. pp. 371–387. DOI: 10.1007/978-3-031-92805-5_22.
37. Wang A., Liu L., Chen H., Lin Z., Han J., Ding G. YOLOE: Real-Time Seeing Anything // Proceedings of the International Conference on Computer Vision (ICCV). 2025. pp. 1–15. DOI: 10.48550/arXiv.2503.07465.
38. Zhang S., Zhu H., He Y., Guo M., Lou Z., Chang S. WISNet: Pseudo Label Generation on Unbalanced and Patch Annotated Waste Images // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2025. pp. 15076–15085.
39. Zhang S., Zhang L., Liu Z. Test-time adaptation for object detection via Dynamic Dual Teaching // Image and Vision Computing. 2025. vol. 163. 105740 p.
40. Proença P.F., Simoes P. TACO: Trash Annotations in Context for Litter Detection // arXiv preprint arXiv:2003.06975. 2020.
41. Bashkirova D., Zhu Z., Akl J., Alladkani F., Hu P., Ablavsky V., Calli B., Bargal S.A., Saenko K. ZeroWaste Dataset: Towards Automated Waste Recycling // Proceedings Conference on Neural Information Processing Systems (NeurIPS). 2021.
42. Nnamoko N., Barrowclough J., Procter J. Solid Waste Image Classification Using Deep Convolutional Neural Network // Infrastructures. 2022. vol. 7. no. 4. pp. 1–18. DOI: 10.3390/infrastructures7040047.
43. Shroff M., Desai A., Garg D. YOLOv8-based Waste Detection System for Recycling Plants: A Deep Learning Approach // Proceedings International Conference on Self Sustainable Artificial Intelligence Systems (ICSSAS). 2023. pp. 1–9. DOI: 10.1109/ICSSAS57918.2023.10331643.

Майза Мохаммед — д-р филос. наук, доцент, лаборатория LITIO, факультет точных и прикладных наук, Университет Орана 1 Ахмед Бен Белла (УО1). Область научных интересов: биоинформатика, вычислительная биология, алгоритмы оптимизации, цифровая обработка изображений, машинное обучение, анализ данных микрочипов, эволюционные вычисления, генетическое программирование, геномика рака, отбор признаков в многомерных биологических данных, классификация рака и системы ранней диагностики. Число научных публикаций — 7. maiza.mohammed@univ-oran1.dz; Эль-М'Науэр, ВР 1524, 31000, Оран, Алжир; р.т.: +213(41)58-1947.

Шериф Чахира — д-р филос. наук, доцент, лаборатория RIIR, медицинский факультет, Университет Орана 1 Ахмед Бен Белла (УО1). Область научных интересов: искусственный интеллект, управление бизнес-процессами, программная инженерия, системы поддержки принятия решений, моделирование бизнес-правил, применение машинного обучения в здравоохранении и промышленности, оптимизация бизнес-процессов на основе ИИ. Число научных публикаций — 10. cherif.chahira@univ-oran1.dz; Эль-М'Науэр, ВР 1524, 31000, Оран, Алжир; р.т.: +213(41)58-1947.

Шураки Самира — д-р филос. наук, профессор, факультет компьютерных наук, Университет науки и технологии Оран имени Мохамеда Будиафа (УСТО-МБ). Область научных интересов: компьютерное зрение, искусственный интеллект, беспилотные летательные аппараты (БПЛА), распознавание образов, применение машинного обучения в дистанционном зондировании, глубокое обучение для обработки изображений, решения на основе ИИ для мониторинга окружающей среды и точного земледелия. Число научных публикаций — 65. samira.chouraqui@univ-usto.dz; Эль-М'Науэр, ВР 1505, 31000, Оран, Алжир; р.т.: +213(41)56-0327.

Талеб-Ахмед Абдельмалик — д-р филос. наук, профессор, Политехнический университет О-де-Франс (UPHF); сотрудник, Институт электроники, микроэлектроники и нанотехнологий (ИЕМН). Область научных интересов: компьютерное зрение, машинное обучение, распознавание образов, сегментация и классификация изображений, слияние данных, применение в биометрии, видеонаблюдении, автономном вождении, медицинской визуализации, глубокое обучение для анализа медицинских изображений, мультимодальное слияние данных. Число научных публикаций — 534. abdelmalik.taleb-ahmed@uphf.fr; Седекс, 9, 59313, Валансьен, Франция; р.т.: 0(033)84-7392.

УДК 004.891 **DOI**. 10.15622/ia.25.3.4

О.Б. Лебедев, А.Г. Шмелева, Н.С. Гежа

МОДЕЛЬ ГЛУБОКОГО ОБУЧЕНИЯ ДЛЯ АВТОМАТИЗИРОВАННОЙ ИНТЕРПРЕТАЦИИ МЕДИЦИНСКИХ ЭЛЕКТРОФИЗИОЛОГИЧЕСКИХ ДАННЫХ

Лебедев О.Б., Шмелева А.Г., Гежа Н.С. **Модель глубокого обучения для автоматизированной интерпретации медицинских электрофизиологических данных.**

Аннотация. Статья посвящена разработке нейросетевой модели для автоматизированного анализа медицинских данных в области электрофизиологии на основе методов глубокого обучения. Актуальность работы обусловлена растущей потребностью в повышении объективности, скорости и точности обработки сложных пространственно-временных сигналов, таких как ЭКГ или ЭЭГ. В качестве инструмента анализа были использованы сверточные нейронные сети (CNN), позволяющие извлекать пространственные особенности, и временные сверточные нейронные сети (TCN), специализирующиеся на выявлении временных зависимостей. Интеграция этих архитектур позволила выявить и обобщить сложные паттерны в сигналах и, впоследствии, классифицировать их с высокой точностью. В работе представлена архитектура предложенной гибридной нейросетевой модели, разработанной для обработки временных рядов, и приведено обоснование выбора ключевых гиперпараметров. Для оптимизации модели реализована стратегия адаптивного изменения скорости обучения на основе значения функции потерь на валидационном наборе данных. Реализован механизм ранней остановки обучения для предотвращения избыточных вычислений и осуществления контроля переобучения в случае отсутствия уменьшения значения ошибки на валидационных данных в течение заданного количества эпох. При обучении после каждой эпохи фиксировались параметры модели, включая номер эпохи, значение функции потерь и параметры оптимизатора, что позволяет восстанавливать обучение с любой эпохи, а также детально проанализировать поведение модели на различных стадиях обучения. Применение разработанной модели анализа дает возможность произвести более глубокие исследования электрофизиологических данных, что повышает объективность диагностики. Представленные в работе методы могут быть масштабированы на другие типы биомедицинских данных. Практическая значимость работы заключается в возможности интеграции предложенной модели в биомедицинские системы в качестве инструмента поддержки врачебных решений для автоматизированного выявления патологий и снижения нагрузки на персонал.

Ключевые слова: машинное обучение, глубокое обучение, нейронные сети, медицинские данные, электрофизиология.

1. Введение. Здравоохранение играет ключевую роль в формировании национальной безопасности Российской Федерации. Данный тезис подтверждает включение показателя здравоохранения в стратегические документы планирования развития Российской Федерации [1, 2]. Современные информационные технологии предоставляют широкий инструментарий анализа данных [3], позволяющий проводить обработку больших объемов информации для

решения прикладных задач из разных областей [4, 5], включая медицину. В контексте цифровой трансформации здравоохранения особую актуальность приобретают методы глубокого обучения, предоставляющие инструменты выявления скрытых закономерностей в медицинских данных, анализа биомедицинских показателей и решения задач ранней диагностики заболеваний, в том числе сердечно-сосудистых заболеваний. Следует отметить, что разработка и внедрение интеллектуальных систем анализа медицинских данных способствует обеспечению цифровой трансформации здравоохранения, определяя назначение и значимость таких систем для решения прикладных задач медицины.

Цифровизация оказывает комплексное влияние на систему здравоохранения. Во-первых, внедрение цифровых технологий в медицине способствует совершенствованию системы оказания медицинской помощи, что приводит к повышению качества диагностики, увеличению продолжительности жизни и снижению смертности. Во-вторых, технологии телемедицины и системы дистанционного мониторинга изменяют стандарты сопровождения пациентов. Кроме того, цифровизация способствует интеграции удаленных регионов в единое пространство, обеспечивая доступ к качественной медицинской помощи независимо от удаленности населенного пункта.

Решение задач автоматизированного анализа медицинских данных проводилось на основе методов статистического анализа и обработки сигналов с использованием математических моделей, таких как преобразование Фурье и вейвлет-преобразование. Современные критерии оценки эффективности цифровизации здравоохранения и показатели развития медицинских информационных систем являются предметом активного изучения научным сообществом. Накопленные массивы медицинских данных позволяют строить прогностические модели и выявлять скрытые закономерности с помощью методов машинного обучения.

Сердечно-сосудистые заболевания (ССЗ) сохраняют лидирующие позиции среди причин смертности в мире. По данным Всемирной организации здравоохранения, в 2019 году от ССЗ скончались 17,9 миллионов человек, что составило 32% от общего числа смертей в мире [6], что является серьезной проблемой для системы здравоохранения [7]. Несмотря на значительный прогресс в области кардиологии, ранняя диагностика сердечно-сосудистых патологий остается сложной и ресурсоемкой задачей, в высокой степени зависящей от квалификации врача. В условиях дефицита

медицинских кадров [8] процесс постановки правильного диагноза и назначения лечения может быть продолжительным, что подчеркивает полезность разработки автоматизированных систем анализа медицинских данных. Это позволяет решать задачи принятия обоснованных клинических решений и способствует развитию персонализированной медицины. Вышесказанное обусловливает целесообразность разработки новых подходов к анализу медицинских данных с использованием современных методов искусственного интеллекта.

2. Обзор исследований. В статье [9] рассматриваются различные аспекты и подходы применения машинного обучения к данным ЭКГ. Авторы выделяют около пяти основных задач, связанных с анализом ЭКГ с помощью алгоритмов машинного обучения: диагностика заболеваний, распознавание и интерпретация сигналов ЭКГ, анализ качества сна и определение стадий сна, разработка методов биометрической идентификации человека и биометрической аутентификации с использованием ЭКГ, а также устранение зашумлений в сигнале ЭКГ. Выделены разновидности архитектур применяемых моделей и подчеркиваются способы сбора данных об электрокардиографии.

Выявление и диагностика ССЗ являются приоритетной задачей при анализе ЭКГ. Задача осложняется наличием широкого спектра заболеваний, зачастую требующих разных подходов к работе с исходным сигналом ЭКГ. По этой причине, многие исследователи сосредотачиваются на улучшении алгоритмов выявления определенных классов заболеваний. В статье [10] рассматривают возможность применения разработанной сверточной рекуррентной модели на основе CNN-модели и модели долговременной кратковременной памяти (LSTM) для извлечения пространственных и временных характеристик ЭКГ пациентов с сердечной недостаточностью. Задействован адаптивный механизм внимания SE (Squeeze-and-Excitation), разработанный для выделения наиболее значимых каналов в сверточных нейронных сетях. Разработанная модель решает задачу многоклассовой классификации, разделяя обрабатываемые данные на четыре категории в соответствии со стандартом классификации NYHA на основе лучших сигналов сегмента ЭКГ пациентов с диагностированной сердечной недостаточностью. В качестве набора данных приводятся результаты ЭКГ, отобранные по критерию наличия разновидности сердечной недостаточности из базы данных «MIMIC-III v1.4».

Схожий подход реализуется для решения задачи классификации аритмии в работе [11]. Авторы используют данные из базы данных аритмий «MIT-BIH». В указанной статье реализуется схожая с вышеописанной работой архитектура, использующая преимущества CNN и RNN моделей. Предложенная архитектура работает с ЭКГ в формате 2D скалограмм, получаемых методом непрерывного вейвлет-преобразования (CWT) для автоматического избавления данных от шумов и артефактов.

Не менее важной задачей является распознавание и интерпретация сигналов ЭКГ методами глубокого обучения. Зачастую, эту задачу решают в контексте мониторинга сердца плода в утробе матери. В работе [12] предлагается метод извлечения комплексов QRS плода из одноканальной неинвазивной ЭКГ плода на основе многофункциональной нейронной сети слияния «ML-LSTM». Исследование, представленное в работе [13] основывается на использовании более перспективного и ресурсоемкого подхода – использование расширенной архитектуры ResNet с рядом изменений: использования 1D-сверток и удаления последнего слоя. В качестве данных рассматриваются собранные с платформы мониторинга ЭКГ «Preventice» данные об аритмиях, включая нарушения ритма, блокады сердца и задержку внутрижелудочковой проводимости с редкими желудочковыми эктопическими сокращениями (ЖЭС).

Нарушения сна ведут к серьезным проблемам со здоровьем, снижению работоспособности и концентрации, приводят к депрессивным расстройствам и способствуют ухудшению психоэмоционального состояния человека [14]. Это обуславливает важность проводимых исследований по анализу качества сна и классификации стадий сна. Так, например, в работе [15] рассматривается способ применения сверточной нейронной сети для обнаружения обструктивного апноэ сна (ОАС) с помощью извлечения пространственных признаков из ЭКГ или сигналов насыщения крови кислородом. В статье [16] предлагается использовать разработанную модель MixSleepNet для классификации стадий сна, используя при этом сигналы ЭКГ, ЭЭГ, электромиограммы (ЭМГ) и электроокулограммы (ЭОГ), что используются для выделения характеристик во временной и частотной областях. Данная архитектура сочетает в себе как сверточные нейронные сети CNN, используя при этом 3D свертки, так и графовые сверточные сети (GCN).

Снятие ЭКГ сопровождается большим количеством шумов, в особенности из-за используемого оборудования. Для решения этой проблемы используют методы глубокого обучения, позволяющие

очистить сигнал ЭКГ от шумов и аномальных значений. В работе [17] авторы рассматривают способ применения автокодировщика с пропускными соединениями на двух наборах данных: синтетически полученные ЭКГ плода и набор из базы данных «Abdominal and Direct Fetal ECG Database». В рамках анализа применяются одномерные сверточные нейронные сети с ядрами сверток по временным измерениям ЭКГ. В свою очередь, двумерные ядра позволяют анализировать 2D-спектрограммы [11] и изображения [18]. Тем не менее, для решения более комплексных задач, затрагивающих анализ не только ЭКГ, ЭЭГ и других разделов электрофизиологии, могут применяться 3D-свертки [16]. Подход, рассматривающий применение двумерных сверточных нейронных сетей, помимо своих очевидных преимуществ, имеет ряд недостатков, самым крупным из которых является необходимость преобразования одномерных данных в двумерные. Из этого проистекает проблема роста вычислительных затрат, так как полученные тензоры изображений по большей части состоят из нулей, задействованных в процессе обработки. Этим обуславливается популярность методов, основанных на использовании 1D-CNN архитектур.

Другой разновидностью глубоких сетей, применяемых для анализа ЭКГ сигналов, являются рекуррентные нейронные сети (RNN, Recurrent Neural Networks). Основная особенность рекуррентных сетей заключается в передаче выходных сигналов предыдущего шага в качестве входных данных для следующего, что позволяет последовательно запоминать данные. Несмотря на ряд преимуществ, использование RNN сетей связано с рядом проблем, основной из которых является проблема затухающего градиента. Для решения этой проблемы была разработана наиболее часто используемая в исследованиях последних лет разновидность RNN архитектур – LSTM (Long Short-Term Memory). В работе [19] авторы применяют LSTM архитектуру для прогнозирования долгосрочных ЭКГ сигналов на наборе данных «MIT-BIH Arrhythmia».

Сверточно-рекуррентные нейронные сети (CRNN, Convolutional Recurrent Neural Networks) сочетают в себе оба описанных выше подхода, используя преимущества двух архитектур для анализа данных с пространственными и временными зависимостями [10, 20]. Возможность эффективно комбинировать анализ пространственных и временных характеристик данных делает данный подход универсальным для решения задач, требующих глубокого анализа. В ряде публикаций предпочитают реализовывать CNN-модели с большим количеством слоев, также позволяющие анализировать

временные зависимости, при этом экономя ресурсы на обучении по сравнению с аналогичными CRNN решениями [21]. Еще одним распространенным методом в анализе ЭКГ данных является использование автокодировщиков. Так, в работе [22] авторы используют маскированный автокодировщик (MAE) для самоконтролируемого обучения модели типа Vision Transformer (ViT) для улучшения ее производительности. В работе [23] авторами используется вариационный автокодировщик (VAE) для анализа данных ЭКГ с целью прогнозирования риска возникновения миокардиальной вентрикулярной аритмии (MVA) у носителей мутации гена PLN p(Arg14del). VAE-автокодировщик был предварительно обучен на 1,1 миллионе ЭКГ и использовался для сжатия ЭКГ данных и последующего сведения их к 32 объясняемым факторам. Для прогнозирования были задействованы модели на основе метода регрессии Кокса. Особенность VAE автокодировщиков заключается в возможности генерировать с их помощью новые данные на основе исходных. Это достигается благодаря обучению вероятностного распределения в латентном пространстве вместо фиксированных точек, а также использованию стохастического процесса для выборки из этого распределения. Пример такой генерации описан в работе [24], где авторы с помощью разработанного вариационного автокодировщика синтезируют ЭКГ в 12 отведениях. При этом, автокодировщик был предварительно обучен на данных с комплексного симулятора сердца на основе метода конечных элементов (FEM).

Помимо VAE автокодировщиков, для решения задачи синтеза ЭКГ сигналов также применяют генеративно-состязательные нейронные сети (GAN, Generative Adversarial Networks). Они состоят из генератора и дискриминатора, при этом каждый является отдельной нейронной сетью, обучающейся в состязательной манере с другой. В статье [23] разрабатывают GAN сеть на основе U-Net для последующего синтезирования фотоплетизмограмм из парных ЭКГ, нужных для диагностики тревожной фибрилляции предсердий.

Полносвязные слои (FC) применяются слоев в более сложных архитектурах, таких как CNN или CRNN. В статье [26] представлен метод классификации сердцебиения на основе эволюционной оптимизации признаков с использованием дифференциальной эволюции (DE) и классификации с использованием вероятностной нейронной сети (PNN) для выявления нормальных и аритмичных сердечных сокращений. В ряде работ используются и гибридные подходы, как в работе [27], где используется модель глубокого обучения с тремя

подходами (CNN, LSTM и Transformer), а также традиционное методы машинного обучения, такие как метод опорных векторов (SVM), логистическую регрессию и алгоритм случайного леса.

Обобщая, отметим, что в области применения машинного обучения для анализа медицинских данных, сформировались направления, обусловленные доступностью и структурой данных. Значительное количество исследований сосредоточено на обработке ЭКГ, что отражает преобладание соответствующих наборов данных в открытом доступе. Анализ литературы демонстрирует, что машинное обучение, и в особенности глубокое обучение, успешно применяется для решения широкого спектра задач: от диагностики сердечно-сосудистых заболеваний и интерпретации сигналов до анализа сна, биометрической аутентификации и устранения шумов.

Методологически данная область характеризуется активным развитием гибридных нейросетевых архитектур, комбинирующих различные подходы для максимизации эффективности. Наиболее распространенными являются сверточные (CNN) и рекуррентные (RNN, в частности LSTM) нейронные сети, а также их комбинации (CRNN), позволяющие одновременно извлекать пространственные и временные зависимости в данных. Помимо этого, значительный исследовательский интерес вызывают генеративные модели, такие как вариационные автокодировщики (VAE) и генеративно-состязательные сети (GAN), используемые как для синтеза новых данных, так и для сжатия и выделения латентных признаков.

Таким образом, развитие направления определяется двумя взаимосвязанными факторами: спецификой решаемой прикладной задачи (диагностика, классификация, синтез) и постоянной эволюцией архитектур моделей, направленной на преодоление вычислительных сложностей, повышение точности и интерпретируемости результатов. Несмотря на явный прогресс, область применения нейронных сетей для решения прикладных задач требует дополнительных исследований, направленных на разработку более эффективных, надежных и интерпретируемых методов анализа.

Проведенный анализ существующих подходов выявил особенность, ограничивающую прямое сравнение результатов исследований с полученными в работе. Упомянутые выше научные работы сфокусированы на решении тематических задач, таких как классификация отдельных видов заболеваний (например, разные виды аритмии), используя для этого специализированные наборы данных. Такая концентрация на нескольких классах заболеваний позволяет достичь более высокой точности, но существенно сужает

область практического применения моделей. Настоящая работа ставит перед собой сложную и практикоориентированную задачу – разработку универсальной модели для мультиметочной классификации широкого спектра заболеваний. Работа с расширенным набором данных, содержащим сравнительно большее число разнородных патологий, является более комплексной задачей технологий интеллектуального анализа.

3. Анализ исходных данных. Основой для обучения модели послужили электрокардиографические (ЭКГ) данные. Ключевыми критериями выбора послужили неинвазивность метода, его повсеместная распространенность (что позволяет сформировать репрезентативный набор данных) и способность объективно фиксировать и графически отображать разность потенциалов, возникающую в процессе сокращения миокарда. В качестве исходных данных используются ЭКГ, загруженные из базы кардиологических данных «The PhysioNet / Computing in Cardiology Challenge 2021» последней редакции. Представленный в работе набор включает 88254 12-канальных записей ЭКГ (Рис. 1).

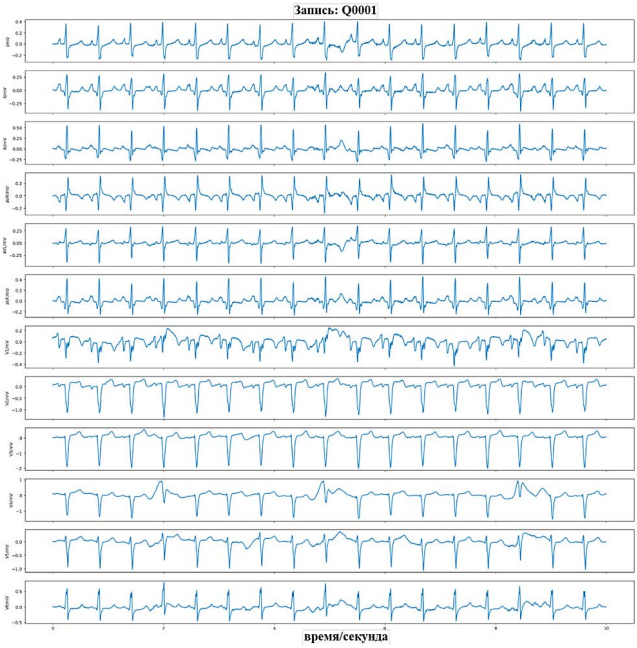


Рис. 1. Пример используемой записи ЭКГ с 12 стандартными отведениями

База данных была сформирована на основе следующих источников: CPSC и CPSC-Extra, INCART, PTB и PTB-XL, Chapman-Shaoxing и Ningbo, база данных Мичиганского университета (UMich), база данных 12-канальной ЭКГ-программы Джорджии (G12EC).

Для минимизации влияния дисбаланса классов в данных в условиях мультиметочной классификации, связанного с малым количеством примеров по ряду диагностических категорий, в работе рассматривается подвыборка из 29 наиболее репрезентативных заболеваний и одной нормы (Рис. 2).

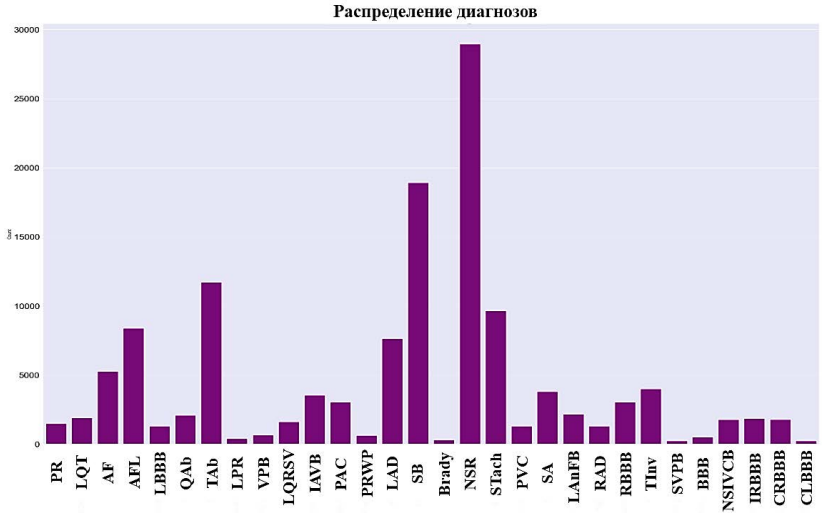


Рис. 2. Распределение выборки данных ЭКГ по диагностическим критериям

На рисунке 2 использовались аббревиатуры: PR – ритм с кардиостимулятором, LQT – удлинённый интервал QT (синдром), AF – фибрилляция предсердий, AFL – трепетание предсердий, LBBB – блокада левой ножки пучка Гиса, QAb – паталогический зубец Q, Tab – паталогический зубец T, LPR – удлинённый интервал PR, VPB – преждевременные сокращения желудочков, LQRSV – низкий вольтаж комплекса QRS, IAVB – атриовентрикулярная блокада 1 степени, PAC – преждевременные сокращения предсердий, PRWP – слабое нарастание зубца R, LAD – отклонение электрической оси сердца влево, SB – синусовая брадикардия, Brady – брадикардия, NSR – синусовый ритм, STach – синусовая тахикардия, PVC – желудочковая экстрасистолия, SA – синусовая аритмия, LAnFB – блокада левой передней ветви пучка Гиса, RAD – отклонение электрической оси

сердца вправо, RBBB – блокада правой ножки пучка Гиса, TInv – инверсия зубца T, SVPB – предсердная экстрасистолия, BBB – блокада ножек пучка Гиса, NSIVCB – неспецифическое нарушение внутрижелудочковой проводимости, IRBBB – неполная блокада правой ножки пучка Гиса, CRBBB – полная блокада правой ножки пучка Гиса, CLBBB – полная блокада левой ножки пучка Гиса.

В полученных данных предварительно заменяются пустые значения и выбросы, после чего сигнал стандартизируется. Для уменьшения вычислительных затрат на последующую обработку сигнала разработанной моделью, в стандартизированном сигнале находятся локальные максимумы каждые 180 временных шагов, после чего отбираются такие максимумы, вокруг которых можно извлечь сегмент длиной в 128 временных шагов. Полученные сегменты объединяются в один массив, после чего все сигналы выравниваются до 2700 временных шагов, что необходимо для стандартизации длины сигналов в каждом ЭКГ по всем отведениям, и передаются на вход модели. В модели, 90% данных отводится на обучающую выборку, а оставшиеся 10% – на валидационную.

Для оценки обобщающей способности предложенного метода на другие типы электрофизиологических данных модель была адаптирована для обработки полисомнографических сигналов и оценивалась на задаче классификации 5 стадий сна (N1, N2, N3, N4 и REM фаза быстрого сна) с использованием базы данных «Sleep-EDF Database Expanded», откуда было взято 153 записи. На рисунке 3 представлен пример данных, где цвет отражает разные стадии сна. Каждая запись представляет полисомнографию ночного сна, синхронно регистрируемую с четырёх отведений с частотой 100 Гц: два биполярных отведения энцефалограммы ЭЭГ (с электродов Fpz-Cz и Pz-Oz), горизонтальная электроокулограмма ЭОГ и электромиограмма ЭМГ подбородочной области. Все записи были разделены на тридцатисекундные отрезки (Рис. 3).

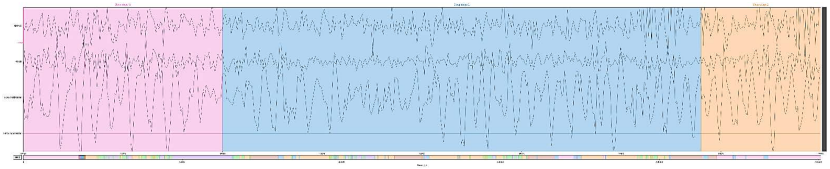


Рис. 3. Пример используемой записи ЭЭГ с 4 стандартными отведениями

Таким образом, суммарная выборка составила 86160 отрезков. В данной реализации 90% данных отводится на обучающую выборку

(2584 батча по тридцать тридцатисекундных отрезков), а оставшиеся 10% – на валидационную (288 батчей по тридцать тридцатисекундных отрезков).

4. Разработка модели. Рассмотрим архитектуру разработанной гибридной нейросетевой модели (Рис. 4).

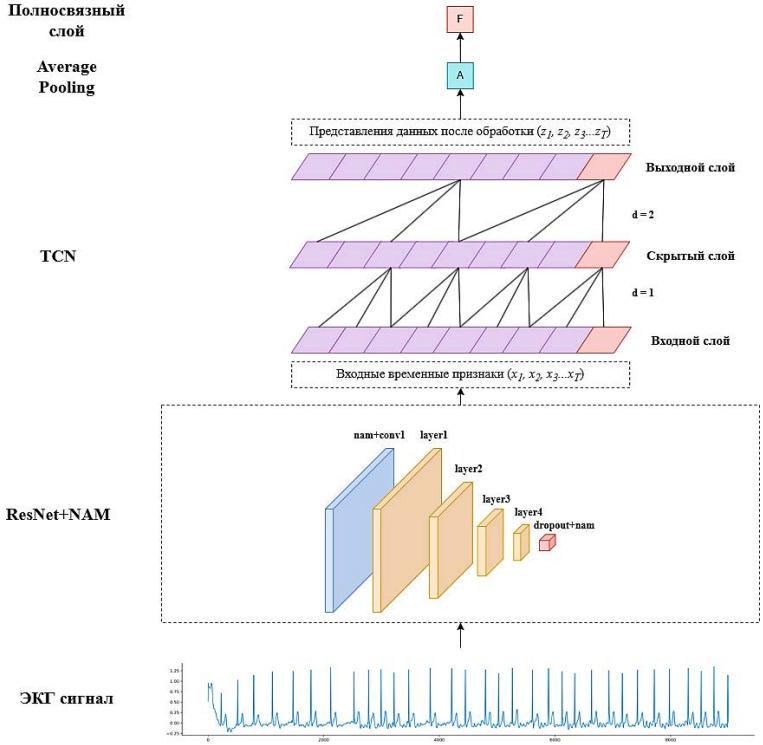


Рис. 4. Архитектура разработанной гибридной нейросетевой модели

Представленная архитектура состоит из трех частей: сверточная нейронная сеть (CNN) в виде модифицированной ResNet с механизмом внимания NAM, временная сверточная нейронная сеть TCN и выходная часть в виде среднего пуллинга и полносвязного слоя. Отчасти такую архитектуру можно назвать трансформером. Новизна заключается в интеграции пространственного и временного механизмов внимания со сверточными временными сетями: модифицированная ResNet с NAM-модулем динамически выделяет

наиболее информативные фрагменты в многоканальных сигналах, а TCN с увеличенной областью восприятия моделирует долгосрочные контекстные зависимости между последовательными эпохами. Такое сочетание позволяет одновременно захватывать локальные паттерны и глобальные циклические закономерности в данных. Особенностью архитектуры является использование промежуточного среднего пуллинга, который, в отличие от распространённого глобального усреднения, сохраняет информацию о временной динамике внутри эпох, что критически важно для классификации переходных и нестационарных состояний.

Рассмотрим архитектуру сверточной нейронной сети (Рис. 5).

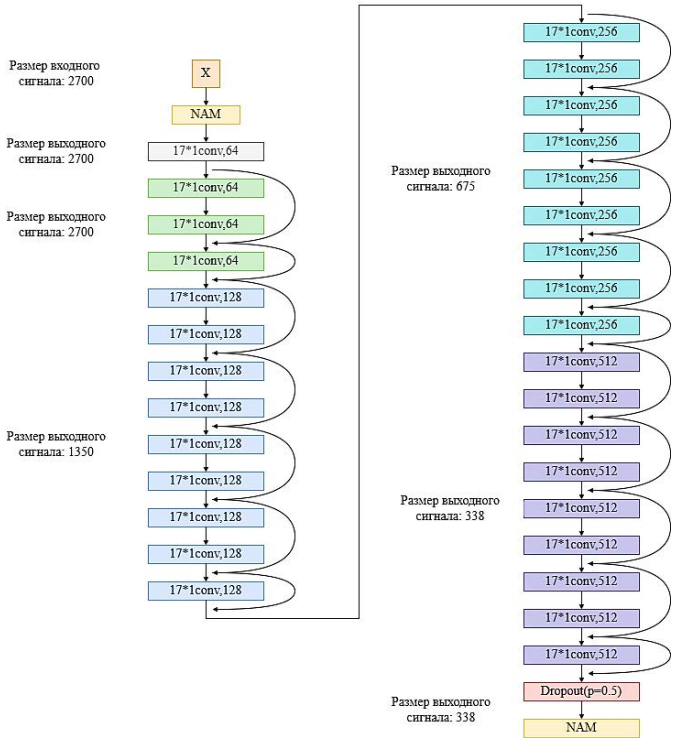


Рис. 5. Архитектура модифицированной ResNet с механизмом внимания

Изогнутые стрелки указывают на остаточные соединения между блоками, позволяющие решать проблему затухающего градиента, что

дает возможность выстраивать нейронные сети с большим количеством слоев.

Механизм внимания NAM (normalized attention module или нормализованный механизм внимания), будучи адаптивным механизмом, используется в модели для подавления шумов и незначительных признаков в данных, усиления значимых признаков, а также для улучшения представления данных. Сам механизм состоит из двух модулей: CAM (channel attention module или канальный модуль внимания) и TAM (temporal attention module или временной модуль внимания) (Рис. 6).

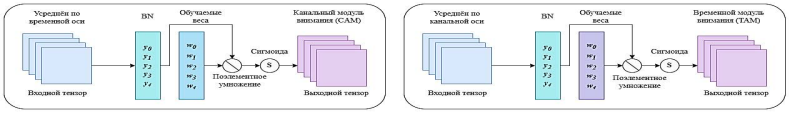


Рис. 6. Структура нормализованного механизма внимания NAM

Входной тензор $x \in R^{B\times C\times T}$, размерности (размер батча, количество каналов, длина временной последовательности) передается на вход канального модуля внимания CAM, где первоначально усредняется по временной оси:

$$\bar{x}_{b,c} = \frac{1}{T}\sum_{t=1}^{T} x_{b,c,t}, \tag{1}$$

где $\bar{x}_c \in R^{B\times C}$ – усредненные значения для каждого канала. По канальным осям применяется батч-нормализация (BN), для стандартизации значения по каналам:

$$\tilde{x}_c = \mathrm{BN}(\bar{x}_c). \tag{2}$$

Для вычисления BN предварительно для каждого канала рассчитывается среднее значение и дисперсия:

$$\begin{aligned} \mu_c &= \frac{1}{B}\sum_{b=1}^{B} \bar{x}_c^{(b)} \\ \sigma_c^2 &= \frac{1}{B}(\bar{x}_c^{(b)} - \mu_c)^2. \end{aligned} \tag{3}$$

Затем входной тензор нормализуется по батчу для каждого канала:

$$\hat{x}_c^{(b)} = \frac{\tilde{x}_c^{(b)} - \mu_c}{\sqrt{\sigma_c^2 + \varepsilon}}. \quad (4)$$

Далее применяется аффинное преобразование:

$$y_c^{(b)} = \gamma_c \hat{x}_c^{(b)} + \beta_c, \quad (5)$$

где γ_cи β_c– обучаемые параметры масштаба и сдвига.

Для расчета внимания по каналам нормализованные значения умножаются на обучаемые веса внимания, после чего к результату применяется сигмоидная функция активации:

$$s_c = \sigma(\tilde{x}_c \cdot \omega_{cam}), \quad (6)$$

где s_c– коэффициент внимания для канала, σ– сигмоидная функция активации, ω_{cam}– обучаемые веса внимания.

Полученный результат масштабируется для дальнейшего применения к исходному тензору:

$$x' = x \cdot s_{cam}, \quad (7)$$

где s_{cam}– веса канального внимания, расширенные вдоль временной оси.

Для расчета внимания по временной оси исходный тензор x передается на вход временному модулю внимания TAM, где аналогично усредняется по каналам:

$$\tilde{x}_{b,t} = \frac{1}{C}\sum_{c=1}^{C} x_{b,c,t}, \quad (8)$$

где $\tilde{x}_t \in R^{B\times T}$– усредненные значения для каждого временного шага.

По временной оси применяется батч-нормализация:

$$\tilde{x}_t = \mathrm{BN}(\tilde{x}_t). \quad (9)$$

Аналогично модулю CAM, рассчитывается внимание по временным шагам с использованием обучаемых весов:

$$s_t = \sigma(\tilde{x}_t \cdot \omega_{tam}). \quad (10)$$

Результат масштабируется и применяется к исходному тензору:

$$x^{'} = x \cdot s_{tam}. \tag{11}$$

Итоговая формула имеет вид:

$$x^{'} = x \cdot s_{cam} \cdot s_{tam} = x \cdot \sigma(\mathrm{BN}(\bar{x}_c) \cdot \omega_{cam}) \cdot \sigma(\mathrm{BN}(\bar{x}_t) \cdot \omega_{tam}). \tag{12}$$

Полученный тензор передается на вход сверточной нейронной сети, состоящей из слоя свертки и 4 слоев с разным количеством блоков с остаточными соединениями.

Каждая свертка вычисляется по следующей формуле одномерной свертки:

$$x^{'}(c,t) = \sum_{c^{'}} \sum_{k} \tilde{x}(c^{'}, S \cdot t - k) \cdot \omega(c, c^{'}, k), \tag{13}$$

где $x^{'}(c,t)$ – выходной тензор, $c^{'}$– входной канал, c– выходной канал, t – временной шаг, k – индекс смещения, $\omega(c, c^{'}, k)$– веса ядра свертки, S – страйд или шаг перемещения фильтра свертки, $\tilde{x}$ – входной тензор, к которому применили паддинг:

$$\tilde{x}(c,t) = \begin{cases} x(c^{'},t), \textit{ если } P \le t < T + P \\ 0, \quad \textit{иначе} \end{cases}, \tag{14}$$

где P – значение паддинга. Паддинг нужен для того, чтобы при применении операции свертки не терялась информация на краях данных.

Для сохранения размерности данных паддинг рассчитывается по формуле:

$$P = \frac{K-1}{2}, \tag{15}$$

где K – размер ядра свертки. Таким образом, в указанной модели при применении операции свертки с ядром 17 используется паддинг 8.

Итоговая формула применения одномерной свертки к исходному тензору в матричном виде имеет вид:

$$X^{'} = X \cdot W, \tag{16}$$

где W – матрица свертки, X – входной тензор, развернутый в виде матрицы, Теплица.

В отличие от стандартной архитектуры ResNet (остаточная нейронная сеть), в которой используется базовый остаточный блок для всех вычислений, в представленной архитектуре используются так называемые ResStage блоки (Рис. 7).

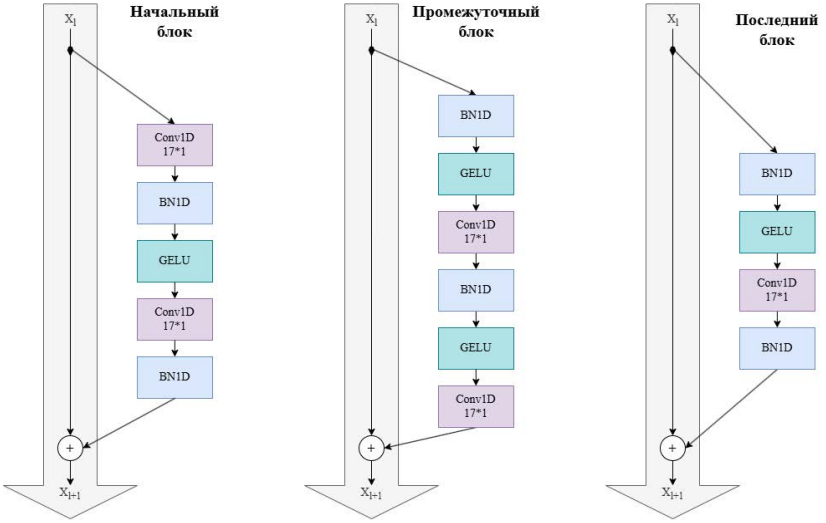


Рис. 7. Структура ResStage блоков сверточной нейронной сети

Данный подход заключается в использовании разных блоков в процессе вычисления: начального, промежуточного и заключительного. В зависимости от количества блоков в слое количество промежуточных блоков может варьироваться. На рисунке 7 кругом со знаком плюс обозначаются остаточные соединения, работающие по принципу:

$$y = F(x) + x, \tag{17}$$

где x – входной тензор блока, $F(x)$ – тензор после математических преобразований внутри блока, y – выходной тензор блока.

В представленной модели используется простое (identity shortcut) соединение. Данный механизм позволяет повышать точность обучения, решать проблему затухающих градиентов и упрощать обучение глубоких нейронных сетей. В качестве функции активации в блоках используется GELU (Гауссов линейный блок ошибки):

$$\mathrm{GELU}(x) = x \cdot \Phi(x) = \frac{1}{2}x(1 + \mathrm{erf}(\frac{x}{\sqrt{2}})), \quad (18)$$

где erf – функция ошибки из распределения Гаусса.

Использование данной функции активации объясняется рядом преимуществ: она более адаптивна, не имеет резкого порога (в отличие от ReLU), благодаря чему модель легче поддается оптимизации, обрабатывает малозначимые признаки, адаптивно уменьшая их влияние и тем самым предотвращая ситуацию, когда нейрон перестает активироваться и обновлять свои веса в процессе обучения. Перечисленные факторы обуславливают применение GELU в современных моделях глубокого обучения и архитектурах типа трансформер.

После вычисления всех скрытых резидуальных слоев применяется Dropout, являющийся эффективным методом регуляризации, предотвращающим совместную адаптацию нейронов. При применении Dropout к входному тензору, случайным образом обнуляются некоторые элементы входного тензора с вероятностью p на время обучения. Обнуляемые элементы выбираются независимо для каждого прямого вызова и выбираются из распределения Бернулли. Остальные элементы входного тензора масштабируются на $\frac{1}{1-p}$ на время обучения.

После преобразования полученный тензор передается на вход временной сверточной нейронной сети TCN (Рис. 8).

Основной принцип TCN сетей заключается в использовании остаточных соединений и каузальных разреженных сверток. Формула каузальной разреженной свертки имеет вид:

$$x_t^{'} = \sum_{i=0}^{k-1} w_i \cdot x_{t-i \cdot d}, \quad (19)$$

где k – размер ядра свертки, w_i – веса сверточного фильтра, d – коэффициент разрежения (диляция), $x_{t-i \cdot d}$ – входное значение, взятое с разрежением.

В каузальных свертках индексы смещены так, чтобы при обработке временной последовательности не заглядывать в будущее, в отличие от обычных сверток в которых используется симметричное окно. Это позволяет модели принимать решения на основе текущих и прошлых данных, что достигается за счет использования следующего паддинга:

$$P = (K - 1) \cdot d, \quad (20)$$

где K – размер ядра свертки, d – коэффициент разрежения (диляции).

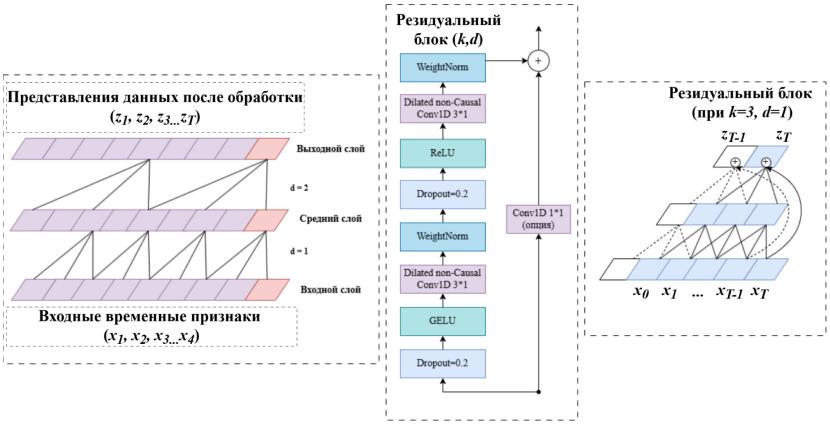
Тем не менее, в работе используются некаузальные разреженные свертки, позволяющие анализировать данные более глобально, но хуже справляющиеся с анализом в реальном времени. Использование диляции позволяет экспоненциально увеличивать область восприятия без роста числа слоев, тем самым становится возможным захватывать долгосрочные зависимости без увеличения вычислительных затрат. В качестве нормализации в архитектуре резидуального блока TCN используется WeightNorm (нормализация весов):

$$W = g\frac{v}{\|v\|}, \tag{21}$$

где g – скалярный параметр масштаба (обучаемый параметр, контролирующий обучение весов), v – вектор параметров (обучаемый вектор направления весов), W – нормализованные веса, $\|v\|$ – L2-норма вектора v, вычисляемая по формуле:

$$\|v\| = \sqrt{\sum_i v_i^2}, \tag{22}$$

где v_i – компонента вектора v.



а) б) с)

Рис. 8. Архитектура временной сверточной нейронной сети: а) схема TCN сети; б) архитектура резидуального блока TCN сети; с) схема работы резидуального блока при использовании каузальной свертки с разрежением с ядром 3 и диляцией 1

Нормализованные веса в виде единичного вектора умножаются на нормализованный вектор g, что позволяет контролировать размер весов. Использование нормализации весов в TCN архитектурах позволяет стабилизировать обучение за счет решения проблемы с резким изменением градиентов и уменьшения зависимости от начальной инициализации весов, ускорять сходимость обучения и решать проблему переобучения, благодаря более равномерной распределенности весов. Помимо функции активации GELU в резидуальном блоке TCN сети также используется стандартная функция ReLU:

$$\mathrm{ReLU}(x) = max(0, x), \tag{23}$$

где x– входное значение.

Функция возвращает x если он положительный и 0 если отрицательный. Подход с использованием двух функций активации продемонстрировал более высокий результат при тестировании.

Выбор TCN-архитектур в современных задачах машинного обучения обуславливается возможностью параллельных вычислений на графических ускорителях благодаря использованию сверток (в то время как скрытое состояние в реккурентных сетях, таких как LSTM, передается последовательно), отсутствием проблемы затухающих градиентов при обработке длинных последовательностей, возможностью обрабатывать последовательности разной длины, а также отсутствием утечки будущей информации за счет использования специализированного паддинга, в отличие от сверточных нейронных сетей. Благодаря вышеописанным факторам TCN-архитектуры хорошо зарекомендовали себя в текущих задачах глубокого обучения и демонстрируют высокий уровень точности и эффективности при обработке последовательностей данных. После TCN модели тензор передается слой адаптивного усредняющего пуллинга:

$$x_{b,c,1} = \frac{1}{T}\sum_{t=1}^{T} x_{b,c,t}, \tag{24}$$

где T – исходная длина последовательности.

Адаптивный усредняющий пулинг вычисляет среднее значение по всей временной оси для каждого канала в тензоре, после чего масштабирует исходный тензор. Он применяется для уменьшения размерности и агрегации глобальных временных признаков перед полносвязным слоем, что необходимо для сохранения временных признаков. После применения пулинга канальная и временная

размерности тензора разворачиваются в одномерный вектор, тем самым получая тензор размерностью $x_{b,c}$, который затем передается на полносвязный слой:

$$x' = xW^T + b, \tag{25}$$

где x – входной тензор, W^T – транспонированная матрица весов, b – вектор смещения.

На выходе имеем тензор с выходными значениями для каждого класса, которые затем необходимо преобразовать в бинарные значения, соответствующие классифицированным заболеваниям, с помощью пороговых значений.

В модели применяется байесовская оптимизация для нахождения приближенных оптимальных пороговых значений и локальная минимизация методом Пауэлла – для их уточнения. Выбранная метрика для оптимизации – точность, которая будет максимизироваться. Для этого, оптимизатор строит аппроксимацию выбранной целевой функции ее нахождения с помощью гауссова процесса:

$$p(f(t)|D) = N(\mu(t), \sigma^2(t)), \tag{26}$$

где $\mu(t)$ – среднее (оценка целевой функции в точке t), $\sigma^2(t)$ – дисперсия в точке t, $D = \{(t_i, f(t_i))\}$ – наблюдаемые точки.

Таким образом, находятся те пороги, в которых целевая функция может иметь лучшее значение. Для дальнейшего поиска среди выбранных порогов применяется функция полезности $\mathrm{EI}(t)$:

$$\mathrm{EI}(t) = (\mu(t) - f^* - \xi)\Phi(Z) + \sigma(t)\phi(Z), \tag{27}$$

где f^* – наилучшее значение целевой функции, ξ – коэффициент исследования для баланса между выбором и проверкой порогов, $\Phi(Z)$ – кумулятивная функция распределения, $\phi(Z)$ – плотность нормального распределения. В данном случае параметр Z рассчитывается по формуле:

$$Z = \frac{\mu(t) - f^* - \xi}{\sigma(t)}. \tag{28}$$

В случае функции полезности $\mathrm{EI}(t)$, если $\mu(t)$ намного больше значения функции f^*, порог будет проверяться. Если $\sigma(t)$ имеет большое значение, алгоритм не уверен в ответе и данный порог также

будет проверяться на возможность дальнейшего улучшения. При $\mu(t) \approx f^*$ и низком значении $\sigma(t)$ порог исключается. Найденные пороги передаются в оптимизатор для построения новой аппроксимации целевой функции.

После нахождения оптимальных порогов с помощью байесовской оптимизации точки передаются в локальный минимизатор, уточняющий значения порогов методом Пауэлла, являющимся безградиентным методом оптимизации на основе принципа координатного спуска, последовательно улучшающим параметры вдоль разных направлений. Для этого первоначально выбирается минимизируемая функция, в данном случае функция точности (accuracy), взятая со знаком минус. Таким образом, решается задача $min_T\, f\,(T)$, где $f(T)$:

$$f(T) = -\text{accuracy\_score}(y_{true}, (y_{pred} \geq T)). \quad (29)$$

Далее выполняется одномерная оптимизация по каждому направлению для каждого порога:

$$\alpha_i = \min_\alpha f(T^{(i-1)} + \alpha d_i), \quad (30)$$

где d_i – текущее направление (изначально берутся базисные единичные векторы), α – шаг в выбранном направлении.

После оптимизации по направлению точка обновляется:

$$T^{(i)} = T^{(i-1)} + \alpha_i d_i. \quad (31)$$

После прохода по всем направлениям добавляется новое направление:

$$d_{new} = T^{(n)} - T^{(0)}, \quad (32)$$

где $T^{(0)}$ – начальная точка, полученная после байесовской оптимизации, $T^{(n)}$ – точка, полученная после прохода по всем направлениям.

Алгоритм выполняется снова, после чего проверяется сходимость с новой найденной точкой. Алгоритм завершает работу если достигает предельного количества итераций или если выполняется условие сходимости для всех пороговых значений:

$$\sqrt{\sum_{i=1}^{m}(T_i^{(new)} - T_i^{(n)})^2} < \varepsilon, \quad (33)$$

где ε – некоторое пороговое значение, определяющее сходимость.

Полученные пороги применяются к выходам модели, преобразовывая их в бинарные значения.

В качестве функции потерь в работе используется бинарная кросс-энтропия с логитами (BCEWithLogitsLoss), применяемая в задачах бинарной многоклассовой и мультиметочной классификации:

$$L = \frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{C} -\left[y_{ij}\, log\, \sigma\,(z_{ij}) + (1 - y_{ij})\, log(\,1 - \sigma(z_{ij}))\right], \quad (34)$$

где N – количество примеров в батче, C – количество каналов, z_{ij} – логиты модели (непреобразованные выходы модели), y_{ij} – истинные бинарные метки, σ – сигмоидная функция активации.

Оптимизатором модели был выбран стандартный Adam (оценка адаптивного метода). При его использовании веса модели обновляется согласно представленному ниже алгоритму. В первую очередь, вычисляется градиент функции потерь:

$$g_t = \nabla_\theta L(\theta_t), \quad (35)$$

где θ_t – параметры модели на шаге t, ∇_θ – оператор градиента, $L(\theta_t)$ – функция потерь.

Затем рассчитывается экспоненциальное скользящее среднее первого момента (или среднее значение градиента):

$$m_t = \beta_1 m_{t-1} + (1 - \beta_1) g_t, \quad (36)$$

где β_1 – коэффициент сглаживания.

Затем, вычисляется экспоненциальное скользящее среднее второго момента (или среднее значение квадрата градиента):

$$v_t = \beta_2 v_{t-1} + (1 - \beta_2) g_t^2, \quad (37)$$

где β_2 – также коэффициент сглаживания.

На первом шаге выполняется коррекция смещения:

$$\widehat{m}_t = \frac{m_t}{1-\beta_1^t}, \widehat{v}_t = \frac{v_t}{1-\beta_2^t}. \quad (38)$$

Наконец, параметры модели обновляются:

$$\theta_{t+1} = \theta_t - \frac{\eta}{\sqrt{\hat{v}_t}+\varepsilon}\hat{m}_t, \quad (39)$$

где η – скорость обучения, ε – некоторое малое число для предотвращения деления на ноль.

Преимущества использования оптимизатора Adam заключаются в том, что он автоматически адаптирует скорость обучения для каждого параметра модели, сочетает в себе разные подходы, что делает его более универсальным и стабилизированным, Adam не требует сложной настройки гиперпараметров, а также демонстрирует быструю и стабильную сходимость в сложных задачах.

Вся реализация модели выполнена с использованием фреймворка PyTorch. Обучение модели осуществлялось с использованием объектно-ориентированного подхода, в котором основной класс инкапсулирует всю логику тренировочного и валидационного циклов. Для повышения устойчивости обучения была реализована стратегия адаптивного изменения скорости обучения на основе значения функции потерь на валидационном наборе данных. Для контроля переобучения и предотвращения избыточных вычислений был реализован механизм ранней остановки, срабатывающий при отсутствии улучшения ошибки на валидационных данных в течение заданного количества эпох. После каждой эпохи состояние модели, включая параметры оптимизатора, номер текущей эпохи и значение функции потерь, сохранялось на диск в виде контрольной точки. Это позволяло при необходимости восстановить обучение с нужной эпохи, а также проанализировать поведение модели на различных стадиях оптимизации.

5. Верификация модели. Для оценки качества модели использовались несколько метрик, каждая из которых отражает разные стороны работы модели. Полученные результаты оценки модели на валидационной выборке представлены в таблице 1.

Таблица 1. Результаты оценки модели на валидационной выборке для 30 классов

Метрика качества	Результат
Accuracy	47%
F1-score	59%
Roc-Auc	64%
Precision	67%
Recall	59%

Отметим, что для оценки моделей анализа медицинских данных необходимо применять несколько метрик, так как обучение моделей производится на несбалансированных выборках, где присутствует дисбаланс классов, поэтому, несмотря на низкий показатель глобальной точности (Accuracy = 47%), что является закономерным следствием экстремального дисбаланса при 30 классах, модель демонстрирует эффективность. Ключевые метрики, устойчивые к дисбалансу, подтверждают её работоспособность: высокое значение Precision (67%) указывает на точность предсказаний для редких категорий, а F1-score (59%) – на баланс между точностью и полнотой их обнаружения. ROC-AUC (64%) подтверждает общую способность модели к содержательной дискриминации всех классов. Таким образом, представленные метрики свидетельствуют о практической применимости модели для задач классификации в условиях сильного дисбаланса, где критически важным является корректное выявление редких, но значимых событий.

Дополнительно разработанная модель была обучена на задаче классификации 12 наиболее репрезентативных заболеваний. Полученные результаты оценки модели на валидационной выборке представлены в таблице 2.

Таблица 2. Результаты оценки модели на валидационной выборке для 12 классов

Метрика качества	Результат
Accuracy	53%
F1-score	64%
Roc-Auc	72%
Precision	65%
Recall	60%

Также, разработанная модель была адаптирована и обучена для решения задачи многоклассовой классификации 5 стадий сна по данным ЭЭГ. Полученные результаты оценки модели на валидационной выборке представлены в таблице 3.

Таблица 3. Результаты оценки модели на валидационной выборке при решении задачи классификации 5 стадий сна

Метрика качества	Результат
Accuracy	78%
F1-score	78%
Precision	79%
Recall	78%

Представленная верификация подтверждает гипотезу о высокой практической применимости предложенного в работе подхода для анализа сложных медицинских данных. Разработанная гибридная архитектура, основанная на комбинации сверточных и временных сверточных нейронных сетей с механизмом внимания, продемонстрировала высокую результативность в задачах автоматизированной диагностики на разнородных физиологических сигналах, в частности, для выявления патологий по данным ЭКГ и классификации стадий сна на основе многоканальных записей ЭЭГ с включением ЭОГ и ЭМГ. Универсальность и эффективность данных методов открывает перспективы для их адаптации при проектировании других специализированных архитектур для анализа медицинских данных.

6. Заключение. В рамках данной статьи показана принципиальная возможность и обоснована целесообразность применения алгоритмов машинного обучения для анализа многомерных медицинских данных. Использование методов интеллектуальной обработки, в частности, глубоких сверточных сетей, позволило достичь высоких показателей точности в задачах автоматизированной диагностики и многоклассовой классификации патологий. Разработанная архитектура гибридной нейросетевой модели реализует полный конвейер обработки данных, включающий этапы предобработки сигналов, селекции информативных признаков, а также их последующего анализа с помощью современных адаптивных моделей. Применение данного подхода продемонстрировало высокую эффективность, что подтверждается результатами оценки модели.

Полученные результаты обуславливают перспективность использования предложенных методов для создания систем поддержки принятия решений в медицине. Разработанная архитектура обладает значительным потенциалом для дальнейшего совершенствования и развития. Пути повышения эффективности включают в себя увеличение объема и качества размеченных данных для обучения, применение методов аугментации, а также тонкую настройку гиперпараметров моделей и использование более сложных ансамблевых алгоритмов. Научная и практическая значимость полученных результатов подтверждает целесообразность дальнейших исследований в области создания автоматизированных информационных систем обработки и анализа медицинских данных.

Литература

1. Президент России. Указ Президента РФ от 02.07.2021 № 400 «О Стратегии национальной безопасности Российской Федерации». URL: http://kremlin.ru/acts/bank/47046 (дата обращения: 20.02.2025).
2. Президент России. Указ Президента РФ от 13.05.2017 № 208 «О Стратегии экономической безопасности Российской Федерации на период до 2030 года». URL: http://www.kremlin.ru/acts/bank/41921 (дата обращения: 20.02.2025).
3. Шевцов Е.С., Шамин Р.В. Логическая интеграция информационных систем на основе экспертных систем // Российский технологический журнал. 2025. Т. 13. №2. С. 27–35. DOI: 10.32362/2500-316X-2025-13-2-27-35.
4. Mityakov E.S., Lapaev D.N., Beskhmelnov M.I., Ladynin A.I., et al. Knowledge-intensive Enterprises Multi-criteria Decision-making Energy Security Management Information System // Proceedings of the 2022 Conference of Russian Young Researchers in Electrical and Electronic Engineering (ElConRus). 2022. pp. 1711–1713. DOI: 10.1109/ElConRus54750.2022.9755655.
5. Красноперова А.С., Твердохлебов А.С., Карташов А.А., Вебер В.И., Куприц В.Ю. Исследование эффективности применения моделей нейронных сетей YOLO для распознавания объектов на радиолокационных изображениях // Российский технологический журнал. 2025. Т. 13. №4. С. 25–36. DOI: 10.32362/2500-316X-2025-13-4-25-36.
6. Всемирная организация здравоохранения. Сердечно-сосудистые заболевания (ССЗ). URL: https://www.who.int/ru/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds) (дата обращения: 25.02.2025).
7. Министерство здравоохранения Российской Федерации. Федеральный проект «Борьба с сердечно-сосудистыми заболеваниями». URL: https://minzdrav.gov.ru/poleznye-resursy/natsproektzdravoohranenie/bssz (дата обращения: 25.02.2025).
8. Шишкин С.В., Шейман И.М., Алмазов А.А., Бирюкова А.И., Власов В.В., Потапчик Е.Г., Сажина С.В., Салахутдинова С.К. Российское здравоохранение: перспективы развития. Доклад НИУ ВШЭ / Под ред. Шишкина С.В. // М.: Изд. дом ВШЭ. 2024. 60 с.
9. Hong S., Zhou Y., Shang J., Xiao C., Sun J. Opportunities and challenges of deep learning methods for electrocardiogram data: A systematic review // Computers in Biology and Medicine. 2020. vol. 122. pp. 1–16. DOI: 10.1016/j.compbiomed.2020.103801.
10. Zhang C.-J., Yuan L., Tang F.-Q., Cai H.-P., Qian Y.-F., Wang C. Heart failure classification using deep learning to extract spatiotemporal features from ECG // BMC Medical Informatics and Decision Making. 2024. vol. 24. no. 1. pp. 1–19. DOI: 10.1186/s12911-024-02415-4.
11. Madan P., Singh V., Singh D.P., Diwakar M., Pant B., Kishor A. A Hybrid Deep Learning Approach for ECG-based Arrhythmia Classification // Bioengineering. 2022. vol. 9. no. 4. DOI: 10.3390/bioengineering9040152.
12. Huang Z., Yu J., Shan Y., Wang X. A Non-Invasive Fetal QRS Complex Detection Method Based on a Multi-Feature Fusion Neural Network // Applied Sciences. 2024. vol. 14. no. 19.
13. Teplitzky B.A., McRoberts M., Ghanbari H. Deep learning for comprehensive ECG annotation // Heart Rhythm. 2020. vol. 17. no. 5. pp. 881–888. DOI: 10.1016/j.hrthm.2020.02.015.
14. Riemann D., Krone L.B., Wulff K., Nissen C. Sleep, insomnia, and depression // Neuropsychopharmacology. 2020. vol. 45. no. 1. pp. 74–89. DOI: 10.1038/s41386-019-0411-y.

15. Paul T., Hassan O., Islam S.K., Mosa A.S.M. Real-time obstructive sleep apnea detection from raw ECG and SpO_2 signal using convolutional neural network // AMIA Joint Summits on Translational Science Proceedings. 2024. pp. 662–669.
16. Ji X., Li Y., Wen P., Barua P., Acharya U.R. MixSleepNet: A Multi-Type Convolution Combined Sleep Stage Classification Model // Computer Methods and Programs in Biomedicine. 2024. vol. 244. 107992 p. DOI: 10.1016/j.cmpb.2023.107992.
17. Mvuh F.L., Ebode Ko'a C.O.V., Bodo B. Multichannel high noise level ECG denoising based on adversarial deep learning // Scientific Reports. 2024. vol. 14. no. 1. DOI: 10.1038/s41598-023-50334-7.
18. Aversano L., Bernardi M.L., Cimitile M., Montano D., Pecori R. Characterization of Heart Diseases per Single Lead Using ECG Images and CNN-2D // Sensors. 2024. vol. 24. no. 11. DOI: 10.3390/s24113485.
19. Zacarias H., Marques J.A.L., Felizardo V., Pourvahab M., Garcia N.M. ECG Forecasting System Based on Long Short-Term Memory // Bioengineering. 2024. vol. 11. no. 1. DOI: 10.3390/bioengineering11010089.
20. Kamanditya B., Fuadah Y.N., Mahardika T.N.Q., Lim K.M. Continuous blood pressure prediction system using Conv-LSTM network on hybrid latent features of photoplethysmogram (PPG) and electrocardiogram (ECG) signals // Scientific Reports. 2024. vol. 14. no. 1.
21. Khan F., Yu X., Yuan Z., Rehman A.U. ECG classification using 1-D convolutional deep residual neural network // PLOS One. 2023. vol. 18. no. 4. e0284791 p. DOI: 10.1371/journal.pone.0284791.
22. Sawano S., Kodera S., Setoguchi N., et al. Applying masked autoencoder-based self-supervised learning for high-capability vision transformers of electrocardiographies // PLOS One. 2024. vol. 19. no. 8. pp. 1–17. DOI: 10.1371/journal.pone.0307978.
23. van de Leur R.R., de Brouwer R., Bleijendaal H., et al. ECG-only explainable deep learning algorithm predicts the risk for malignant ventricular arrhythmia in phospholamban cardiomyopathy // Heart Rhythm. 2024. vol. 21. no. 7. pp. 1102–1112. DOI: 10.1016/j.hrthm.2024.02.038.
24. Nishikimi R., Nakano M., Kashino K., Tsukada S. Variational autoencoder-based neural electrocardiogram synthesis trained by FEM-based heart simulator // Cardiovascular Digital Health Journal. 2023. vol. 5. no. 1. pp. 19–28. DOI: 10.1016/j.cvdhj.2023.12.002.
25. Sohn J., Shin H., Lee J., Kim H.C. Validation of Electrocardiogram Based Photoplethysmogram Generated Using U-Net Based Generative Adversarial Networks // Journal of Healthcare Informatics Research. 2024. vol. 8. no. 1. pp. 140–157. DOI: 10.1007/s41666-023-00156-z.
26. Nasim A., Kim Y.S. DE-PNN: Differential Evolution-Based Feature Optimization with Probabilistic Neural Network for Imbalanced Arrhythmia Classification // Sensors. 2022. vol. 22. no. 12. 4450 p. DOI: 10.3390/s22124450.
27. Din S., Qaraqe M., Mourad O., Qaraqe K., Serpedin E. ECG-based cardiac arrhythmias detection through ensemble learning and fusion of deep spatial-temporal and long-range dependency features // Artificial Intelligence in Medicine. 2024. vol. 150. 102818 p.

Лебедев Олег Борисович — д-р техн. наук, доцент, профессор, кафедра информатики, Федеральное государственное бюджетное образовательное учреждение высшего образования "МИРЭА – Российский технологический университет" (РТУ МИРЭА). Область научных интересов: искусственный интеллект, машинное обучение, алгоритмы роевой оптимизации, системы поддержки принятия решений, разработка и применение

математических и программных инструментов. Число научных публикаций — 270. lebedev_o@mirea.ru; пр. Вернадского, 78, 119454, Москва, Россия; р.т.: +7(499)215-6565.

Шмелева Анна Геннадьевна — канд. физ.-мат. наук, доцент, доцент, кафедра информатики, Федеральное государственное бюджетное образовательное учреждение высшего образования "МИРЭА – Российский технологический университет" (РТУ МИРЭА). Область научных интересов: искусственный интеллект, машинное обучение, системы поддержки принятия решений. Число научных публикаций — 107. ann_shmeleva@mail.ru; пр. Вернадского, 78, 119454, Москва, Россия; р.т.: +7(499)215-6565.

Гежа Николай Сергеевич — студент, кафедра информатики, Федеральное государственное бюджетное образовательное учреждение высшего образования "МИРЭА – Российский технологический университет" (РТУ МИРЭА). Область научных интересов: машинное обучение, искусственный интеллект, анализ биомедицинских данных, автоматизация поддержки принятия решений. Число научных публикаций — 1. n.s.gezha@gmail.com; пр. Вернадского, 78, 119454, Москва, Россия; р.т.: +7(499)215-6565.

DOI. 10.15622/ia.25.3.4

O. LEBEDEV, A. SHMELEVA, N. GEZHA

MEDICAL ELECTROPHYSIOLOGICAL DATA AUTOMATED INTERPRETATION DEEP LEARNING MODEL

Lebedev O., Shmeleva A., Gezha N. **Medical Electrophysiological Data Automated Interpretation Deep Learning Model.**

Abstract. The article is devoted to the development of a neural network model for automated analysis of medical data in the field of electrophysiology based on deep learning methods. The presented study is of high relevance due to the growing need for improvement of complex spatiotemporal signal processing. Such signals include electrocardiograms and electroencephalograms in clinical practice. Convolutional neural networks (CNNs), which extract spatial features, and temporal convolutional neural networks (TCNs), which specialize in identifying long-term time dependencies, were used as analysis tool. The integration of such architectures made it possible to identify and generalize complex patterns in signals and, subsequently, classify them with high accuracy. The paper presents a hybrid neural network model architecture developed for processing one-dimensional time series and provides a rationale for choosing key hyperparameters. To optimize the model, a strategy for adaptively changing the learning rate based on the value of the loss function on the validation dataset was implemented. An early training stopping mechanism was implemented to prevent redundant computations and to control overfitting in the event of no decrease in the error value on the validation data over a specified number of epochs. During training, model parameters were recorded after each epoch, including the epoch number, loss function value, and optimizer parameters. This allows for the reconstruction of training from any epoch and a detailed analysis of the model's behavior at various stages of training. The application of the developed analysis model enables more in-depth analysis of electrophysiological data, improving the objectivity of diagnostics. The methods presented in this paper can also be scaled to different biomedical data. The practical significance of the paper lies in the integration of the developed model into biomedical systems as a tool for supporting medical decisions, utilizing automated detection of pathologies and reducing the burden on doctors.

Keywords: machine learning, deep learning, neural networks, medical data, electrophysiology.

References

1. Ukaz Prezidenta RF ot 02.07.2021 № 400 [On the National Security Strategy of the Russian Federation]. Available at: http://kremlin.ru/acts/bank/47046 (accessed 20.02.2025). (In Russ.).
2. Ukaz Prezidenta RF ot 13.05.2017 № 208 [On the Strategy of Economic Security of the Russian Federation for the period up to 2030]. Available at: http://www.kremlin.ru/acts/bank/41921 (accessed 20.02.2025). (In Russ.).
3. Shevtsov E.S., Shamin R.V. [Logical integration of information systems based on expert systems]. Rossiyskiy tekhnologicheskiy zhurnal – Russian Journal of Technology. 2025. vol. 13. no. 2. pp. 27–35. DOI: 10.32362/2500-316X-2025-13-2-27-35. (In Russ.).
4. Mityakov E.S., Lapaev D.N., Beskhmelnov M.I., Ladynin A.I., et al. Knowledge-intensive Enterprises Multi-criteria Decision-making Energy Security Management Information System. Proceedings of the 2022 Conference of Russian Young Researchers in Electrical and Electronic Engineering (ElConRus). 2022. pp. 1711–1713. DOI: 10.1109/ElConRus54750.2022.9755655.

5. Krasnoperova A.S., Tverdokhlebov A.S., Kartashov A.A., Weber V.I., Kuprits V.Yu. [Investigation of the effectiveness of YOLO neural network models for object recognition in radar images]. Rossiyskiy tekhnologicheskiy zhurnal – Russian Journal of Technology. 2025. vol. 13. no. 4. pp. 25–36. DOI: 10.32362/2500-316X-2025-13-4-25-36. (In Russ.).
6. Vsemirnaya organizatsiya zdravookhraneniya. [Cardiovascular diseases (CVD)]. Available at: https://www.who.int/ru/news-room/fact-sheets/detail/ cardiovascular-diseases-(cvds) (accessed 25.02.2025). (In Russ.).
7. Ministerstvo zdravookhraneniya Rossiyskoy Federatsii. [Federal project «Fight against cardiovascular diseases»]. Available at: https://minzdrav.gov. ru/poleznye-resursy/natsproektzdravoohranenie/bssz (accessed 25.02.2025). (In Russ.).
8. Shishkin S.V., Sheiman I.M., Almazov A.A., Biryukova A.I., Vlasov V.V., Potapchik E.G., Sazhina S.V., Salakhutdinova S.K. Rossiyskoe zdravookhranenie: perspektivy razvitiya. Doklad NIU VShE [Russian healthcare: development prospects. Report of the Higher School of Economics] (Pod red. Shishkina S.V.) M.: Izd. dom VShE. 2024. 60 p. (In Russ.).
9. Hong S., Zhou Y., Shang J., Xiao C., Sun J. Opportunities and challenges of deep learning methods for electrocardiogram data: A systematic review. Computers in Biology and Medicine. 2020. vol. 122. pp. 1–16. DOI: 10.1016/j.compbiomed.2020.103801.
10. Zhang C.-J., Yuan L., Tang F.-Q., Cai H.-P., Qian Y.-F., Wang C. Heart failure classification using deep learning to extract spatiotemporal features from ECG. BMC Medical Informatics and Decision Making. 2024. vol. 24. no. 1. pp. 1–19. DOI: 10.1186/s12911-024-02415-4.
11. Madan P., Singh V., Singh D.P., Diwakar M., Pant B., Kishor A. A Hybrid Deep Learning Approach for ECG-based Arrhythmia Classification. Bioengineering. 2022. vol. 9. no. 4. DOI: 10.3390/bioengineering9040152.
12. Huang Z., Yu J., Shan Y., Wang X. A Non-Invasive Fetal QRS Complex Detection Method Based on a Multi-Feature Fusion Neural Network. Applied Sciences. 2024. vol. 14. no. 19.
13. Teplitzky B.A., McRoberts M., Ghanbari H. Deep learning for comprehensive ECG annotation. Heart Rhythm. 2020. vol. 17. no. 5. pp. 881–888. DOI: 10.1016/j.hrthm.2020.02.015.
14. Riemann D., Krone L.B., Wulff K., Nissen C. Sleep, insomnia, and depression. Neuropsychopharmacology. 2020. vol. 45. no. 1. pp. 74–89. DOI: 10.1038/s41386-019-0411-y.
15. Paul T., Hassan O., Islam S.K., Mosa A.S.M. Real-time obstructive sleep apnea detection from raw ECG and SpO_2 signal using convolutional neural network. AMIA Joint Summits on Translational Science Proceedings. 2024. pp. 662–669.
16. Ji X., Li Y., Wen P., Barua P., Acharya U.R. MixSleepNet: A Multi-Type Convolution Combined Sleep Stage Classification Model. Computer Methods and Programs in Biomedicine. 2024. vol. 244. 107992 p. DOI: 10.1016/j.cmpb.2023.107992.
17. Mvuh F.L., Ebode Ko'a C.O.V., Bodo B. Multichannel high noise level ECG denoising based on adversarial deep learning. Scientific Reports. 2024. vol. 14. no. 1. DOI: 10.1038/s41598-023-50334-7.
18. Aversano L., Bernardi M.L., Cimitile M., Montano D., Pecori R. Characterization of Heart Diseases per Single Lead Using ECG Images and CNN-2D. Sensors. 2024. vol. 24. no. 11. DOI: 10.3390/s24113485.
19. Zacarias H., Marques J.A.L., Felizardo V., Pourvahab M., Garcia N.M. ECG Forecasting System Based on Long Short-Term Memory. Bioengineering. 2024. vol. 11. no. 1. DOI: 10.3390/bioengineering11010089.

20. Kamanditya B., Fuadah Y.N., Mahardika T.N.Q., Lim K.M. Continuous blood pressure prediction system using Conv-LSTM network on hybrid latent features of photoplethysmogram (PPG) and electrocardiogram (ECG) signals. Scientific Reports. 2024. vol. 14. no. 1.
21. Khan F., Yu X., Yuan Z., Rehman A.U. ECG classification using 1-D convolutional deep residual neural network. PLOS One. 2023. vol. 18. no. 4. e0284791 p. DOI: 10.1371/journal.pone.0284791.
22. Sawano S., Kodera S., Setoguchi N., et al. Applying masked autoencoder-based self-supervised learning for high-capability vision transformers of electrocardiographies. PLOS One. 2024. vol. 19. no. 8. pp. 1–17. DOI: 10.1371/journal.pone.0307978.
23. van de Leur R.R., de Brouwer R., Bleijendaal H., et al. ECG-only explainable deep learning algorithm predicts the risk for malignant ventricular arrhythmia in phospholamban cardiomyopathy. Heart Rhythm. 2024. vol. 21. no. 7. pp. 1102–1112. DOI: 10.1016/j.hrthm.2024.02.038.
24. Nishikimi R., Nakano M., Kashino K., Tsukada S. Variational autoencoder-based neural electrocardiogram synthesis trained by FEM-based heart simulator. Cardiovascular Digital Health Journal. 2023. vol. 5. no. 1. pp. 19–28. DOI: 10.1016/j.cvdhj.2023.12.002.
25. Sohn J., Shin H., Lee J., Kim H.C. Validation of Electrocardiogram Based Photoplethysmogram Generated Using U-Net Based Generative Adversarial Networks. Journal of Healthcare Informatics Research. 2024. vol. 8. no. 1. pp. 140–157. DOI: 10.1007/s41666-023-00156-z.
26. Nasim A., Kim Y.S. DE-PNN: Differential Evolution-Based Feature Optimization with Probabilistic Neural Network for Imbalanced Arrhythmia Classification. Sensors. 2022. vol. 22. no. 12. 4450 p. DOI: 10.3390/s22124450.
27. Din S., Qaraqe M., Mourad O., Qaraqe K., Serpedin E. ECG-based cardiac arrhythmias detection through ensemble learning and fusion of deep spatial-temporal and long-range dependency features. Artificial Intelligence in Medicine. 2024. vol. 150. 102818 p.

Lebedev Oleg — Ph.D., Dr. Sci., Associate professor, Professor, Department of Computer Science, Federal State Budgetary Educational Institution of Higher Education "MIREA – Russian Technological University" (RTU MIREA). Research interests: artificial intelligence, machine learning, swarm optimization algorithms, decision support systems, development and application of mathematical and software tools. The number of publications — 270. lebedev_o@mirea.ru; 78, Vernadsky Ave., 119454, Moscow, Russia; office phone: +7(499)215-6565.

Shmeleva Anna — Ph.D., Associate professor, Department of Computer Science, Federal State Budgetary Educational Institution of Higher Education "MIREA – Russian Technological University" (RTU MIREA). Research interests: artificial intelligence, machine learning, decision support systems. The number of publications — 107. ann_shmeleva@mail.ru; 78, Vernadsky Ave., 119454, Moscow, Russia; office phone: +7(499)215-6565.

Gezha Nikolay — Student, Department of Computer Science, Federal State Budgetary Educational Institution of Higher Education "MIREA – Russian Technological University" (RTU MIREA). Research interests: machine learning, artificial intelligence, biomedical data analysis, and decision support automation. The number of publications — 1. n.s.gezha@gmail.com; 78, Vernadsky Ave., 119454, Moscow, Russia; office phone: +7(499)215-6565.

УДК 004.008 **DOI** 10.15622/ia.25.3.5

И.В. Нетай , Е.П. Панкратов , К.И. Корнилов , А.Н. Голубинский

ИДЕНТИФИКАЦИЯ ЛИЧНОСТИ ПО КИНЕМАТИКЕ ХОДЬБЫ В ТРЁХМЕРНОМ ПРОСТРАНСТВЕ НА ОСНОВЕ ИНВАРИАНТНЫХ ПРИЗНАКОВ

Нетай И.,В., Панкратов Е.,П., Корнилов К.,И., Голубинский А.,Н. **Идентификация личности по кинематике ходьбы в трёхмерном пространстве на основе инвариантных признаков.**

Аннотация. Предлагается система сбора, предварительной обработки и анализа данных для идентификации личности по походке с использованием методов машинного обучения и анализа изображений. Система включает механизм предварительной обработки для нормализации биометрических данных движения человека и этап постобработки для выделения признаков, инвариантных относительно ортогональных преобразований пространства и изменения положения камеры, к перспективе. Реализован вычислительно эффективный метод распознавания походки с использованием одной стереокамеры. Проведен анализ пространства признаков для определения значимых характеристик, а также сравнительный анализ используемого набора признаков. Протестированы несколько архитектур машинного обучения, включая модели глубокого обучения, с анализом их точности и вычислительной эффективности. Экспериментально исследована производительность системы на различных вычислительных устройствах, измерены временные характеристики обработки данных. Результаты позволяют сравнить применимость и эффективность в зависимости от доступных вычислительных мощностей. Исходные данные формируются как временные ряды 3D-координат ключевых точек, реконструируемых по глубине на основе 2D-детекций позы (COCO, 17 точек), при этом пропуски детекций кодируются нулевым вектором. Для идентификации по походке точки головы исключаются как неинформативные, а пропуски восстанавливаются с последующим сглаживанием координат. Инвариантное признаковое пространство строится по геометрии скелета (длины сегментов, углы и их производные). Далее признаки агрегируются в единый вектор по скользящему окну, то есть каждый вектор описывает движение на интервале из нескольких последовательных кадров. На этапе идентификации личности используется фильтрация по порогу и агрегирование предсказаний на уровне кадров. Показано, что качество и скорость работы зависят от выбранной модели машинного обучения и доступных вычислительных ресурсов.

Ключевые слова: машинное обучение, компьютерное зрение, биометрия, нейронные сети, кинематика, 3D-видеобработка, инвариантные признаки

1. Введение. Результаты измерений крупной и мелкой моторики человека имеют широкий спектр применения в различных областях. Наиболее распространены маркерные системы измерений движений, используемые в компьютерной графике, биометрической идентификации и верификации, медицинской диагностике [1] и спортивной науке.

Маркерные системы обладают ограничениями [2]. Они требуют размещения отражающих индикаторов на теле, калибровки нескольких камер и контролируемой среды. Для работы, как

правило, требуются высокоскоростные камеры, специализированное программное обеспечение (ПО) и вычислительные ресурсы. Точность зависит от регулярной калибровки, что увеличивает трудоёмкость их эксплуатации.

Безмаркерные системы анализа движений уменьшают влияние некоторых недостатков. Хотя они по-прежнему уступают маркерным системам в точности, современные алгоритмы компьютерного зрения и глубокого обучения позволяют сократить этот разрыв благодаря развитию технологий трёхмерной оценки позы [3].

В данной работе предлагается метод безмаркерного анализа походки для идентификации личности с акцентом на построение инвариантного 3D-пространства признаков, отражающего уникальные кинематические характеристики каждого испытуемого. Предлагается система, сочетающая точность и простоту использования, с минимальными требованиями к оборудованию для съемки (одна стереокамера).

Предлагаемый метод основан на обработке видеозаписи походки с использованием моделей машинного обучения для оценки позы [4]. После записи и предобработки данных получаются 3D-координаты ключевых точек тела, что обеспечивает достаточно точное представление кинематики движений. Дальнейшая обработка включает выделение признаков из координат. Для обучения моделей вычисляются геометрические и кинематические ортогонально инвариантные признаки [5].

В отличие от более распространенных методов, работающих с 2D-данными или упрощенными 3D-моделями [6], данный подход обеспечивает пространственное и инвариантное представление движений. Использование ортогонально инвариантных параметров исключает возможность влияния на результаты положения и направления, угла съёмки камеры. Это потенциально повышает надёжность методов, применяемых в биометрической идентификации, медицинской диагностике и анализе двигательных нарушений.

2. Подготовка данных. Экспериментальные исследования проводились на собственном наборе видеоданных движений человека, разделённом на обучающую, валидационную и тестовую части. В эксперименте участвовали 19 испытуемых. Для каждого испытуемого было записано по три стереовидеофрагмента: две записи использовались для формирования обучающей и валидационной выборок, третья запись – в качестве тестовой. Таким образом, общий объём набора данных составил 57 видеозаписей (19×3).

Съёмка выполнялась стереоскопической камерой Intel RealSense Depth Camera D457 при разрешении 1280×720 и частоте 30 кадров/с [8]. Типичная продолжительность каждой из двух обучающих записей составляла около 90 с на одного испытуемого, тестовой записи – около 45 с. Во всех записях испытуемый двигался обычной походкой по освещённому коридору длиной 13 м, не приближаясь к камере ближе чем на 2 м. Для повышения воспроизводимости измерений и при этом обеспечения естественной вариативности каждая из трёх записей выполнялась в отдельную сессию через 2-5 недель после предыдущей; при этом допускались естественные изменения внешнего вида (например, одежда), а также небольшие вариации условий съёмки, неизбежные при повторной записи (незначительные изменения освещённости и фона, небольшие отличия в положении/наклоне камеры при установке). Такое разделение обеспечивало проверку устойчивости метода к межсессионным изменениям при сохранении сопоставимости условий [7].

Принцип разделения на обучающую/валидационную/тестовую части. Тестовая выборка формировалась целиком из третьей (отложенной) видеозаписи каждого испытуемого и не использовалась ни при настройке параметров, ни при выборе модели.

Для оценки обобщающей способности модели и выбора гиперпараметров применялась 5-кратная стратифицированная кросс-валидация. Каждая из двух обучающих видеозаписей каждого испытуемого разбивалась на 5 последовательных временных фрагментов (фолдов), внутри которых кадры сохраняли исходный временной порядок; в совокупности это давало 10 фолдов на одного испытуемого. Распределение фолдов между обучающей и валидационной подвыборками выполнялось с перемешиванием ($\text{seed} = 42$): в каждой итерации 2 из 10 фолдов составляли валидационную подвыборку, остальные 8 использовались для обучения. Сохранение временного порядка внутри фолдов уменьшает риск утечки информации между выборками при наличии сильной временной корреляции соседних кадров и при использовании скользящих окон, а перемешивание фолдов обеспечивает проверку устойчивости классификации: модель должна демонстрировать стабильное качество независимо от того, какие участки записей (начало, середина или конец движения) используются для обучения, а какие – для валидации.

Валидационная выборка применялась для:

– подбора и сравнения вариантов предварительной обработки (в частности, способов заполнения пропусков и параметров сглаживания);

– выбора архитектуры и гиперпараметров модели;

– настройки эвристик принятия решения (например, порога уверенности T при фильтрации кадров/окон);

– контроля переобучения при обучении нейросетевых моделей (ранняя остановка, планировщик шага обучения).

После фиксации всех настроек итоговая оценка качества проводилась однократно на тестовой выборке.

Процедура видеосъемки. Для обеспечения качества и сопоставимости данных соблюдались следующие условия видеосъемки:

– место съёмки должно быть хорошо освещено, в случае наличия отражающих поверхностей порог уверенности детектора YOLO (параметр `conf`) повышается до $\mathrm{conf} = 0.5$ для исключения ложных детекций (отражений);

– присутствие посторонних лиц в кадре во время формирования обучающих данных исключается; используемый метод рассчитан на распознавание одного человека в кадре;

– испытуемый должен на всём протяжении записи оставаться в пределах поля зрения видеокамеры; все части тела должны непрерывно находиться в кадре;

– движение должно начинаться при полностью видимом теле в кадре (испытуемый начинает движение, находясь целиком в поле зрения камеры);

– длительность обучающей записи должна быть не менее 30 с;

Указанные условия видеосъёмки обеспечивают получение устойчивых трёхмерных данных, необходимых для построения моделей движения и корректной оценки качества предложенного метода.

3. Последовательность обработки данных. Последовательность этапов обработки данных реализована в виде конвейера (Рис. 1) и получает на вход стереовидеозапись с камеры Intel RealSense D457, включающую RGB-поток (формат mp4) и синхронизированные данные глубины.

На первом этапе из RGB-потока формируется двумерный видеоряд, к которому применяется предобученная модель YOLOv11x-pose [9–11] для оценки позы и извлечения двумерных координат 17 ключевых точек в каждом кадре. Модель использовалась в режиме процедуры вывода без дообучения, с весами, предоставленными авторами библиотеки Ultralytics (предобучение на наборе данных COCO-pose). Все параметры детектора сохранены по умолчанию, за исключением порога уверенности, установленного на уровне $\mathrm{conf} = 0.3$ для обеспечения приемлемого баланса между полнотой обнаружения и уровнем ложных срабатываний в условиях реальной видеосъемки. Исходные кадры разрешением

1280×720 автоматически масштабировались к внутреннему размеру модели 640 пикселей по длинной стороне с сохранением пропорций; порог немаксимального подавления (NMS) $\mathrm{iou} = 0.7$. Координаты обнаруженных ключевых точек пересчитывались обратно в исходное разрешение.

На втором этапе для каждой обнаруженной суставной точки восстанавливаются трёхмерные координаты. Камера Intel RealSense D457 одновременно записывает RGB-поток и синхронизированный стереопоток, содержащий трёхмерные координаты каждого пикселя в системе координат камеры. Для перехода от двумерных детекций к трёхмерным координатам выполняется сопоставление пиксельных координат (u, v) суставных точек, полученных моделью YOLOv11x-pose из RGB-кадров, с соответствующими пикселями стереопотока, из которого извлекаются трёхмерные координаты (x, y, z).

Пространственное соответствие между RGB- и стереопотоками обеспечивается заводской калибровкой камеры: внутренние и внешние параметры хранятся в энергонезависимой памяти устройства, и выравнивание потоков выполняется автоматически средствами Intel RealSense SDK 2.0 [12]. Дополнительная калибровка камеры пользователем не проводилась.

Итоговое представление данных имеет вид массива размерности $N_{\text{кадров}} \times 17 \times 3$, содержащего трёхмерные координаты ключевых точек.

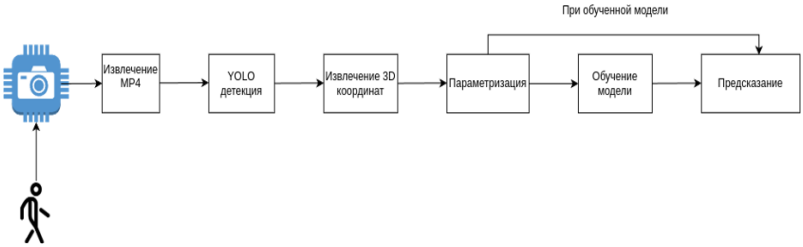


Рис. 1. Последовательность вычислений

После восстановления трёхмерных координат ключевых точек выполняется предварительная обработка, в ходе которой формируются инвариантные признаки: углы между смежными рёбрами графа ключевых точек, скорости, ускорения, длины градиентов и их первые и вторые производные. Для идентификации личности по построенным признакам используется многослойный персептрон с двумя скрытыми слоями, обеспечивающий удовлетворительное качество при малой вычислительной задержке и поддерживающий обработку в реальном

времени. При применении обученной модели выполняется та же предварительная обработка, после чего классификатор определяет личность по кинематическим признакам и выдаёт оценку уверенности. Далее приводится детальное описание всех этапов параметризации.

3.1. Предобработка трёхмерных координат и формирование признакового пространства. Исходные данные представляют собой временной ряд трёхмерных координат ключевых точек, полученных путём восстановления по данным глубины на основе двумерных детекций модели YOLOv11x-pose [11]. Данные организованы в виде массива размерности $N_{\text{кадров}} \times 17 \times 3$ в соответствии со схемой COCO (17 ключевых точек). Для задачи идентификации по походке исключаются ключевые точки головы (нос, глаза, уши) как неинформативные; в дальнейшем используется подмножество из 12 точек: плечи, локти, запястья, бёдра, колени и лодыжки. Исходные траектории содержат шум, обусловленный как внутренней вариативностью оценок детектора YOLO, так и внешними факторами, влияющими на качество детектирования [13]. Пропуски в данных (кодируемые вектором $[0, 0, 0]$) возникают в тех кадрах, где детектор не смог с достаточной уверенностью локализовать соответствующие ключевые точки. Такие случаи могут возникать при поворотах во время движения.

Предварительная обработка включает восстановление пропущенных значений, сглаживание координат, вычисление геометрических и кинематических признаков, а также их первых и вторых производных; далее выполняется повторное сглаживание, направленное на повышение устойчивости признаков. Для обеспечения инвариантности к жёстким преобразованиям сцены (поворотам, отражениям и переносам), а также к изменению положения и ориентации камеры, признаки формируются на основе относительных расстояний, углов и норм временных градиентов. Абсолютные координаты ключевых точек не используются, поскольку они зависят от глобального положения человека и камеры и, следовательно, не являются инвариантными.

3.2. Заполнение пропущенных значений. Их доля в общем объёме трёхмерных координат (после исключения ключевых точек головы – носа, глаз и ушей; используется подмножество из $K = 12$ точек) составляет 2.24%. Распределение пропусков неравномерно: наибольший процент наблюдается для суставных точек верхних конечностей – запястий (до 8.5%) и локтей (до 5.7%), что обусловлено махами рук во время ходьбы; для точек тазобедренных суставов и коленей доля пропусков не превышает 0.3%. Подробное распределение по ключевым точкам приведено в таблице 1.

Таблица 1. Доля пропусков детекции по ключевым точкам тела

Ключевая точка	Пропусков (кадров)	% от кадров
Левое плечо	283	0.77
Правое плечо	122	0.33
Левый локоть	2099	5.69
Правый локоть	1198	3.24
Левое запястье	3135	8.49
Правое запястье	2268	6.14
Левый тазобедр. сустав	81	0.22
Правый тазобедр. сустав	62	0.17
Левое колено	113	0.31
Правое колено	72	0.20
Левая лодыжка	236	0.64
Правая лодыжка	243	0.66

Далее рассматриваются пропуски в трёхмерных координатах ключевых точек, которые кодируются вектором $[0, 0, 0]$. Ключевые точки головы (нос, глаза, уши) исключаются из обработки как малозначимые для идентификации по походке. Обозначим через $\mathbf{p}_{t,k} \in \mathbb{R}^3$ координаты k-й точки в кадре t, где $t = 1, \ldots, T$, $k = 1, \ldots, K$. Введём бинарный индикатор достоверности координаты:

$$m_{t,k} = \begin{cases} 1, & \mathbf{p}_{t,k} \neq \mathbf{0}, \\ 0, & \mathbf{p}_{t,k} = \mathbf{0}, \end{cases}$$

где $\mathbf{0} = (0, 0, 0)^\top$. Восстановление выполнялось только для тех пар (t, k), для которых $m_{t,k} = 0$, с использованием одним из методов, описанных далее.

Базовый метод (forward fill). Наибольшую точность показало заполнение предыдущим валидным значением:

$$\mathbf{p}_{t,k} \leftarrow \begin{cases} \mathbf{p}_{t-1,k}, & \text{если } m_{t,k} = 0 \text{ и } m_{t-1,k} = 1, \\ \mathbf{p}_{t,k}, & \text{иначе}, \end{cases} \qquad t = 2, \ldots, T.$$

Для $t = 1$ заполнение не выполняется (первый кадр используется как исходный).

Альтернативные тестировавшиеся методы.

– **Кубическая сплайн-интерполяция по времени.** Для каждой ключевой точки k и каждой координаты $c \in \{x, y, z\}$ рассматривается одномерный временной ряд $\{p_{t,k}^{(c)}\}_{t=1}^{T}$. Обозначим через $V_k = \{t_1 < \ldots < t_m\}$ множество индексов кадров, где значение не равно $[0, 0, 0]$. Если $m \geqslant 2$, по точкам $\left(t_i,\, p_{t_i,k}^{(c)}\right)$ строится кубический сплайн $f_k^{(c)}(t)$, и для каждого пропущенного кадра $t_p \notin V_k$ координата восстанавливается как $p_{t_p,k}^{(c)} = f_k^{(c)}(t_p)$ (при $m < 2$ интерполяция для данной компоненты не выполняется).

– **KNN-заполнение по кадрам.** Каждый кадр t представляется плоским вектором $\mathbf{x}_t \in \mathbb{R}^D$, $D = 3K$, полученным конкатенацией координат всех ключевых точек: $\mathbf{x}_t = \left[\mathbf{p}_{t,1}^{\top}, \ldots, \mathbf{p}_{t,K}^{\top}\right]^{\top}$. Введём маску валидности координат в плоском представлении $\mathbf{q}_t \in \{0, 1\}^D$: для каждой точки k три соответствующие компоненты $\mathbf{q}_t$ равны $m_{t,k}$. Для кадра t выбираются n ближайших кадров $\{t_j\}_{j=1}^{n}$ по расстоянию:

$$d_j = \left\| \left(\mathbf{q}_t \odot \mathbf{q}_{t_j}\right) \odot \left(\mathbf{x}_t - \mathbf{x}_{t_j}\right) \right\|_2,$$

где $\odot$ – покомпонентное произведение. Затем отсутствующее значение $\mathbf{p}_{t,k}$ восстанавливается как взвешенное среднее по этим соседям [14]:

$$\mathbf{p}_{t,k} \leftarrow \frac{\sum_{j=1}^{n} w_j\, \mathbf{p}_{t_j,k}}{\sum_{j=1}^{n} w_j}, \qquad w_j = \frac{1}{d_j + \varepsilon},$$

где $n \in \mathbb{N}$ – число соседей, $\varepsilon > 0$ – малая константа для предотвращения деления на ноль.

– **Скользящее среднее по времени.** Пусть $w \in \mathbb{N}$ – ширина окна (в кадрах), $\lfloor \cdot \rfloor$ – целая часть. Для кадра t рассматривается окно индексов:

$$\mathcal{W}_t = \{t - \lfloor w/2 \rfloor, \ldots, t + \lfloor w/2 \rfloor\} \cap \{1, \ldots, T\}.$$

Из $\mathcal{W}_t$ выбираются только кадры с валидным значением точки k: $V = \{v \in \mathcal{W}_t : m_{v,k} = 1\}$. Пропуск в кадре t заполняется средним по этим значениям:

$$\mathbf{p}_{t,k} \leftarrow \frac{1}{|V|} \sum_{v \in V} \mathbf{p}_{v,k},$$

где $|V|$ – мощность множества V. В отличие от линейной интерполяции между двумя граничными точками пропуска, данный метод усредняет все доступные значения в окрестности и действует как локальное сглаживание траектории.

– **Итеративное модельное заполнение.** Последовательность рассматривается как матрица $\mathbf{X} \in \mathbb{R}^{T \times D}$ («кадры $\times$ координаты»), $D = 3K$, где пропуски заменены на NaN. Для каждого столбца $f \in \{1, \ldots, D\}$ с пропусками обучается регрессионная модель R_f, предсказывающая значения этого столбца по остальным признакам $\mathbf{X}_{-f}$ (матрица $\mathbf{X}$ без f-го столбца). Итерации продолжаются до сходимости по критерию:

$$\left\|\mathbf{X}^{(i)} - \mathbf{X}^{(i-1)}\right\|_F < \delta,$$

где $\mathbf{X}^{(i)}$ – матрица после i-й итерации, $\| \cdot \|_F$ – норма Фробениуса, $\delta > 0$ – порог остановки. В реализации в качестве R_f использовалось дерево решений, т.к. позволяет учитывать нелинейные взаимосвязи между координатами.

Выбор в рабочей конфигурации. Экспериментальная проверка показала, что на валидационной и тестовой выборках наибольшая точность идентификации достигается при использовании базового каузального способа заполнения пропусков (перенос последнего корректного значения вперёд, *forward fill*). Данный способ в меньшей степени вносит фазовые и амплитудные искажения, не приводит к переусреднению траектории, вычислительно малозатратен и не вызывает избыточного сглаживания данных.

Распределение длин непрерывных пропусков. Для оценки структуры пропусков была проанализирована длина каждого непрерывного пропуска – число последовательных кадров, в которых данная ключевая точка не была детектирована. Пусть для точки k на участке кадров $[t_{\text{start}}, t_{\text{end}}]$ выполняется $m_{t,k} = 0$ (пропуск), а в кадрах $t_{\text{start}} - 1$ и $t_{\text{end}} + 1$ значения валидны. Тогда длина пропуска определяется как:

$$g = t_{\text{end}} - t_{\text{start}} + 1. \quad (1)$$

Из 1374 зафиксированных непрерывных пропусков 37.2% составляют одиночные кадры ($g = 1$), медианная длина – 3 кадра, средняя – 7.2 кадра, 95-й перцентиль – 22 кадра. Максимальная длина пропуска составила 256 кадров (1 случай из 1374, связанный с длительным поворотом испытуемого во время движения). При частоте дискретизации 30 кадров/с одиночный пропуск соответствует интервалу 33 мс, за который координаты ключевой точки изменяются незначительно, что обосновывает выбор forward fill в качестве базового метода заполнения.

Абсолютные длины сегментов скелета. Для оценки качества трёхмерной реконструкции были вычислены средние длины сегментов скелетного графа $G = (V, E)$ в метрах. Пусть $e = (i, j) \in E$ – ребро скелета. Длина сегмента e в кадре t (после заполнения пропусков, до сглаживания) определяется как:

$$L_t^{(e)} = \|\mathbf{p}_{t,i} - \mathbf{p}_{t,j}\|_2 . \tag{2}$$

Из рассмотрения исключались кадры с нефизичными значениями длин ($L_t^{(e)} < 0.05$ м или $L_t^{(e)} > 2.0$ м), обусловленными выбросами стереореконструкции глубины. Обозначим множество валидных кадров для сегмента e и записи s как $\mathcal{V}_s^{(e)}$.

Среднее значение длины сегмента для записи s:

$$\bar{L}_s^{(e)} = \frac{1}{\left|\mathcal{V}_s^{(e)}\right|} \sum_{t \in \mathcal{V}_s^{(e)}} L_t^{(e)}. \tag{3}$$

Для испытуемых с несколькими записями значения $\bar{L}_s^{(e)}$ предварительно усреднялись:

$$\bar{L}_u^{(e)} = \frac{1}{|S_u|} \sum_{s \in S_u} \bar{L}_s^{(e)}, \tag{4}$$

где S_u – множество записей испытуемого u. Итоговые статистики вычислялись по $U = 19$ испытуемым. Среднее значение длины сегмента:

$$\mathrm{Mean}^{(e)} = \frac{1}{U}\sum_{u=1}^{U} \bar{L}_u^{(e)}. \tag{5}$$

Межсубъектное среднеквадратическое отклонение, характеризующее антропометрический разброс:

$$\mathrm{Std}^{(e)} = \sqrt{\frac{1}{U}\sum_{u=1}^{U}\left(\bar{L}_u^{(e)} - \mathrm{Mean}^{(e)}\right)^2}. \tag{6}$$

Результаты приведены в таблице 2. Средние длины сегментов соответствуют антропометрическим нормам: голень – 0.44 м, бедро – 0.45 м, торс (плечо–бедро) – 0.55 м, плечо – 0.32 м, предплечье – 0.29 м. Межсубъектное СКО составляет 0.03–0.04 м, что отражает естественный антропометрический разброс между испытуемыми и свидетельствует о том, что длины сегментов скелета несут индивидуальную идентификационную информацию.

Таблица 2. Средние длины сегментов скелета (м) (до сглаживания координат). Mean – среднее значение длины, усреднённое по испытуемым; Std – СКО средних длин между испытуемыми (антропометрический разброс)

Сегмент	Mean (м)	Std (м)
Лодыжка–колено (лев.)	0.442	0.034
Колено–бедро (лев.)	0.445	0.029
Лодыжка–колено (прав.)	0.442	0.033
Колено–бедро (прав.)	0.445	0.032
Бедро–бедро (таз)	0.233	0.033
Плечо–бедро (лев.)	0.547	0.037
Плечо–бедро (прав.)	0.547	0.042
Плечо–плечо	0.346	0.037
Плечо–локоть (лев.)	0.319	0.028
Плечо–локоть (прав.)	0.320	0.028
Локоть–запястье (лев.)	0.294	0.038
Локоть–запястье (прав.)	0.293	0.031

Оценка искажения длины сегментов. Для количественной оценки влияния заполнения пропусков на восстановление траекторий были рассчитаны относительные изменения длин сегментов скелета после

заполнения. Пусть $G = (V, E)$ – скелетный граф ключевых точек без головы, $|V| = K = 12$, а рёбра $E = \{(u_i, v_i)\}_{i=1}^{N}$, $N = 12$, задают сегменты скелета. Длину i-го сегмента в кадре t определим как:

$$L_t^i = \left\| \mathbf{p}_{t,u_i} - \mathbf{p}_{t,v_i} \right\|_2, \qquad t = 1, \ldots, T.$$

Относительное изменение длины сегмента вычисляется по последовательным кадрам:

$$\zeta_t^i = \frac{L_t^i - L_{t-1}^i}{L_{t-1}^i}, \qquad t = 2, \ldots, T.$$

Значения ζ_t^i учитывались только в тех кадрах, где обе длины определены, т.е. $m_{t,u_i} = m_{t,v_i} = m_{t-1,u_i} = m_{t-1,v_i} = 1$. Обозначим множество таких индексов кадров через:

$$\mathcal{T}_i = \left\{ t \in \{2, \ldots, T\} : \; m_{t,u_i} m_{t,v_i} m_{t-1,u_i} m_{t-1,v_i} = 1 \right\}, \qquad T_i = |\mathcal{T}_i|.$$

По всем полученным значениям $\{\zeta_t^i : \; t \in \mathcal{T}_i, \; i = 1, \ldots, N\}$ рассчитаны статистические характеристики (для способа *forward fill*, показавшего наилучшие результаты при классификации):

$$\mu = \frac{1}{\sum_{i=1}^{N} T_i} \sum_{i=1}^{N} \sum_{t \in \mathcal{T}_i} \zeta_t^i = -7.8 \times 10^{-5},$$

$$\sigma^2 = \frac{1}{\sum_{i=1}^{N} T_i} \sum_{i=1}^{N} \sum_{t \in \mathcal{T}_i} \left(\zeta_t^i - \mu \right)^2 = 1.34 \times 10^{-4},$$

где μ – среднее относительное изменение, σ^2 – дисперсия. Малое значение μ указывает на отсутствие выраженного систематического смещения длин сегментов после заполнения пропусков. Дисперсия σ^2 (стандартное отклонение $\sigma \approx 0.0116$) характеризует высокую устойчивость восстановленных траекторий: в большинстве случаев относительные изменения длин сегментов не превышают $\pm 1.2\%$.

3.3. Сглаживание координат. После заполнения пропусков применяется фильтр экспоненциального забывания. Сглаживание выполняется покадрово для каждой ключевой точки и каждой координаты:

$$p'_{t,j,c} = \alpha\, p_{t,j,c} + (1-\alpha)\, p'_{t-1,j,c}, \quad t = 2, \ldots, T,\ j = 1, \ldots, K,\ c \in \{x, y, z\},$$

где T – число кадров в последовательности, $K = 12$ – число используемых ключевых точек, $p_{t,j,c}$ – исходная координата компоненты c точки j в кадре t, $p'_{t,j,c}$ – сглаженная координата, $\alpha \in (0, 1)$ – коэффициент сглаживания (в первичном сглаживании использовалось $\alpha = 0.007$). Инициализация задаётся как $p'_{1,j,c} = p_{1,j,c}$.

Для компенсации начальной предвзятости при коротких фрагментах может применяться поправка:

$$\widehat{p}'_{t,j,c} = \frac{p'_{t,j,c}}{1 - (1-\alpha)^t}, \quad t = 1, \ldots, T.$$

Связь коэффициента сглаживания с частотой среза. Связь α с постоянной времени τ при частоте кадров f_s ($T_s = 1/f_s$) задаётся:

$$\tau = -\frac{T_s}{\ln(1-\alpha)}, \qquad f_c \approx \frac{1}{2\pi\tau} = -\frac{\ln(1-\alpha)}{2\pi T_s} = -\frac{f_s\, \ln(1-\alpha)}{2\pi}.$$

Значения коэффициентов сглаживания определены исходя из физических свойств фильтруемых величин. При $f_s = 30$ кадр/с и $\alpha = 0{,}007$ получаем $\tau \approx 4{,}74$ с и $f_c \approx 0{,}034$ Гц. Данная частота среза существенно ниже частоты шагового цикла (0,8-1,0 Гц). Такой выбор параметра является намеренным: целью первичного сглаживания является не сохранение временной динамики координат, а стабилизация геометрии скелета для последующего извлечения инвариантных признаков – длин сегментов и углов между рёбрами скелетного графа. Для жёстких сегментов скелета длины должны быть приблизительно постоянными во времени, и агрессивное сглаживание координат подавляет покадровый шум 3D-реконструкции, обусловленный погрешностями оценки глубины камерой Intel RealSense D457 и вариативностью детекций позы. Динамическая информация о походке кодируется на следующем

этапе – через временные производные извлечённых признаков (углов и длин).

Дополнительно, после вычисления признаков применялось сглаживание признаковых рядов с коэффициентом $\alpha_f = 0{,}05$ (при той же частоте дискретизации), для которого $\tau \approx 0{,}65$ с и $f_c \approx 0{,}245$ Гц. Поскольку признаки извлекаются из уже сглаженных координат, остаточный шум в них невелик и достаточно мягкой фильтрации; при этом данный фильтр сохраняет динамику шагового цикла.

Таким образом, двухуровневая схема сглаживания реализует разделение задач: первое сглаживание ($f_c \approx 0{,}034$ Гц) стабилизирует геометрию скелета для извлечения инвариантных признаков, второе ($f_c \approx 0{,}245$ Гц) стабилизирует сами признаки при сохранении временной динамики движений.

Также были протестированы свёрточные методы: окно Ханна, медианный фильтр и скользящее среднее. Оконные свёртки требуют центрирования окна (что повышает задержку), и некоторые, использующие оконные функции с разрывами, подвержены краевым искажениям; медианный фильтр эффективен против импульсного шума, но искажает гладкие траектории и их производные. Фильтр экспоненциального забывания показал наилучший баланс между подавлением шумов и сохранением динамики: оно инвариантно по времени, линейно и даёт малую задержку при заданной степени сглаживания.

3.4. Формирование признаков. Первоначально рассматривался вариант прямого использования сглаженных трёхмерных координат ключевых точек без явной параметризации. Пусть $\mathbf{p}_{t,j} \in \mathbb{R}^3$ – вектор координат j-й ключевой точки в кадре t, где $t = 1, \ldots, T$, $j = 1, \ldots, K$. Результат, полученный после заполнения пропусков и сглаживания, обозначим $\mathbf{p}'_{t,j}$. Для базового варианта по последовательности $\{\mathbf{p}'_{t,j}\}_{t=1}^{T}$ формировались скользящие окна ширины L:

$$\mathbf{W}_t = \big[\mathbf{p}'_t,\, \mathbf{p}'_{t+1},\, \ldots,\, \mathbf{p}'_{t+L-1}\big],$$

где $\mathbf{p}'_t = [(\mathbf{p}'_{t,1})^\top, \ldots, (\mathbf{p}'_{t,K})^\top]^\top \in \mathbb{R}^{3K}$. Полученные окна подавались на вход классификатора в предположении, что модель самостоятельно выделит информативные характеристики движения. На практике такой способ представления обеспечил точность не выше 25%, что послужило основанием для перехода к явному построению инвариантных геометрико-кинематических признаков.

Далее используется скелетный граф ключевых точек $G = (V, E)$, заданный схемой COCO, с подмножеством точек без головы $|V| = K = 12$ и множеством связей E ($|E| = N_E = 12$ рёбер). Дополнительно введём множество троек $\mathcal{A} \subseteq V^3$ для вычисления углов между смежными рёбрами; $|\mathcal{A}| = N_\theta = 16$, а каждая тройка $(i, j, r) \in \mathcal{A}$ соответствует углу при вершине j между рёбрами (i, j) и (r, j).

Набор признаков. Признаки вычисляются покадрово, после чего агрегируются во временные окна (раздел 3.5).

1. **Длины сегментов.** Для каждого ребра $e = (i, j) \in E$ и кадра t длина сегмента определяется как:

$$\ell_{t,e} = \left\|\mathbf{p}'_{t,i} - \mathbf{p}'_{t,j}\right\|_2,$$

что даёт $N_E = 12$ признаков на кадр.

2. **Углы между смежными рёбрами.** Для каждой тройки $(i, j, r) \in \mathcal{A}$ положим $\mathbf{u}_{t,(i\to j)} = \mathbf{p}'_{t,i} - \mathbf{p}'_{t,j}$ и $\mathbf{u}_{t,(r\to j)} = \mathbf{p}'_{t,r} - \mathbf{p}'_{t,j}$. Тогда:

$$c_{t,(i,j,r)} = \frac{\mathbf{u}_{t,(i\to j)} \cdot \mathbf{u}_{t,(r\to j)}}{\|\mathbf{u}_{t,(i\to j)}\|_2 \, \|\mathbf{u}_{t,(r\to j)}\|_2}, \qquad \theta_{t,(i,j,r)} = \arccos\bigl(c_{t,(i,j,r)}\bigr),$$

что даёт $N_\theta = 16$ признаков на кадр.

3. **Динамические величины.** Для учёта динамики вводятся центральные разностные операторы первого и второго порядков по времени, применяемые к каждой точке покоординатно:

$$\Delta_1 \mathbf{p}'_{t,j} = \frac{1}{2}\bigl(\mathbf{p}'_{t+1,j} - \mathbf{p}'_{t-1,j}\bigr), \qquad \Delta_2 \mathbf{p}'_{t,j} = \mathbf{p}'_{t+1,j} - 2\mathbf{p}'_{t,j} + \mathbf{p}'_{t-1,j},$$

где $t \in \{2, \ldots, T-1\}$, $j = 1, \ldots, K$.

На границах $t \in \{1, T\}$ значения $\Delta_1 \mathbf{p}'_{t,j}$ и $\Delta_2 \mathbf{p}'_{t,j}$ задаются копированием ближайшего определённого значения (из кадра $t = 2$ или $t = T - 1$).

– *«Скорости» и «ускорения» сегментов.* Для каждого ребра $e = (i, j) \in E$ и $r \in \{1, 2\}$ определим:

$$\ell_{t,e}^{(r)} = \left\| \Delta_r \mathbf{p}'_{t,i} - \Delta_r \mathbf{p}'_{t,j} \right\|_2,$$

что даёт $N_E + N_E = 24$ признака на кадр.

– *Производные углов.* Вектор углов в кадре обозначим $\boldsymbol{\theta}_t \in \mathbb{R}^{N_\theta}$, где компоненты $\boldsymbol{\theta}_t$ равны $\theta_{t,(i,j,r)}$ для $(i,j,r) \in \mathcal{A}$. Тогда разностные аппроксимации первого и второго порядков по времени вычисляются как $\Delta_1 \boldsymbol{\theta}_t$ и $\Delta_2 \boldsymbol{\theta}_t$, что добавляет $N_\theta + N_\theta = 32$ признака на кадр.

4. **Маски и финальное сглаживание.** Пусть $m_{t,j} \in \{0,1\}$ – индикатор валидности координаты ключевой точки (раздел 3.2). Для каждого признака f в кадре t вводится маска $m_{t,f} \in \{0,1\}$, равная произведению индикаторов $m_{\tau,j}$ по всем точкам j и кадрам τ, которые используются при вычислении данного признака в кадре t. Например, для длины сегмента $e = (i,j)$ в кадре t: $m_{t,\ell_e} = m_{t,i} m_{t,j}$; для угла (i,j,r): $m_{t,\theta_{(i,j,r)}} = m_{t,i} m_{t,j} m_{t,r}$; для динамических признаков с центральной разностью в кадре t: дополнительно требуется валидность в кадрах $t-1$ и $t+1$.

Непосредственно перед финальным сглаживанием недостоверные компоненты зануляются покомпонентно:

$$\tilde{\mathbf{x}}_t = \mathbf{m}_t \odot \mathbf{x}_t, \qquad \mathbf{m}_t = [m_{t,1}, \ldots, m_{t,F}]^\top,$$

где $\mathbf{x}_t \in \mathbb{R}^F$ – вектор всех признаков кадра t, F – размерность признакового пространства. Далее к непрерывным признаковым рядам применяется фильтр экспоненциального забывания:

$$z'_{t,f} = \alpha_f \, \tilde{z}_{t,f} + (1 - \alpha_f) \, z'_{t-1,f}, \qquad f = 1, \ldots, F, \quad t = 2, \ldots, T,$$

с коэффициентом $\alpha_f = 0.05$ покомпонентно и инициализацией $z'_{1,f} = \tilde{z}_{1,f}$. Разностные ряды $\Delta_1(\cdot)$ и $\Delta_2(\cdot)$ вычисляются по времени из уже сглаженных базовых последовательностей.

Итоговое пространство признаков имеет размерность $F = 116$ (12 длин сегментов + 16 углов + 24 разностных признака длин + 32 разностных признака углов + 32 скалярные связи) и, за счёт использования относительных расстояний, углов и скалярных произведений, инвариантно к ортогональным преобразованиям и сдвигам сцены.

3.5. Оконная обработка для учёта временной динамики. Для учёта временной структуры движений применялись схемы скользящих окон, обеспечивающие агрегирование покадровых признаков во временные фрагменты фиксированной длины [15]. Пусть $\mathbf{x}_t \in \mathbb{R}^F$ – вектор из F признаков, вычисленных для кадра $t = 1, \ldots, T$. Маска валидности признаков $\mathbf{m}_t \in \{0,1\}^F$ определяется по индикаторам валидности координат $m_{t,j}$ (раздел 3.2) следующим образом: компонент $m_{t,f} = 1$ тогда и только тогда, когда для вычисления f-го признака в кадре t доступны (валидны) все требуемые ключевые точки и, при необходимости, соседние кадры (например, $t-1$ и $t+1$ для центральных разностей). При отсутствии масок полагаем $\mathbf{m}_t = \mathbf{1}$. Далее используется вектор $\tilde{\mathbf{x}}_t = \mathbf{m}_t \odot \mathbf{x}_t$, где $\odot$ – покомпонентное произведение.

Плотное скользящее окно (dense sliding window). Пусть $L \in \mathbb{N}$ – длина окна (в кадрах), $h \in \mathbb{N}$ – шаг между окнами. Для $t \in \{1,\, 1+h,\, \ldots,\, 1+(N_{\text{win}}^{\text{dense}}-1)h\}$ определим окно:

$$\mathbf{W}_t^{\text{dense}} = \big[\tilde{\mathbf{x}}_t, \tilde{\mathbf{x}}_{t+1}, \ldots, \tilde{\mathbf{x}}_{t+L-1}\big] \in \mathbb{R}^{F\times L}, \qquad \mathbf{w}_t^{\text{dense}} = \operatorname{vec}\big(\mathbf{W}_t^{\text{dense}}\big) \in \mathbb{R}^{LF},$$

где $\operatorname{vec}(\cdot)$ – оператор векторизации (конкатенация столбцов матрицы). Число окон равно:

$$N_{\text{win}}^{\text{dense}} = \left\lfloor \frac{T-L}{h} \right\rfloor + 1,$$

а определение корректно при условии $T \geqslant L$.

Прореженное окно (sparse sliding window). Пусть $n \in \mathbb{N}$ – число отобранных кадров в окне, $s \in \mathbb{N}$ – шаг прореживания внутри окна. Для начала окна в кадре t вводятся индексы:

$$\mathcal{I}_t = \{t,\, t+s,\, \ldots,\, t+(n-1)s\}, \qquad t+(n-1)s \leqslant T.$$

Тогда:

$$\mathbf{W}_t^{\text{dec}} = \big[\tilde{\mathbf{x}}_t, \tilde{\mathbf{x}}_{t+s}, \ldots, \tilde{\mathbf{x}}_{t+(n-1)s}\big] \in \mathbb{R}^{F\times n}, \qquad \mathbf{w}_t^{\text{dec}} = \operatorname{vec}\big(\mathbf{W}_t^{\text{dec}}\big) \in \mathbb{R}^{nF}.$$

При сдвиге начала окна на один кадр число таких окон равно:

$$N_{\text{win}}^{\text{dec}} = T - (n-1)s,$$

а определение корректно при условии $T \geqslant (n-1)s + 1$.

По результатам экспериментов на валидационной и тестовой выборках наилучшие показатели продемонстрировало плотное скользящее окно шириной $L = 30$ и шагом $h = 3$, а также прореженное окно при $s \in \{1, 2, 3\}$; с небольшим преимуществом лидировало плотное окно, которое было принято в качестве основного варианта в дальнейших экспериментах и обучении.

4. Обучение и тестирование моделей машинного обучения. Для решения задачи идентификации личности по кинематическим признакам походки рассматривались методы градиентного бустинга (CatBoost, XGBoost, LightGBM) и нейросетевые модели (MLP, LSTM) [16–18]; обучение и оценка выполнялись на объединённом корпусе с разделением на обучающие/тестовые выборки и последующей k-fold кросс-валидацией для робастной оценки качества.

Нормализация признаков и процедура обучения. Оконные векторы признаков всех обучающих записей объединялись в единую матрицу $X \in \mathbb{R}^{N \times D}$, где N – общее число окон, $D = LF$ – размерность векторизованного окна признаков (в рабочей конфигурации $L = 30$, $F = 116$, $D = 3480$). Масштабирование выполнялось стандартной нормализацией:

$$\mathbf{X}' = \frac{\mathbf{X} - \boldsymbol{\mu}}{\boldsymbol{\sigma}}, \quad \boldsymbol{\mu}, \boldsymbol{\sigma} \in \mathbb{R}^D.$$

Разделение на обучающую и валидационную подвыборки выполнялось в соответствии с процедурой 5-кратной кросс-валидации, описанной в разделе 2. На каждом фолде модель обучалась на обучающей подвыборке, а валидационная подвыборка использовалась для ранней остановки и сохранения лучшего состояния модели (по минимуму валидационной функции потерь).

По завершении всех пяти фолдов итоговой выбиралась модель того фолда, который обеспечил наибольшее значение Accuracy на своей валидационной подвыборке. Именно эта модель затем применялась к тестовому набору (третьи видеозаписи всех испытуемых).

В таблице 3 колонки «Обучение» и «Валидация» соответствуют обучающей и валидационной подвыборкам лучшего фолда, а колонка

«Тестирование» – результатам применения этой модели к отложенному тестовому набору.

Архитектура многослойного персептрона (MLP).

В качестве основной модели классификации использовался глубокий многослойный персептрон (MLP), реализованный на PyTorch. Модель принимает на вход нормализованный вектор признаков $\mathbf{X}' \in \mathbb{R}^{N \times D}$, где $D = 3480$ (116 признаков $\times$ 30 временных шагов после препроцессинга), и выполняет классификацию на C классов (число персон в задаче идентификации).

Архитектура состоит из четырёх линейных слоёв со следующими скрытыми размерностями:

– вход $\rightarrow$ 512 нейронов
– 512 $\rightarrow$ 256 нейронов
– 256 $\rightarrow$ 128 нейронов
– 128 $\rightarrow C$ (выходной слой)

Каждый скрытый блок имеет структуру:

$$\textbf{Linear} \rightarrow \textbf{BatchNorm1d} \rightarrow \textbf{GELU} \rightarrow \textbf{Dropout}$$

Вероятности dropout увеличиваются по глубине слоя:

$$d_0 = 0.10, d_1 = 0.15, d_2 = 0.30$$

Математически прямой проход описывается следующим образом:

$$\begin{aligned}
\mathbf{H}_0 &= \mathbf{X}', \\
\mathbf{H}_1 &= \mathrm{Dropout}_{0.10}\big(\mathrm{GELU}(\mathrm{BN}(\mathbf{H}_0\mathbf{W}_0 + \mathbf{b}_0))\big), \quad \mathbf{W}_0 \in \mathbb{R}^{D\times 512}, \\
\mathbf{H}_2 &= \mathrm{Dropout}_{0.15}\big(\mathrm{GELU}(\mathrm{BN}(\mathbf{H}_1\mathbf{W}_1 + \mathbf{b}_1))\big), \quad \mathbf{W}_1 \in \mathbb{R}^{512\times 256}, \\
\mathbf{H}_3 &= \mathrm{Dropout}_{0.30}\big(\mathrm{GELU}(\mathrm{BN}(\mathbf{H}_2\mathbf{W}_2 + \mathbf{b}_2))\big), \quad \mathbf{W}_2 \in \mathbb{R}^{256\times 128}, \\
\mathbf{O} &= \mathbf{H}_3\mathbf{W}_{\text{out}} + \mathbf{b}_{\text{out}}, \quad \mathbf{W}_{\text{out}} \in \mathbb{R}^{128\times C}.
\end{aligned}$$

Инициализация весов линейных слоёв выполнена методом **Kaiming Normal** с режимом `fan_in` и нелинейностью `relu` (рекомендуется для GELU). Смещения инициализированы нулями. Модель обучалась в режиме 5-fold стратифицированной кросс-валидации на всём доступном

обучающем наборе данных. Лучшая модель по валидационной потере сохранялась для последующей оценки.

Функция потерь. Основная функция потерь — многоклассовая кросс-энтропия:

$$\mathcal{L}_{\mathrm{CE}} = -\frac{1}{N}\sum_{i=1}^{N} \log p_{i,y_i}, \quad p_{i,c} = \frac{\exp(O_{i,c})}{\sum_{j=1}^{C}\exp(O_{i,j})}.$$

Также использовалась фокальная потеря (Focal Loss) для снижения вклада легко классифицируемых примеров и учёта дисбаланса классов [19]:

$$\mathcal{L}_{\mathrm{focal}} = -\frac{1}{N}\sum_{i=1}^{N} \alpha_{y_i}\,(1-p_{i,y_i})^{\gamma}\,\log p_{i,y_i},$$

где $\gamma = 2$, а α_c задавались пропорционально эмпирическим долям классов в текущем валидационном фолде. Однако результаты использования этой функции потерь не дали значительного прироста по точности предсказания.

Обучение, планировщик и ранняя остановка. Обучение выполнялось с использованием одного большого батча данных (размер батча $10\,000$ объектов), включающего наблюдения по всем испытуемым, при максимальном числе эпох 400. Фактическое число эпох определялось сочетанием правила адаптивного уменьшения шага обучения ReduceLROnPlateau и критерия ранней остановки. Шаг обучения автоматически уменьшался по правилу ReduceLROnPlateau (factor $= 0.5$, patience $= 3$, $\eta_{\min} = 10^{-5}$); при этом уменьшение происходило при отсутствии улучшения выбранной метрики на валидационной выборке в течение заданного числа эпох [20]. Критерий ранней остановки задавался как отсутствие улучшения значения функции потерь на валидационной выборке в течение 10 последовательных эпох.

Для обеспечения воспроизводимости результатов все источники случайности (инициализация весов, разбиение данных на фолды, порядок формирования пакетов данных) фиксировались при seed $= 42$.

4.1. Процедура вывода и агрегирование решений. При выполнении процедуры вывода к предварительно вычисленным признакам тестовой выборки применялся сохранённый преобразователь StandardScaler, после чего использовалась обученная модель MLP

в рабочей конфигурации. Пусть $\mathbf{x}_i \in \mathbb{R}^D$ – вектор признаков i-го экземпляра данных (окна или кадра) после нормализации, где $i = 1, \ldots, N$, N – общее число экземпляров данных в рассматриваемой выборке, D – размерность признакового вектора. Выходные значения сети (логиты) обозначим $\mathbf{o}_i \in \mathbb{R}^C$, где C – число классов (число испытуемых). Вероятности классов определяются как $\mathbf{p}_i = \mathrm{softmax}(\mathbf{o}_i)$, то есть $p_{i,c}$ – оценка вероятности класса c для экземпляра i, $c = 1, \ldots, C$.

Фильтрация по порогу разности вероятностей. Для каждого экземпляра данных определяется класс с максимальной вероятностью:

$$\hat{y}_i = \arg \max_{c \in \{1,\ldots,C\}} p_{i,c}.$$

и класс, занявший второе место по величине вероятности,

$$\tilde{y}_i = \arg \max_{c \in \{1,\ldots,C\} \setminus \{\hat{y}_i\}} p_{i,c}.$$

Вводится величина разности вероятностей (margin):

$$M_i = p_{i,\hat{y}_i} - p_{i,\tilde{y}_i}.$$

К дальнейшему анализу допускаются только те экземпляры, для которых выполняется условие:

$$\mathcal{I}(\tau) = \big\{ i \in \{1, \ldots, N\} : M_i \geqslant \tau \big\},$$

где $\tau \in [0, 1]$ – фиксированный порог уверенности; в базовой оценке использовалось $\tau = 0.9$.

Агрегирование по видеозаписи. Пусть $\mathcal{F}$ – множество тестовых видеозаписей (файлов), а $F \in \mathcal{F}$ – одна видеозапись. Обозначим через $I(F) \subseteq \{1, \ldots, N\}$ множество индексов экземпляров данных (окон/кадров), сформированных из видеозаписи F. Для каждой видеозаписи F накапливается взвешенная сумма по классам, рассчитанная только по отфильтрованным экземплярам:

$$S_c(F;\tau) \;=\; \sum_{\substack{i\in I(F)\cap\mathcal{I}(\tau)\\ \hat{y}_i=c}} M_i, \qquad c=1,\ldots,C,$$

а итоговое решение по видеозаписи определяется как:

$$\widehat{Y}(F) \;=\; \arg\max_{c\in\{1,\ldots,C\}} S_c(F;\tau).$$

Дополнительно вычисляется усреднённое распределение вероятностей:

$$\bar{\mathbf{p}}(F;\tau) \;=\; \frac{1}{|I(F)\cap\mathcal{I}(\tau)|} \sum_{i\in I(F)\cap\mathcal{I}(\tau)} \mathbf{p}_i,$$

по которому формируется список трёх наиболее вероятных классов для повышения интерпретируемости результатов.

Итоговое решение и показатели качества классификации. На уровне экземпляров данных (окон/кадров) формируется отчёт о классификации по множеству $\mathcal{I}(\tau)$ с метриками $\mathrm{Accuracy}$, $\mathrm{F1}_{\mathrm{weighted}}$, $\mathrm{Precision}_{\mathrm{weighted}}$, $\mathrm{Recall}_{\mathrm{weighted}}$. На уровне видеозаписей оценивается доля корректно распознанных записей:

$$\mathrm{Acc}_{\mathrm{files}} \;=\; \frac{\left|\left\{F\in\mathcal{F}:\ \widehat{Y}(F)=Y(F)\right\}\right|}{|\mathcal{F}|},$$

где $Y(F)\in\{1,\ldots,C\}$ – истинная метка личности для видеозаписи F.

Другие модели, используемые в эксперименте. Для сравнительной оценки в эксперименте были рассмотрены четыре модели машинного обучения, решающие задачу многоклассовой классификации походки по последовательностям предварительно вычисленных признаков.

Модель CatBoost представляет собой градиентный бустинг над решающими деревьями с поддержкой категориальных признаков. Лучшая конфигурация, полученная при случайном поиске гиперпараметров, включала 600 итераций, глубину деревьев 10, скорость обучения 0.01,

тип роста деревьев Depthwise, бутстрэп по схеме Bernoulli с subsample = 0.5, минимальное число объектов в листе min_data_in_leaf = 2, L_2-регуляризацию l2_leaf_reg = 1, random_strength = 0.1 и автоматическое взвешивание классов auto_class_weights = Balanced. Обучение выполнялось с использованием графического процессора.

Модель LightGBM также относится к семейству методов градиентного бустинга. Итоговая конфигурация включала 700 деревьев, num_leaves = 31, максимальную глубину 10, скорость обучения 0.01, min_data_in_leaf = 20, feature_fraction = 1.0, регуляризацию lambda_l1 = 0.1 и lambda_l2 = 0.1, subsample_for_bin = 200000 и тип бустинга gbdt. Как и в случае CatBoost, обучение выполнялось с использованием графического процессора.

Модель XGBoost настраивалась следующим образом: 600 деревьев, максимальная глубина 10, скорость обучения 0.01, subsample = 0.5, min_child_weight = 2, reg_lambda = 1, метод построения деревьев gpu_hist, целевая функция multi:softmax и функция качества mlogloss. Обучение также выполнялось с использованием графического процессора.

Модель LSTM представляет собой рекуррентную нейронную сеть с механизмом долгой краткосрочной памяти. Архитектура включала четыре двунаправленных слоя LSTM с размерностью скрытого состояния 186 и межслойным dropout 0.6. К выходу последнего временного шага применялась полносвязная классифицирующая часть: линейный слой, преобразующий конкатенацию прямого и обратного состояний ($186 \times 2 = 372$) в 128 нейронов с активацией ReLU и dropout = 0.3, а затем финальный линейный слой в пространство размерности N_classes.

Модели градиентного бустинга (CatBoost, LightGBM, XGBoost) обучались на покадровых признаках без явного учёта временной зависимости: каждый кадр рассматривался как независимый объект. В отличие от них, модель LSTM обрабатывала временные последовательности, что позволяло учитывать динамику изменения признаков и, как следствие, описывать временные закономерности походки.

Практические замечания. Порог T задаёт компромисс между точностью и полнотой на уровне кадров и влияет на устойчивость агрегирования по отдельным видеозаписям. В демонстрационных и базовых экспериментах использовалось фиксированное значение $T = 0.9$, тогда как в прикладных сценариях оно может подбираться по валидационной выборке.

Результаты экспериментов. Качество классификации оценивалось на тестовой выборке по двум основным показателям:

взвешенной F_1-мере ($\mathrm{F1}_{\mathrm{weighted}}$) и числу корректных идентификаций личности на уровне видеозаписей (TP – число истинно положительных решений). Для обучающей и валидационной подвыборок приводятся средние значения и стандартные отклонения по пяти фрагментам кросс-валидации; тестовые метрики являются точечными оценками, полученными однократным применением лучшей модели к отложенному тестовому набору. Сравнительные результаты для рассмотренных архитектур представлены в таблице 3.

Таблица 3. Метрики на обучающих, валидационных и тестовых данных (среднее $\pm$ std по 5 фолдам; тест – точечная оценка)

Модель	**Обучение**		**Валидация**		**Тестирование**	
	F1	**TP**	**F1**	**TP**	**F1**	**TP**
MLP	$0{,}99_{\pm 0{,}01}$	19/19	$0{,}98_{\pm 0{,}01}$	19/19	0,69	14/19
LSTM	$0{,}99_{\pm 0{,}01}$	19/19	$0{,}96_{\pm 0{,}02}$	19/19	0,59	11/19
Catboost	$0{,}97_{\pm 0{,}01}$	19/19	$0{,}89_{\pm 0{,}02}$	19/19	0,44	10/19
Xgboost	$0{,}96_{\pm 0{,}01}$	19/19	$0{,}88_{\pm 0{,}02}$	19/19	0,41	8/19
LightGBM	$0{,}95_{\pm 0{,}01}$	19/19	$0{,}88_{\pm 0{,}02}$	19/19	0,44	10/19

Оценка значимости различий между моделями. Стандартные отклонения $\mathrm{F1}_{\mathrm{weighted}}$ по фрагментам кросс-валидации не превышают $0{,}02$ для всех моделей, что свидетельствует о стабильности оценок. Различие между MLP и моделями градиентного бустинга на валидационной подвыборке составляет $0{,}09$–$0{,}10$ (при $\mathrm{std} \leqslant 0{,}02$), что превышает четыре стандартных отклонения и указывает на содержательную значимость преимущества MLP. На тестовой выборке разрыв ещё более выражен: $\mathrm{F1} = 0{,}69$ для MLP против $0{,}41$–$0{,}44$ для бустинговых моделей. Различия между тремя моделями градиентного бустинга (CatBoost, XGBoost, LightGBM) на валидации и тесте лежат в пределах наблюдаемой вариации ($\leqslant 0{,}03$) и не являются статистически значимыми.

Дополнительно были построены матрицы ошибок (confusion matrix) и ROC-кривые в подходе One-vs-Rest для всех моделей на тестовом наборе. Матрицы ошибок и ROC-кривые для лучшей модели (MLP) приведены на рисунках 2 и 3 соответственно.

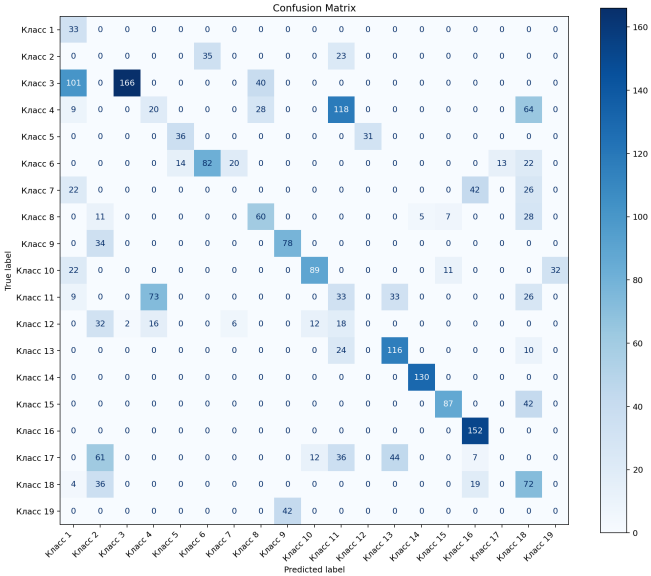


Рис. 2. Матрица ошибок для модели MLP на тестовом наборе

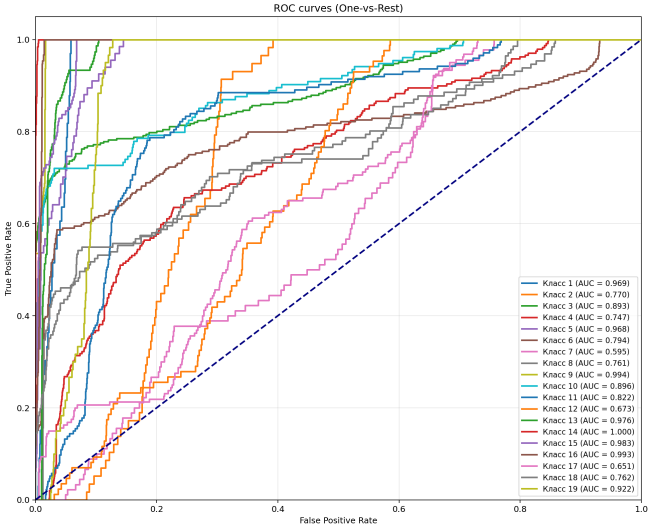


Рис. 3. ROC-кривые One-vs-Rest для модели MLP на тестовом наборе

Причины снижения качества на тестовой выборке. Снижение качества на тестовой выборке по сравнению с обучающей и валидационной (таблица 3) обусловлено тем, что валидация является внутрисессионной: она формируется из последних 30% кадров тех же двух видеозаписей, которые используются для обучения, поэтому распределения обучающих и валидационных данных близки. Тестовая выборка соответствует отдельной (третьей) сессии, выполненной через 2–5 недель, что приводит к межсессионному смещению распределения (изменение внешнего вида, небольшие изменения условий съёмки и установки камеры, вариативность темпа и манеры ходьбы); такие различия являются одной из основных причин деградации качества распознавания походки [6].

Дополнительным фактором выступает ограниченный объём данных (19 классов, по две обучающие записи на класс), из-за чего модель может частично подстраиваться под статистику конкретных сессий (характер шумов 3D-реконструкции, типичные пропуски ключевых точек, небольшие артефакты детекции), сохраняя высокие значения метрик на внутрисессионной валидации, но теряя качество при переносе на межсессионный тест.

Причины низкой производительности LSTM в сравнении с MLP. Обе модели обучались на одних и тех же оконных представлениях признаков и более низкое качество LSTM, по сравнению с MLP, может быть связано с тем, что рекуррентная архитектура оказалась избыточной и хуже «настроенной» под имеющийся объём данных. Признаки уже содержат существенную часть временной информации, поэтому дополнительное моделирование последовательности внутри окна даёт ограниченный выигрыш, а иногда и ухудшает обобщающую способность. При наличии локально шумных участков (ошибки трёхмерной реконструкции, несовершенство алгоритма заполнения пропусков) рекуррентная модель может сильнее накапливать влияние этих ошибок в скрытом состоянии, тогда как MLP, работающий с вектором окна, оказывается устойчивее.

Сравнение методов заполнения пропусков. Для основной модели MLP выполнено сравнение точности классификации в зависимости от выбора способа заполнения пропущенных значений. Результаты по показателю TP (число корректно идентифицированных личностей из 19 на обучающей выборке) приведены в таблице 4.

Таблица 4. Количество верно идентифицированных личностей (TP) для разных методов заполнения пропусков

Метод заполнения	TP (Обучающая выборка)
Forward fill	19/19
Кубическая сплайн-интерполяция	18/19
KNN-заполнение	18/19
Скользящее среднее по времени	19/19
Итеративное модельное заполнение	17/19

5. Сравнение производительности на различном аппаратном обеспечении. Оценка производительности системы проводилась при различных вычислительных ресурсах. Тестирование выполнялось на следующих платформах: NVIDIA A100 (80 ГБ), NVIDIA RTX 4070 Ti SUPER, NVIDIA Jetson Orin AGX 64 и NVIDIA Jetson TX2. Такой выбор позволяет оценить работу системы как на высокопроизводительных серверных решениях, так и на энергоэффективных встраиваемых вычислительных устройствах.

Измерения включали временные затраты на ключевые этапы обработки: вычисление признаков, параметризацию данных и выполнение процедуры вывода модели. На платформе NVIDIA Jetson TX2 этап подготовки данных не выполнялся, поскольку из-за ограниченного объёма внутренней памяти отсутствовала возможность загрузить стереовидеозаписи обучающей выборки. Все испытания проводились на одном и том же наборе данных для двух видеозаписей при фиксированных гиперпараметрах базовой модели, что обеспечивало сопоставимость результатов. Полученные значения времени обработки приведены в таблице 5 для каждого из тестируемых устройств. Анализ результатов позволяет выявить зависимость производительности системы от аппаратных характеристик используемой платформы.

Таблица 5. Скорость системы на устройствах разного типа

Устройства	Временные затраты		
	Подготовка данных, сек	Обучение модели на 19 классах, сек	Тестирование модели на 19 классах, сек
RTX 4070 TI SUPER	29	14	2
A100 80GB	43	31	11
Nvidia Jetson Orin AGX 64	62	35	13
Nvidia Jetson TX2	–	76	34

Программная реализация и окружение. Программная реализация выполнена на Ubuntu 22.04 с использованием Python 3.10. Для обеспечения воспроизводимости ниже приведены версии основных библиотек, использованных в экспериментальной части (обработка видеоданных, оценка позы, построение признаков и обучение моделей).

– **Базовые научные библиотеки:** numpy==1.26.4, scipy==1.14.0, pandas==2.2.2, numba==0.61.2.

– **Компьютерное зрение и работа с данными:** opencv-python==4.10.0.84, scikit-image==0.25.2, imageio==2.37.0, pyrealsense2==2.55.1.6486, ultralytics==8.3.54, ultralytics-thop==2.0.14.

– **Машинное обучение:** scikit-learn==1.5.1, scikit-learn-intelex==2024.6.0, torch==1.13.1, torchvision==0.14.1, torchaudio==0.13.1, catboost==1.2.7, lightgbm==4.5.0, xgboost==3.0.0.

– **Визуализация и интерпретация:** matplotlib==3.10.1, seaborn==0.13.2, plotly==5.23.0, shap==0.47.2, lime==0.2.0.1, tqdm==4.67.1.

– **GPU-библиотеки:** nvidia-cublas-cu11==11.10.3.66, nvidia-cuda-nvrtc-cu11==11.7.99, nvidia-cuda-runtime-cu11==11.7.99, nvidia-cudnn-cu11==8.5.0.96, nvidia-nccl-cu12==2.26.5.

6. Сравнение с опубликованными подходами.

6.1. Метрики качества. В работах по идентификации личности по походке широко используется *точность ранга-k* (Rank-k) – доля испытаний, в которых истинная личность попала в список из k наиболее близких кандидатов. Частный случай Rank-1 совпадает с обычной точностью классификации (правильный ответ на первом месте), Rank-5 – правильный ответ в первой пятёрке.

Метод авторов (трёхмерная кинематика, 1 стереокамера).

Задача: Идентификация личности по походке (биометрия).
Съёмка и ракурсы: 1 стереокамера RealSense D457 (стерео + карта глубины), 1 ракурс на испытуемого, $\sim$225 с на человека, 3 сессии, межсессионная проверка.
Вход: Трёхмерный скелет и инвариантные геометрико-кинематические признаки; классификатор – многослойный перцептрон.
Инвариантность: Явно вводится инвариантность признаков к геометрическим преобразованиям (перенос/поворот системы координат) за счёт конструкции признаков.
Качество: Rank-1 $= 14/19$ (73.7%) в межсессионном режиме.

Соколова А., Конушин А. [6]: двумерный оптический поток и сверточная сеть.

Задача: Идентификация личности по походке; метрики: Rank-1/Rank-5.
Съёмка и ракурсы: CASIA-B (Chinese Academy of Sciences, Institute of Automation, Dataset B): 124 человека, 11 ракурсов (0–180° с шагом 18°), на каждого человека по 10 видеопоследовательностей на *каждый* ракурс (6 – обычная ходьба, 2 – с сумкой, 2 – в верхней одежде), длительность 3–5 с; в «обобщаемом» протоколе фиксируется ракурс 90° (обучение на 60 людях, тест на 64).
Вход: Двумерный оптический поток $\rightarrow$ сверточная сеть.
Инвариантность: Явная инвариантность к изменению ракурса в приведённом протоколе не обеспечивается (ракурс фиксируется).
Качество: TUM-GAID (TUM Gait from Audio, Image and Depth): Rank-1 $= 97.52\%$, Rank-5 $= 99.89\%$; CASIA-B (90°): Rank-1 $= 74.93\%$.

Bari A.S.M.H., Gavrilova M.L. [8]: KinectGaitNet, трёхмерные суставы (RGB-D) и сверточная сеть.

Задача: Идентификация личности по походке; метрика: Accuracy (эквивалент Rank-1 для классификации).
Съёмка и ракурсы: 1 сенсор Microsoft Kinect (RGB-D); трёхмерные координаты суставов по циклу шага.
Данные: UPCV (University of Patras Computer Vision) – 30 участников, по 5 последовательностей; KGB (Kinect Gait Biometry) – 164 участника, по 5 последовательностей.
Инвариантность: Кинематический подход (работа по трёхмерным суставам), однако явная геометрическая инвариантность (к переносу/повороту системы координат) отдельно не выделена.
Качество: 96.91% (UPCV) и 99.33% (KGB).

Mao M., Song Y. [16]: графовая сверточная сеть по трёхмерному скелету.

Задача: Идентификация личности по походке; метрика: Rank-1.
Съёмка и ракурсы: Многоракурсная постановка (в т.ч. CASIA-B с 11 ракурсами).
Вход: Трёхмерные координаты суставов и костей $\rightarrow$ графовая сверточная сеть; объединение признаков; оптимизация: Softmax + центрирующая функция потерь (center loss).
Инвариантность: Строго кинематический подход (скелет); устойчивость к ракурсу достигается нормировкой скелета и обучаемым представлением, без явного введения инвариантных скалярных признаков.

Качество: Rank-1 варьируется в диапазоне от 66,3% до 87,7% на CASIA-B в зависимости от условий съёмки, что свидетельствует о чувствительности к ракурсу и внешним факторам.

Luo J., Xu B., Tjahjadi T., Yi J. [5]: параметрическая трёхмерная модель походки (3DGait), устойчивая к сумкам/одежде.
Задача: Идентификация личности по походке; основная метрика: Rank-1.
Съёмка и ракурсы: Оценка на CASIA-B (11 ракурсов) и др. наборах; в статье отдельно рассматриваются многоракурсная идентификация и вариации условий (сумка/одежда).
Вход: Дескрипторы 3DGait: форма (3D-Shape), поза/движение (3D-Pose) и внешние факторы (3D-eFactors); устойчивость к внешним условиям задаётся параметрической моделью и обучением отображения 2D→3D-дескрипторов.
Инвариантность: Авторы формулируют использование интерпретируемых трёхмерных дескрипторов как инвариантных признаков для устойчивости к переносимым предметам и вариативности одежды.
Качество (средние): CASIA-B, многоракурсная идентификация (среднее по тестовым ракурсам): 89,8%; CASIA-B, с вариациями условий: 86,5%.

Итоговые выводы. В отличие от большинства рассмотренных работ, наш метод изначально разрабатывался под применение стереокамеры, позволяющей получать трёхмерные координаты и строить признаки, устойчивые к геометрическим преобразованиям. Мы используем вычислительно экономичную архитектуру (классификатор на многослойном перцептроне), тогда как в ряде трёхмерных подходов применяются более тяжёлые нейросетевые методы. С точки зрения условий съёмки предлагаемый нами сценарий сложнее по ограничению наблюдения: одна камера и один ракурс на испытуемого, тогда как в типовых публичных наборах для сравнения часто доступны многоракурсные наблюдения и/или многокамерная регистрация. При этом наши записи на человека относительно длинные, но число сессий/примеров ограничено, что делает задачу более чувствительной к межсессионным изменениям и статистической недостаточности данных. Отдельно важно, что в исследовании оценка проводится именно в межсессионном режиме, и проблема деградации качества при смене сессии/условий наблюдается и в опубликованных работах (например, в экспериментах с временным разрывом на TUM GAID точность Rank-1 заметно ниже).

7. Заключение. В работе разработана и экспериментально исследована система идентификации личности по походке на основе

безмаркерной реконструкции трёхмерных координат ключевых точек и последующего построения инвариантного пространства признаков. Предложенный подход ориентирован на использование минимального набора оборудования (одна стереокамера) и на извлечение именно кинематической информации, без опоры на внешние признаки (лицо, одежда, фактура и т. п.), которые могут неявно повышать качество распознавания в лабораторных условиях, но снижают воспроизводимость и переносимость метода.

Ключевые научные выводы работы состоят в следующем:

– Показано, что прямое использование сглаженных трёхмерных координат ключевых точек без явной параметризации (окна из координат $\mathbf{W}_t$ на входе классификатора) обеспечивает точность не выше 25%, что указывает на недостаточность такого представления для надёжной идентификации и мотивирует переход к явному построению признаков.

– Построение геометрико-кинематических признаков, инвариантных к ортогональным преобразованиям и переносам, позволяет существенно повысить качество идентификации по сравнению с базисом из «сырых» координат.

– Использование скользящих окон признаков и последующего агрегирования решений по видеозаписи повышает устойчивость итогового результата по сравнению с покадровыми решениями, поскольку сглаживает влияние случайных ошибок детектора позы и локальных артефактов траекторий.

– Проведено сравнение нескольких моделей машинного обучения (градиентный бустинг и нейросетевые модели) и выполнена оценка вычислительных затрат на разных аппаратных платформах, что позволяет выбирать конфигурацию системы в зависимости от доступных вычислительных ресурсов и требований к времени обработки.

Дальнейшие исследования предполагают расширение набора данных за счёт дополнительных видеозаписей для каждого испытуемого, а также повышение точности на тестовой выборке посредством внедрения архитектур на основе трансформеров [21] и дополнительного обучения модели компьютерного зрения. В качестве направления развития рассматривается применение модели разметки с расширенным набором ключевых точек, описывающих положение и позу человека в пространстве (например, 35 точек вместо используемых 17) [22]. Кроме того, планируется переход от чисто классификационной постановки с кросс-энтропийной функцией потерь к обучению векторных представлений походки в скрытом пространстве, что позволит масштабировать число идентифицируемых персон без изменения архитектуры нейросетевой

модели и эффективнее использовать существенно более крупные выборки; при этом возможно расширение числа классов на основе таких векторных представлений без полного переобучения модели [23].

Литература

1. Ramesh S.H., Lemaire E.D., Tu A., Cheung K., Baddour N. Automated Implementation of the Edinburgh Visual Gait Score (EVGS) Using OpenPose and Handheld Smartphone Video // Sensors. 2023. vol. 23. pp. 1–36.
2. Moro M., Marchesi G., Hesse F., Odone F., Casadio M. Markerless vs. Marker-Based Gait Analysis: A Proof of Concept Study // Sensors. 2022. vol. 22. pp. 1–15. DOI: 10.3390/s22052011.
3. Han X., Guffanti D., Brunete M.A. A Comprehensive Review of Vision-Based Sensor Systems for Human Gait Analysis // Sensors. 2025. vol. 25. pp. 1–33. DOI: 10.3390/s25020498.
4. Sapkota R., Karkee M. YOLO11 and Vision Transformers Based 3D Pose Estimation of Immature Green Fruits in Commercial Apple Orchards for Robotic Thinning // arXiv preprint arXiv:2410.19846. 2024. pp. 1–18.
5. Luo J., Xu B., Tjahjadi T., Yi J. A Novel 3D Gait Model for Subject Identification Robust against Carrying and Dressing Variations // Computers, Materials Continua. 2024. vol. 82. no. 1. pp. 194–215. DOI: 10.32604/cmc.2024.050018.
6. Sokolova A., Konushin A. Gait Recognition Based on Convolutional Neural Networks // The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences. 2017. vol. XLII-2/W4. pp. 207–212. DOI: 10.5194/isprs-archives-XLII-2-W4-207-2017.
7. Natraj S., Messmer T., Fujii Y., Suzuki K., Riener R., Eriks-Hoogland I., Paez-Granados D. 3D Pose Estimation for Scalable Remote Gait Kinematics Assessment // npj Digital Medicine. 2026. vol. 9. no. 37. DOI: 10.1038/s41746-025-02211-y.
8. Bari A.S.M.H., Gavrilova M.L. KinectGaitNet: Kinect-Based Gait Recognition Using Deep Convolutional Neural Network // Sensors. 2022. vol. 22. pp. 1–15. DOI: 10.3390/s22072631.
9. Cai S., Xu H., Cai W., et al. A human pose estimation network based on YOLOv8 with efficient multi-scale receptive field and expanded feature pyramid network // Scientific Reports. 2025. vol. 15. DOI: 10.1038/s41598-025-00259-0.
10. Kwon J., Lee Y., Lee J. Comparative Study of Markerless Vision-Based Gait Analyses for Person Re-Identification // Sensors. 2021. vol. 21. pp. 1–25. DOI: 10.3390/s21248208.
11. Ultralytics YOLO11 // GitHub. URL: www.github.com/ultralytics/ultralytics (дата обращения: 10.11.2024).
12. Intel RealSense SDK 2.0 (librealsense) // GitHub URL: www.github.com/IntelRealSense/librealsense.
13. Yoon H., Jo E., Ryu S., Yoo J.-I., Kim M., Kim J.H. Noise-robust markerless video gait anomaly detection via two-stage acquisition and LSTM autoencoders // Scientific Reports. 2025. vol. 15. DOI: 10.1038/s41598-025-26169-9.
14. Huo Z., Ji T., et al. DynImp: Dynamic Imputation for Wearable Sensing Data Through Sensory and Temporal Relatedness // arXiv preprint arXiv:2209.15415. 2022. pp. 1–25.
15. Jun K., Lee K., Lee S., Lee H., Kim M.-K. Hybrid Deep Neural Network Framework Combining Skeleton and Gait Features for Pathological Gait Recognition // Bioengineering. 2023. vol. 10. 1133 p. DOI: 10.3390/bioengineering10101133.

16. Mao M., Song Y. Gait Recognition Based on 3D Skeleton Data and Graph Convolutional Network // IEEE International Joint Conference on Biometrics (IJCB). 2020. DOI: 10.1109/IJCB48548.2020.9304916.
17. Rao H., Miao C. A Survey on 3D Skeleton-Based Person Re-Identification: Approaches, Designs, Challenges, and Future Directions // arXiv preprint arXiv:2401.15296. 2024.
18. Khaliluzzaman M., Uddin A., et al. Person Recognition Based on Deep Gait: A Survey // Sensors. 2023. vol. 23. pp. 1–36. DOI: 10.3390/s23104875.
19. Lin T.-Y., Goyal P., Girshick R., He K., Dollar P. Focal Loss for Dense Object Detection // arXiv preprint arXiv:1708.02002. 2017. pp. 1–10.
20. Yun S., Jeong M., Kim R., Kang J., Kim H.J. Graph Transformer Networks // arXiv preprint arXiv:1911.06455. 2019. pp. 1–11.
21. Shi L.-F., Liu Z.-Y., Zhou K.-J., Shi Y., Jing X. Novel Deep Learning Network for Gait Recognition Using Multimodal Inertial Sensors // Sensors. 2023. vol. 23. pp. 1–17. DOI: 10.3390/s23020849.
22. Maji D., Nagori S., Mathew M., Poddar D. YOLO-Pose: Enhancing YOLO for Multi Person Pose Estimation Using Object Keypoint Similarity Loss // arXiv preprint arXiv:2204.06806. 2022. pp. 1–10.
23. Netay I.V. Series of quasi-uniform scatterings with fast search, root systems and neural network classifications // arXiv preprint arXiv:2512.04865. 2025. pp. 1–13.

Нетай Игорь Витальевич — канд. физ.-мат. наук; старший научный сотрудник ИППИ РАН. Область научных интересов: алгебра, алгебраическая геометрия, теория представлений, теория инвариантов, численные методы, машинное обучение и нейросети, цифровая обработка сигналов, оптимизация вычислений. Число научных публикаций — 15. i.netay@kryptonite.ru, www.iitp.ru/ru/about; ИППИ РАН, Большой Каретный переулок, д. 19, стр. 1, г. Москва, 127051, РФ; р.т. +7(903)778-6644.

Панкратов Евгений Павлович — младший научный сотрудник ИППИ РАН. Область научных интересов: машинное обучение, обработка естественного языка, большие языковые модели, нейросетевое моделирование. Число научных публикаций — 2. pankratov.ep@phystech.edu, www.iitp.ru/ru/about; ИППИ РАН, Большой Каретный переулок, д. 19, стр. 1, г. Москва, 127051, РФ; р.т. +7(977)424-9237.

Корнилов Константин Игоревич — стажер-исследователь ИППИ РАН. Область научных интересов: генеративные модели, прогнозирование временных рядов, прикладные исследования больших языковых моделей. Число научных публикаций — 1. kornilov.ki@phystech.edu, www.iitp.ru/ru/about; ИППИ РАН, Большой Каретный переулок, д. 19, стр. 1, г. Москва, 127051, РФ; р.т. +7(910)531-8381.

Голубинский Андрей Николаевич — д-р техн. наук, доцент; начальник отдела, Российский научный фонд. Область научных интересов: машинное обучение, нейросетевое моделирование, автоматизированные системы управления с элементами искусственного интеллекта, обработка речевых сигналов. Число научных публикаций — 250. annikgol@mail.ru, www.rscf.ru; ул. Солянка, д. 14, стр. 3, г. Москва, 109240, РФ; р.т. +7(910)346-6537.

DOI 10.15622/ia.25.3.5

I. V. NETAY , E. P. PANKRATOV , K. I. KORNILOV , A. N. GOLUBINSKIY

IDENTIFICATION OF A PERSON BY GAIT KINEMATICS IN THREE-DIMENSIONAL SPACE BASED ON INVARIANT FEATURES

Netay I. V., Pankratov E. P., Kornilov K. I., Golubinskiy A. N. **Identification of a Person by Gait Kinematics in Three-Dimensional Space Based on Invariant Features.**

Abstract. A system for data acquisition, preprocessing, and analysis for gait-based personal identification using machine learning and image analysis methods is proposed. The system includes a preprocessing mechanism for normalizing biometric motion data and a post-processing stage for extracting features invariant to orthogonal transformations of space and changes in camera pose, including perspective effects. A computationally efficient gait recognition method using a single stereo camera is implemented. The feature space is analyzed to determine the most informative characteristics, and the employed feature set is comparatively assessed. Several machine-learning architectures, including deep learning models, are evaluated with respect to both accuracy and computational efficiency. System performance is experimentally studied on various computing devices, and processing time characteristics are measured. The results enable comparison of applicability and efficiency depending on available computational resources.

The input data are formed as time series of 3D coordinates of keypoints reconstructed from depth data using 2D pose detections (COCO, 17 keypoints), where missing detections are encoded by a zero vector. For gait identification, head keypoints are excluded as uninformative, while missing values are filled and the resulting trajectories are subsequently smoothed. An invariant feature space is built from the skeleton geometry (segment lengths, joint angles, and their derivatives); then these features are aggregated into a single vector using a sliding temporal window, i.e., each vector describes motion over an interval of several consecutive frames. At the identification stage, threshold-based filtering and aggregation of predictions at the frame level are applied. It is shown that both accuracy and processing speed depend on the chosen machine-learning model and the available computational resources.

Keywords: machine learning, computer vision, biometrics, neural networks, kinematics, 3D video processing, invariant features

References

1. Ramesh S.H., Lemaire E.D., Tu A., Cheung K., Baddour N. Automated Implementation of the Edinburgh Visual Gait Score (EVGS) Using OpenPose and Handheld Smartphone Video. Sensors. 2023. vol. 23. pp. 1–36.
2. Moro M., Marchesi G., Hesse F., Odone F., Casadio M. Markerless vs. Marker-Based Gait Analysis: A Proof of Concept Study. Sensors. 2022. vol. 22. pp. 1–15. DOI: 10.3390/s22052011.
3. Han X., Guffanti D., Brunete M.A. A Comprehensive Review of Vision-Based Sensor Systems for Human Gait Analysis. Sensors. 2025. vol. 25. pp. 1–33. DOI: 10.3390/s25020498.
4. Sapkota R., Karkee M. YOLO11 and Vision Transformers Based 3D Pose Estimation of Immature Green Fruits in Commercial Apple Orchards for Robotic Thinning. arXiv preprint arXiv:2410.19846. 2024. pp. 1–18.
5. Luo J., Xu B., Tjahjadi T., Yi J. A Novel 3D Gait Model for Subject Identification Robust against Carrying and Dressing Variations. Computers, Materials Continua. 2024. vol. 82. no. 1. pp. 194–215. DOI: 10.32604/cmc.2024.050018.

6. Sokolova A., Konushin A. Gait Recognition Based on Convolutional Neural Networks. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences. 2017. vol. XLII-2/W4. pp. 207–212. DOI: 10.5194/isprs-archives-XLII-2-W4-207-2017.
7. Natraj S., Messmer T., Fujii Y., Suzuki K., Riener R., Eriks-Hoogland I., Paez-Granados D. 3D Pose Estimation for Scalable Remote Gait Kinematics Assessment. npj Digital Medicine. 2026. vol. 9. no. 37. DOI: 10.1038/s41746-025-02211-y.
8. Bari A.S.M.H., Gavrilova M.L. KinectGaitNet: Kinect-Based Gait Recognition Using Deep Convolutional Neural Network. Sensors. 2022. vol. 22. pp. 1–15. DOI: 10.3390/s22072631.
9. Cai S., Xu H., Cai W., et al. A human pose estimation network based on YOLOv8 with efficient multi-scale receptive field and expanded feature pyramid network. Scientific Reports. 2025. vol. 15. DOI: 10.1038/s41598-025-00259-0.
10. Kwon J., Lee Y., Lee J. Comparative Study of Markerless Vision-Based Gait Analyses for Person Re-Identification. Sensors. 2021. vol. 21. pp. 1–25. DOI: 10.3390/s21248208.
11. Ultralytics YOLO11. GitHub. Available at: www.github.com/ultralytics/ultralytics (accessed 10.11.2024).
12. Intel RealSense SDK 2.0 (librealsense). GitHub. Available at: www.github.com/IntelRealSense/librealsense.
13. Yoon H., Jo E., Ryu S., Yoo J.-I., Kim M., Kim J.H. Noise-robust markerless video gait anomaly detection via two-stage acquisition and LSTM autoencoders. Scientific Reports. 2025. vol. 15. DOI: 10.1038/s41598-025-26169-9.
14. Huo Z., Ji T., et al. DynImp: Dynamic Imputation for Wearable Sensing Data Through Sensory and Temporal Relatedness. arXiv preprint arXiv:2209.15415. 2022. pp. 1–25.
15. Jun K., Lee K., Lee S., Lee H., Kim M.-K. Hybrid Deep Neural Network Framework Combining Skeleton and Gait Features for Pathological Gait Recognition. Bioengineering. 2023. vol. 10. 1133 p. DOI: 10.3390/bioengineering10101133.
16. Mao M., Song Y. Gait Recognition Based on 3D Skeleton Data and Graph Convolutional Network. IEEE International Joint Conference on Biometrics (IJCB). 2020. DOI: 10.1109/IJCB48548.2020.9304916.
17. Rao H., Miao C. A Survey on 3D Skeleton-Based Person Re-Identification: Approaches, Designs, Challenges, and Future Directions. arXiv preprint arXiv:2401.15296. 2024.
18. Khaliluzzaman M., Uddin A., et al. Person Recognition Based on Deep Gait: A Survey. Sensors. 2023. vol. 23. pp. 1–36. DOI: 10.3390/s23104875.
19. Lin T.-Y., Goyal P., Girshick R., He K., Dollar P. Focal Loss for Dense Object Detection. arXiv preprint arXiv:1708.02002. 2017. pp. 1–10.
20. Yun S., Jeong M., Kim R., Kang J., Kim H.J. Graph Transformer Networks. arXiv preprint arXiv:1911.06455. 2019. pp. 1–11.
21. Shi L.-F., Liu Z.-Y., Zhou K.-J., Shi Y., Jing X. Novel Deep Learning Network for Gait Recognition Using Multimodal Inertial Sensors. Sensors. 2023. vol. 23. pp. 1–17. DOI: 10.3390/s23020849.
22. Maji D., Nagori S., Mathew M., Poddar D. YOLO-Pose: Enhancing YOLO for Multi Person Pose Estimation Using Object Keypoint Similarity Loss. arXiv preprint arXiv:2204.06806. 2022. pp. 1–10.
23. Netay I.V. Series of quasi-uniform scatterings with fast search, root systems and neural network classifications. arXiv preprint arXiv:2512.04865. 2025. pp. 1–13.

Netay Igor Vitalievich — Ph.D., Associate professor, Senior researcher, Federal State Budgetary Institution of Science A.A. Kharkevich Institute for Information Transmission Problems of the

Russian Academy of Sciences (IPPI RAS). Research interests: algebra, algebraic geometry, representation theory, invariant theory, numerical methods, machine learning and neural networks, digital signal processing, computational optimization. The number of publications — 15.
i.netay@kryptonite.ru; 19, Bolshoy Karetny Lane, 127051, Moscow, Russia; office phone: +7(903)778-6644.

Pankratov Evgeny Pavlovich — Junior research assistant, Federal State Budgetary Institution of Science A.A. Kharkevich Institute for Information Transmission Problems of the Russian Academy of Sciences (IPPI RAS). Research interests: machine learning, natural language processing, large language models, neural network modeling. The number of publications — 2.
pankratov.ep@phystech.edu; 19, Bolshoy Karetny Lane, 127051, Moscow, Russia; office phone: +7(977)424-9237.

Kornilov Konstantin Igorevich — Intern researcher, Federal State Budgetary Institution of Science A.A. Kharkevich Institute for Information Transmission Problems of the Russian Academy of Sciences (IPPI RAS). Research interests: generative models, time series forecasting, applied research of large language models. The number of publications — 1.
kornilov.ki@phystech.edu; 19, Bolshoy Karetny Lane, 127051, Moscow, Russia; office phone: +7(910)531-8381.

Golubinskiy Andrey Nikolaevich — Ph.D., Dr. Sci., Associate professor, Head of Department, Russian Science Foundation (RSF). Research interests: machine learning, neural network modeling, automated control systems with artificial intelligence elements, speech signal processing. The number of publications — 250.
annikgol@mail.ru; 14, Solyanka St., 109240, Moscow, Russia; office phone: +7(910)346-6537.

УДК 004.82 **DOI** 10.15622/ia.25.3.6

В.К. Пимешков, М.Л. Никонорова, М.Г. Шишаев

ИНФОРМАЦИОННЫЙ ПОИСК В СОЦИАЛЬНЫХ МЕДИА НА БАЗЕ МУЛЬТИДОМЕННОЙ ДИНАМИЧЕСКОЙ СИСТЕМЫ ЗНАНИЙ

Пимешков В.К., Никонорова М.Л., Шишаев М.Г. **Информационный поиск в социальных медиа на базе мультидоменной динамической системы знаний.**

Аннотация. Задача информационного поиска заключается в нахождении информации, наилучшим образом удовлетворяющей информационную потребность пользователя. В условиях социальных медиа информационный поиск осложняется высокой динамичностью контента, тематической разнородностью и разнообразием ментальных моделей пользователей. В данной работе предлагается подход к решению задачи информационного поиска в таких условиях путем построения мультидоменной динамической системы знаний. Ее новизна заключается в объединении трех уровней семантики: проблемно-ориентированной, представленной онтологией метазадачи (описывающей цели поиска); доменно-специфической, реализованной через динамический многослойный граф знаний, построенный на основе пользовательского контента социальных медиа; и доменно-независимой, основанной на лексической базе данных и большой языковой модели. Граф знаний позволяет отразить различные контексты употребления понятий, соответствующие тематическим кластерам в коллекции документов. Такая интеграция позволяет учитывать эволюцию понятий, особенности дискурса и ментальные стереотипы участников коммуникации. Для оценки эффективности предложенной системы проведен эксперимент с использованием датасета публикаций из социальной сети “ВКонтакте” по проблемно-ориентированному мониторингу публикаций, где требуется отбор релевантных публикаций из нетематизированных источников. Для решения данной задачи предложена технология, основанная на использовании метрики расстояния между терминами запроса и терминами публикации в многослойном графе знаний. Результаты эксперимента с применением данной технологии подтверждают эффективность предложенной модели для задач информационного поиска по сравнению со стандартным поиском по ключевым словам и эмбеддинговыми моделями. В продолжении данного исследования планируется сформировать лексическую базу данных, а также рассмотреть возможность расширения модели за счет применения меры точечной совместной информации и методов векторного представления графов.

Ключевые слова: информационный поиск, социальные медиа, системы знаний, многослойный граф знаний, релевантность.

1. Введение. Информационный поиск (ИП) – процесс поиска информации, удовлетворяющей информационной потребности пользователя, является одной из главных составляющих современных процессов обработки данных в различных приложениях, в том числе работающих с данными социальных медиа. Общая схема ИП заключается в том, что пользователь адресует информационной системе запрос и получает в ответ на него некоторую выдачу, содержащую

запрошенную (с точки зрения системы) информацию. Объективным критерием успешности поиска является пертинентность сгенерированной системой выдачи. Пертинентность является трудноформализуемым параметром информационной выдачи, характеризующим разницу между содержимым последней и ожиданиями пользователя, сформулированными посредством запроса. Подчеркнем, что речь идет о разнице между информационными свойствами выдачи и именно ожиданиями, а не сформулированным запросом пользователя (в последнем случае говорят о релевантности выдачи запросу). Практика показывает, что оценка качества поиска при идентичных запросах и выдаче, но для разных пользователей, сформировавших запрос, может существенно отличаться.

Таким образом, для объективной оценки качества ИП нужно оценить пертинентность выдачи, а для адекватной оценки последней необходима модель знаний, описывающая ментальные стереотипы пользователя. С помощью такой модели можно учесть ожидания пользователя при оценке качества поиска. В случае социальных сетей модель стереотипов может описывать пользователя либо непосредственно, либо транзитивно через модель сообщества, частью которого он является. Заметим, что в качестве сообщества может рассматриваться аудитория некоторой тематической дискуссии, и тогда под пользователем будет подразумеваться не субъект, как таковой, а участник дискуссии, то есть текущая роль субъекта.

Существенная проблема заключается в том, что разные пользователи и сообщества характеризуются различными ментальными моделями (набор и способы интерпретации понятий), которые неоднозначны в том смысле, что способы интерпретации зависят от контекста использования, и которые, к тому же, способны меняться со временем. Контекст использования при этом определяется, в первую очередь, метазадачей поиска или, другими словами, прикладной задачей, решаемой пользователем в рамках взаимодействия с информационной системой. Построение исчерпывающе точных моделей знаний в таких условиях маловероятно. Однако ментальные стереотипы находят отражение в генерируемом пользователями социальных медиа контенте и коммуникативных актах, что создает возможность построения моделей, позволяющих учесть ожидания пользователя при формировании результатов поиска и обеспечить тем самым улучшение их качества с точки зрения пертинентности.

В данной работе в качестве такой модели, ориентированной на обеспечение приближения ИП к пертинентному результату,

рассматривается мультидоменная динамическая система знаний (МДСЗ). В предлагаемой модели проблема динамичности знаний решается за счет включения в модель графов знаний (ГЗ), формируемых и обновляемых на основе контента социальных медиа. То есть, в модели отражаются актуальные системы понятий и способы их интерпретации, оперируемые пользователями. Для решения проблемы разнородности знаний (неоднозначности интерпретаций понятий) ГЗ формируется в виде многослойной структуры, где каждый слой соответствует различным контекстам использования или доменам. В качестве связующего компонента в предлагаемой модели выступает онтология метазадачи (предметной области), задающая базовую, независимую от домена систему понятий. Терминологическая база при этом задается с помощью лексической базы данных (БД) и большой языковой модели. Таким образом, МДСЗ интегрирует в себе относительно стабильную проблемно-ориентированную семантику, представленную онтологией, динамичную доменно-специфическую семантику, представленную с помощью многослойного ГЗ, и доменно-независимую семантику, формируемую с помощью лексической БД и большой языковой модели.

Отметим, что в конечном итоге качество результатов поиска, по-прежнему, будет выражаться некоторой оценкой разницы между выдачей и поисковым запросом, поэтому далее в статье будет использоваться для обозначения этой разницы более традиционный термин “релевантность”. Однако предполагается, что в случае учета ментальных стереотипов автора запроса с помощью МДСЗ, такая оценка релевантности будет более близка к пертинентности, то есть к объективной оценке качества ИП. В этом заключается основная гипотеза представленного исследования.

Эффективность использования предложенной модели в ИП проверялась на примере задачи проблемно-ориентированного мониторинга социальных медиа, где требуется отбор релевантных публикаций из нетематизированных источников. Суть задачи состоит в том, чтобы отобрать из потока сообщений те (в данном случае – постов и комментариев текстового формата социальной сети “ВКонтакте”, опубликованных в пределах некоторого ограниченного набора пабликов), которые в наибольшей степени удовлетворяют целям мониторинга (метазадаче поиска). Для проведения эксперимента были использованы данные наблюдений за один год, в которых, изначально, в качестве критерия релевантности использовался простой факт вхождения в тело сообщения ключевых слов из заданного набора. Суть эксперимента заключалась в сравнении исходных результатов отбора с результатами отбора релевантных

сообщений с использованием МДСЗ. Для дополнительной верификации результатов авторского подхода к оценке релевантности использовались экспертно размеченные наборы данных.

Основной вклад (значимость) работы заключается в следующем:

- разработана концептуальная модель предметной области, отражающая мультидоменность и динамичность последней и позволяющая учесть специфику задачи ИП в контексте поиска данных в социальных медиа;
- предложен способ построения МДСЗ, интегрирующей доменно-специфичную, доменно-независимую (общую) и проблемно-ориентированную семантику;
- подтверждена эффективность предложенной модели на примере задачи мониторинга социально-политической обстановки на территории активности крупного предприятия, реализующего социально-ответственную политику;
- предложен подход к формированию многослойного ГЗ на базе тематического моделирования, позволяющий автоматизировать процесс построения мультидоменной модели знаний в условиях отсутствия явной тематизации источников данных;
- предложена технология оценки релевантности документов с помощью многослойного ГЗ, продемонстрировавшая эффективность на примере тестовой прикладной задачи мониторинга социально-политической обстановки на территории.

Наряду с рассмотренным здесь мониторингом социальных сетей, разработанные модели и технологии могут найти применение в задачах пополнения онтологий (пополнение начальной онтологии за счет ГЗ, полученного на данных, предположительно содержащих необходимые знания), а также в широком спектре задач поиска знаний (knowledge mining).

Дальнейший материал имеет следующую структуру. В разделе 2 представлено позиционирование данного исследования относительно современных подходов и направлений развития методологии ИП. В разделе 3 рассмотрены особенности задачи ИП в социальных медиа и формализована постановка данной задачи. В разделе 4 представлена модель МДСЗ, состоящая из проблемно-ориентированной онтологии, динамического многослойного ГЗ и лексической БД. В разделе 5 представлена технология определения релевантности публикаций с помощью многослойного ГЗ, в разделе 6 приведены результаты её экспериментального тестирования.

2. Обзор релевантных исследований в области технологий информационного поиска. Основной целью современного ИП является предоставление конечному пользователю максимально релевантной информации по его запросу. Этот процесс можно разделить на два этапа: поиск и ранжирование [1]. На первом этапе извлекается набор исходных документов, наиболее релевантных запросу, а на втором этапе происходит ранжирование полученных документов на основе их оценки релевантности. На этапе поиска могут использоваться как традиционные методы, например, логический (Boolean) поиск и TF-IDF, так и методы семантического поиска, например, расширение запроса, латентно-семантический анализ и тематическое моделирование, а также методы глубокого обучения и гибридные. На этапе ранжирования также применяются различные методы классического и глубокого машинного обучения, включая обучение с подкреплением, контекстуальные эмбеддинги и механизмы внимания, чтобы научиться ранжировать документы на основе критериев запроса пользователя. Одной из распространенных количественных оценок релевантности являются показатели точности (доля правильных результатов в их общем количестве) и полноты (доля правильных результатов, попавших в результирующую выдачу).

Рассматриваемую в данной работе технологию определения релевантности с помощью многослойного ГЗ можно отнести к методам семантического поиска, а именно к методам расширения запроса. Подобные методы модифицируют исходный поисковый запрос для улучшения точности и полноты результатов поиска [2, 3]. Такая модификация может включать добавление синонимов, использование нормализованных и словообразовательных вариантов терминов из запроса, добавление терминов, встречающихся в том же контексте и т.д. Данные методы, как правило, позволяют повысить качество и полноту поиска, но существует риск избыточного расширения, что приводит к большому числу ложноположительных результатов.

Примером работы, реализующей методы расширения запроса, является [4], где применяют контролируемые словари, онтологии и контекстно-свободные дистрибутивные языковые модели для расширения поискового запроса, исходящего от информационных работников. Для онтологических ресурсов применяется простой поиск связанных терминов в виде гиперонимов, гипонимов и т.д., а языковые модели используются для генерации запросов из связанных терминов.

Во многих системах ИП активно используются ГЗ. Они позволяют лучше понять информационную потребность пользователя, его запрос или расширить представления о документах на таком

уровне, который нельзя достичь с помощью одних лишь ключевых слов как таковых [5]. Дополнительно ГЗ открывают возможность сразу отвечать на запрос пользователя в рамках поиска сущностей (когда данный ответ заложен в понятиях и/или структуре графа в явном виде), а также дают возможность исследовать связанные сущности из запроса или документов, включая справочную информацию по ним или их связям [6].

С вовлечением в процессы обработки сложных, семантически неоднородных данных возникает потребность в структуризации знаний, представленных графом. Это осуществляется с помощью многослойных ГЗ, где их элементы разделены на несколько слоев или измерений, отражающих различные типы связей между сущностями (например, социальные, профессиональные или контекстуальные) или разные точки зрения на предметную область (структура понятий и связей на каждом слое различна). Известным примером многослойного ГЗ является KnowWhereGraph [7], включающий около 13 млрд. триплетов и более 30 слоев (пространственные, демографические, инфраструктурные, сельскохозяйственные данные и др.), что обеспечивает его применимость в широком спектре задач.

Таким образом, можно отметить, что основным подходом к улучшению качества ИП являются различные манипуляции с поисковым запросом (как правило – расширение), а ГЗ предоставляют технологическую базу для формирования в результате таких манипуляций поисковых запросов, в большей степени соответствующих ожиданиям пользователя. Это достигается за счет адекватного формального представления знаний о предметной области, которые используются в процессе генерации поискового запроса и его результата. При этом одним из способов повышения эффективности использования ГЗ в случае сложных неоднородных предметных областей является его разбиение на слои, каждый из которых представляет собой специфическую точку зрения на предметную область. В данной работе развивается подход к ИП с использованием ГЗ в направлении повышения адекватности структуры знаний, представленной многослойным графом, за счет применения тематического моделирования. Использование тематического моделирования позволяет структурировать знания в ситуации, когда нет априорной тематизации источников данных, на базе которых строится ГЗ.

Еще одним способом улучшения качества поиска является использование дополнительной, внешней по отношению к запросу, информации при формировании результатов ИП. Примером такой

информации является структура коммуникаций и различные метаданные, представленные в социальных медиа. Область исследований, в рамках которой для улучшения работы систем поиска используются также показатели взаимодействия пользователей с контентом (лайки, просмотры), поведение пользователей и знания сообществ, называется “социальный информационный поиск” (Social Information Retrieval) [8]. Примером работы в данной области является [9], где авторы применяют технологию расширения запроса, включающую информацию о пользователях для улучшения работы ИП. В начале из документов выбираются важные термины, которые затем кластеризуются с использованием k-means. На первом этапе расширения запроса выбирается самый похожий на запрос кластер, который расширяется терминами из этого кластера. Затем оцениваются комментарии пользователей к документам в выбранном кластере и эти оценки используются для последующего расширения запроса.

Таким образом, существующие работы по проблематике социального информационного поиска показывают, что учет дополнительной информации, доступной при ИП в социальных медиа, позволяет повысить качество результата. В данной работе для повышения эффективности ИП использовалась информация о структуре коммуникаций (причинно-следственная структура публикаций и комментариев) в социальных медиа, представленная в виде соответствующей формальной модели.

Можно также отметить, что в целом современное развитие методов ИП фокусируется на оперировании семантикой запросов и поисковых выдач, как альтернативе более примитивным интерпретациям запроса и его результата как последовательностей символов. Важной проблемой, возникающей при реализации семантического поиска в социальных медиа, является разнородность используемых понятийных систем, представленных предметными (а в ряде случаев – мультипредметными) и общими знаниями, а также соответствующей терминологией. Для решения этой проблемы широко используются модели машинного обучения. Например, в работе [10] применяется глубокое обучение для идентификации разговорных выражений, описывающих симптомы COVID-19, в тексте постов Twitter. Затем они нормализуются и сопоставляются с внешним источником данных для получения словаря, который может использоваться врачами и неспециалистами для лучшего понимания того, как пользователи описывают свои симптомы в социальных медиа.

В работе использовались переобученные модели глубокого обучения для устранения лексической многозначности, характерной для коммуникаций в социальных сетях, при этом устанавливается соответствие между терминами общей лексики и терминами, используемыми в различных доменах.

3. Особенности и формальная постановка задачи информационного поиска в социальных медиа. Решаемая метазадача ИП в социальных медиа задает границы предметной области и таким образом определяет некоторый стабильный домен (задающий способы интерпретации понятий и связей между ними), в рамках которого осуществляется поиск и интерпретация результатов.

В то же время коммуникации в социальных медиа, как правило, охватывают множество доменов. Одно и то же понятие может встречаться в двух доменах, но иметь различные ассоциативные связи (различную семантику). Например, “вакцинация” в контексте COVID-19 будет связана с одними понятиями, а в контексте ветеринарии – с другими. Какие-то домены при этом существуют практически постоянно в социальном дискурсе без особых изменений, другие могут появляться, исчезать и видоизменяться достаточно быстро вместе с тем, как разгорается и угасает интерес к социальной дискуссии на какую-либо тему. Таким образом, домены, по существу, являются отражением ментальных моделей сообществ и пользователей социальных медиа.

Исходя из этого, модель предметной области должна обладать свойствами мультидоменности и динамичности, а точка зрения на решаемую проблему – задавать некоторый постоянный домен:

$$S\big(D(t)\big) = \{Domain^P, Domain(t)_1, \dots, Domain(t)_n\},$$
$$\forall\, Domain_i(t), Domain_j(t) \in S(t), Domain_i(t) \neq Domain_j(t),$$

где $Domain^P$ задает стабильную часть предметных знаний.

Формально определим домен как систему, состоящую из концептов, связей, классов связей, функции, определяющей связи между концептами, и функции, определяющей вес связи и существующую в статичном виде только в некоторый момент времени t:

$$Domain(t) = (Concept^D(t), Relations^D(t),$$
$$RelTypes, r\_assignment, w\_estimating),$$

где $Concept^D(t)$ – множество концептов/терминов домена в момент времени t; $Relations^D(t) \subseteq Concept^D(t) \times Concept^D(t) \times RelTypes$ –

множество связей между вершинами, где каждая связь rel – это тройка (c_i, c_j, rt), где $rt \in RelTypes$ в момент времени t, $RelTypes$ – множество типов связей; $w\_estimating: Relations^D(t) \to R_{>0}$ – функция, присваивающая каждой связи некоторый вес; $r\_assignment: Concept^D(t) \times Concept^D(t) \to B(RelTypes)$ – функция, присваивающая каждой паре концептов некоторый тип или типы неповторяющихся связей из $RelTypes$: $r\_assignment(c_i, c_j) \subseteq RelTypes$ – определяет набор отношений между вершинами c_i и c_j; $r\_assignment(c_i, c_j) = \emptyset$, если связей между терминами нет. $B(RelTypes)$ – множество всех подмножеств $RelTypes$.

Структурные ограничения:

- концепты не могут быть связаны сами с собой:

$$\forall c_i \in Concept^D(t), \forall r \in RelTypes, r\_assignment(c_i, c_i) = \emptyset;$$

- одновременное существование между двумя вершинами только уникальных связей:

$$\forall c_i, c_j \in Concept^D(t), \forall r \in RelTypes, \left|\{c_i, c_j, rt\}\right| \leq 1.$$

Учет времени:

$$Concept^D(t + \Delta t) = Concept^D(t) \cup Concept_{new}(t) - Concept_{removed}(t), Relations^D(t + \Delta t) = Relations^D(t) \cup Relations^D{}_{new}(t) - Relations^D{}_{removed}(t).$$

Различие доменов. Пусть есть два домена – $Domain_1(t)$ и $Domain_2(t)$:

$$Domain_1(t) = (Concept_1^D(t), Relations_1^D(t), RelTypes_1, r\_assignment, w\_estimating),$$

$$Domain_2(t) = (Concept_2^D(t), Relations_2^D(t), RelTypes_2, r\_assignment, w\_estimating).$$

Будем утверждать, что $Domain_1(t) \neq Domain_2(t)$, если выполняется хотя бы одно из следующих условий:

1. Домены различаются по составу множества концептов:

$$Concept_1^D(t) \neq Concept_2^D(t);$$

2. Домены различаются по составу множества связей:

$$Relations_1^D(t) \neq Relations_2^D(t);$$

3. Домены совпадают по составу множеств концептов и связей, но различаются веса этих связей:

$$\exists rel \in Relations_1^D \cap Relations_2^D,$$
$$w\_estimating(rel, Domain_1) \neq w\_estimating(rel, Domain_2).$$

Коммуникации в социальных медиа материализуются в форме публикаций и комментариев, которые далее именуются документами. Все множество документов, являющихся объектами поиска в контексте рассматриваемой метазадачи, образуют коллекцию. Процесс появления публикаций в социальных медиа не случаен и характеризуется наличием причинно-следственной структуры. Так, например, комментарии, как правило, являются реакцией на публикацию или ассоциированные с ней понятия. Или, например, новостная публикация может быть обновлением информации касательно события, уже произошедшего и получившего упоминание в социальных медиа. Для адекватного представления специфики социальных медиа в контексте задач ИП в модельном представлении необходимо отобразить структуру коммуникации на коллекцию документов.

Соответственно, определим мультидоменную динамическую коллекцию документов как:

$$MDDC(t) = \{D(t), L, S(t)\},$$

где $D(t) = \{d_1, \ldots, d_n\}$ – множество документов источника в некоторый момент времени t. Каждый документ $d \in D(t)$ представлен множеством предложений $d = \{snt_1, \ldots, snt_i, \ldots, snt_n\}$, где $snt \in d$, $snt = \{wrd_{i1}, sep_{i1}, \ldots, wrd_{im}, sep_{ik}\}$ – предложение, состоящее из слов wrd_{ij} и знаков препинания sep_{ib}; $L \subseteq D \times D$ – асимметричное транзитивное отношение, задающее последовательность документов в иерархии (причинно-следственную структуру). $d_1 L d_2$ означает, что документ d_2 следует после документа d_1. Такая цепочка состоит из веток, представляющих собой последовательности Br документов из $D(t)$ ($Br \subseteq D(t)$), удовлетворяющие условию:

$$Br = \{d_1, \ldots, d_N\}: \forall i < j, d_i L d_j .$$

Полная цепочка (полное дерево, CS) задается корневым документом d и всеми транзитивно связанными с ним документами $Br(d)$.

$S(t)$ – множество доменов в момент времени t.

Функция определения принадлежности документа d к домену s в момент времени t имеет следующий вид:

$$d\_entry(d,s,t)\colon D(t) \times S(t) \to [0,1].$$

Таким образом, решаемая задача ИП в социальных медиа может быть формально описана следующим образом. Имеется мультидоменная динамическая коллекция документов $MDDC$ и запрос q, сформулированный пользователем. Система должна вернуть упорядоченное подмножество релевантных документов следующего вида:

$$(D', \preceq) = \{d \in D(t) | f_{rel}\,(d,q) > th\},$$
$$f_{rel} : D \times \{q\} \to E,$$
$$d_i \preceq_q d_j \Leftrightarrow f_{rel}\,(d_i\,, q) \leq f_{rel}\,(d_j\,, q),$$

где th – некоторое пороговое значение релевантности, E – множество оценок (например, множество вещественных чисел).

4. Модель мультидоменной динамической системы знаний. Для решения исследуемой, в рамках ИП, задачи определения релевантных публикаций предлагается использовать МДСЗ, состоящую из трех основных частей (Рис. 1):

1. Проблемно-ориентированная часть, представленная онтологией. Она задает некоторую точку зрения на решаемую проблему и данные соответственно, при этом не ограничивая остальную систему.

2. Мультидоменная часть системы, представленная мультидоменным динамическим ГЗ, отражает знания, заложенные в динамических мультидоменных структурированных коллекциях документов, и ментальные стереотипы сообществ или пользователей, породивших эти коллекции соответственно.

3. Доменно-независимая часть, представленная лексической БД, необходима для разрешения лексической многозначности (word sense disambiguation) и является отправной точкой для согласования понятий проблемно-ориентированной и мультидоменной частей. Согласование концептов, хранящихся в системе, предлагается производить с помощью большой языковой модели (LLM), так как она в некотором роде содержит усредненную доменно-независимую семантику.

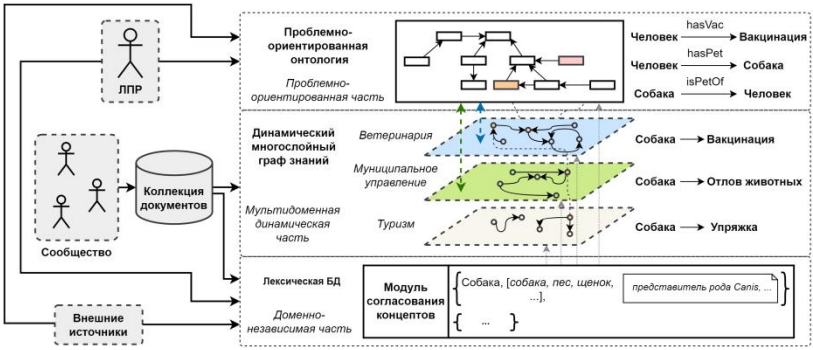


Рис. 1. Концептуальная модель МДСЗ

Процесс построения МДСЗ начинается с формирования лексической БД, которая задается с помощью внешних источников данных, и проблемно-ориентированной онтологии, задаваемой с помощью эксперта или внешних источников. Лексическая БД пополняется также концептами, входящими в онтологию, и наоборот – при создании онтологии могут использоваться концепты, уже заданные ранее в лексической БД. Далее осуществляется построение ГЗ на основе мультидоменной коллекции документов социальных медиа, которые предварительно преобразовываются в структуры коммуникаций с последующим выделением в них концептов. Данный процесс продемонстрирован на рисунке 2.

Пополнение МДСЗ новой информацией происходит при необходимости в процессе её применения. Онтология обновляется при обнаружении неполноты или несоответствия текущей структуры или экземпляров по отношению к решаемой прикладной задаче и ожиданиям пользователя. При этом новые экземпляры могут происходить как и из ГЗ, так и из каких-либо внешних источников, если того требует решаемая задача при условии, что они удовлетворяют заданным в онтологии свойствам и ограничениям. Онтология перестраивается полностью при кардинальной смене предметной области мониторинга, например, при смене пользователя. ГЗ пополняется новой статистической информацией из наблюдаемых в рамках мониторинга документов в пакетном режиме. Перестройка или дополнение ГЗ новыми слоями происходит при значительном изменении семантических зависимостей в наблюдаемых данных, что, как правило, происходит со временем или при смене наблюдаемого множества источников. Лексическая БД пополняется новыми терминами и/или концептами, встречающимися

в обрабатываемых документах, по мере необходимости в течение всего времени работы системы.

Таким образом, определение релевантности с помощью предлагаемой системы может производиться как отдельно на каждом из её уровней, так и совместно для достижения наилучшего результата. Так, лексическая БД обеспечивает решение проблемы лексической многозначности, что позволяет использовать систему для простого поиска по ключевым словам, принимая во внимание различные лексикализации понятий. Многослойный ГЗ позволяет взглянуть на понятия и отношения между ними с точки зрения актуальных тематических представлений сообществ и, при необходимости, комбинировать таковые без перестройки системы. Для определения релевантности с его помощью предлагается использовать метрику расстояния между терминами запроса и документов (см. раздел 5). Онтология, в свою очередь, задает проблемно-ориентированную точку зрения на решаемую проблему с осмысленными отношениями и правилами вывода, с помощью которых, в том числе, можно определять релевантность документов. Далее рассмотрим компоненты МДСЗ более подробно.

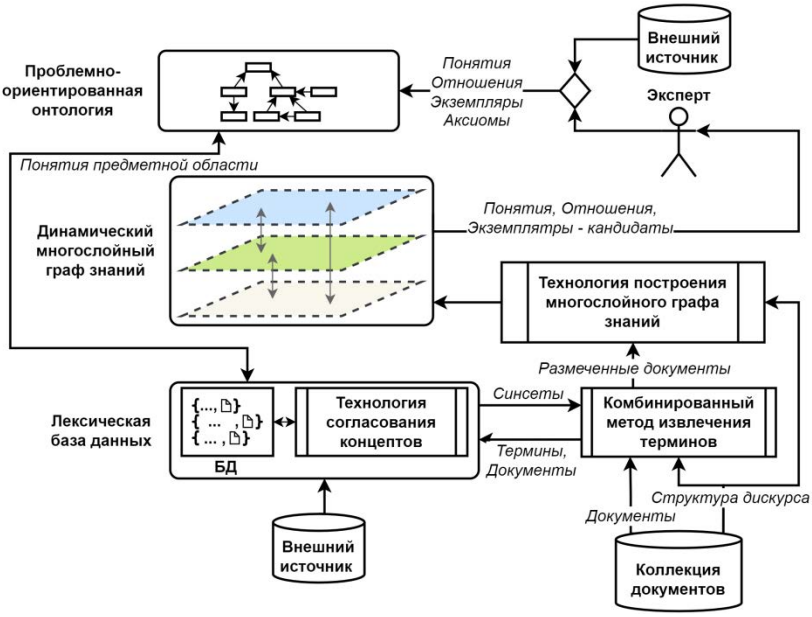


Рис. 2. Функциональная структура МДСЗ

4.1. Лексическая БД. Лексическая БД содержит структурированную информацию о различных словах и отношениях между ними, а также сопутствующую лингвистическую информацию, например, о частоте использования слов или морфологическую информацию. От тезауруса или словаря её отличает ряд признаков (таблица 1).

Таблица 1. Различия между лексической БД и тезаурусом

Признак	Лексическая БД	Тезаурус
Цель	Хранение лексических единиц, отношений между ними, а также сопутствующей лингвистической информации без привязки к конкретной области знаний, чаще с привязкой к конкретному языку.	Хранение лексических единиц конкретной области знаний, семантических отношений между ними (синонимы, антонимы, паронимы, гипонимы, гиперонимы и т.п.) и определений для них.
Структура	Сложная сеть из лексических единиц, их отношений, признаков и сопутствующей лингвистической информации.	Простые наборы лексических единиц с их определениями.
В каких задачах используется	Обработка естественного языка, исследования, касающиеся лингвистики.	Написание различных текстов, для получения справочной информации при переводе, для описания предметных областей с помощью понятий.

Исходя из этого, можно заключить, что тезаурусы предназначены для повседневного использования, в то время как лексические БД используются в исследовательской работе. Например, можно выделить лексическую БД WordNet [11], которая содержит обширную информацию об английских частях речи, включая существительные, глаголы, прилагательные и наречия. Основой структуры WordNet являются синсеты – специальные группы слов, принадлежащих к одной части речи и обладающих высокой степенью синонимичности в конкретном контексте. В качестве примера можно привести синсет, включающий лексемы {car, auto, automotive, machine, motorcar}, которые взаимозаменяемы при обозначении транспортного средства определенного типа. Каждый синсет сопровождается пояснительным комментарием, который содержит краткое описание обозначаемого понятия. Например, для вышеупомянутого синсета приводится определение: “четырехколесный” – обычно приводимый в движение двигателем внутреннего сгорания. Особую значимость в архитектуре WordNet представляет система семантических связей между синсетами, которая включает такие типы отношений, как

гипонимия, гиперонимия и меронимия, что позволяет выстраивать сложные семантические сети между лексическими единицами. Также в качестве примера можно привести CELEX [12], включающий в себя три лексические БД для голландского, английского и немецкого языков, содержащие орфографическую, фонологическую, морфологическую, синтаксическую и частотную информацию.

Формально разрабатываемую лексическую БД можно описать следующим образом:

$$LD = \{Concept^{LD}, Vocabulary, Context\},$$

где $Concept^{LD}$ – множество концептов БД (включая концепты из внешних источников, онтологии и полученные в процессе обработки документов), где каждый отдельный концепт c представлен четверкой $(id, name, synset, context)$, где $id \in ID$ (ID – множество уникальных меток) – уникальная метка концепта в рамках БД; $name \in Vocabulary$ – наименование концепта на естественном языке; $synset \subset Vocabulary$ – множество терминов (словоформ и синонимов) данного концепта; $context$ – некоторый контекст употребления концепта. Для любого термина может существовать несколько контекстов и для любого контекста может существовать несколько терминов, но для одного синсета и, соответственно, концепта существует только один контекст; $Vocabulary$ – множество терминов или лексических представлений (слов или словосочетаний); $Context$ – множество контекстов лексической БД.

Контексты в рамках данной системы выполняют функцию определения некоторого смысла для концепта в виде, приспособленном для обработки компьютером. Соответственно, в качестве способа представления концептов предлагается использовать векторные представления, полученные из больших языковых моделей. Такой выбор обоснован богатым семантическим кодированием, которое обеспечивают подобные векторные представления, фиксируя не только поверхностное лексическое сходство, но и более глубокие контекстные и функциональные связи между концептами. Формирование концептов (сопоставление и распределение терминов по смыслу) в рамках лексической БД можно рассматривать как задачу кластеризации в многомерном семантическом пространстве, что соответствует теории дистрибутивной семантики. Таким образом, для динамической группировки связанных терминов под общими концептами и автоматизированной адаптации к ранее не встретившимся терминам

можно использовать метрики векторного сходства, например, косинусное расстояние. Последние исследования [13 – 15] применения больших языковых моделей и, в частности, векторных представлений из них в области обработки естественного языка подтверждают эффективность их использования для решения подобных задач.

Взаимодействие внутри системы с лексической БД предусматривает наличие функций пополнения, разметки документа, поиска концепта и получения информации.

Пополнение лексической БД необходимо для её поддержания в соответствии с обрабатываемыми данными и производится следующим образом:

1. Если такого термина и контекста еще нет в БД, то добавляем термин в словарь (вместе с его контекстом) и добавляем соответствующий ему концепт.

2. Если такой контекст уже существует, а термин – нет, то добавляем термин в словарь, находим синсет, содержащий этот контекст, и добавляем в этот синсет новый термин.

3. Если такой термин уже существует, а контекст – нет, то добавляем контекст в БД, связываем термин с этим концептом и создаем соответствующий концепт.

Основной задачей лексической БД является разметка документов, которая заключается в сопоставлении терминов документа с известными концептами. Разметку также предполагается производить при помощи векторных представлений аналогично функции пополнения. Альтернативно, подобное сравнение терминов возможно с использованием простых правил сопоставления текстовых представлений, статистических мер или, например, тематического моделирования. При этом, если встречается термин или концепт, не присутствующий в лексической БД, то он добавляются в нее.

Получение информации из лексической БД предполагает обработку конкретных запросов, например, получение конкретного концепта, синсета или набора терминов.

Для управления такой БД необходима соответствующая специфике задачи система управления БД. В нашем случае основными кандидатами являются векторные БД (например, ChromaDB [16], Milvus [17]), созданные специально для эффективной работы с многомерными векторными представлениями и позволяющие интегрировать в процесс обработки данных большие языковые модели.

4.2. Комбинированный метод извлечения терминов. Метод нацелен на извлечение как предметных, так и общеупотребительных терминов. На вход метод принимает экземпляр структуры

коммуникации, состоящий хотя бы из одного корневого документа. Затем для каждого документа производится извлечение терминов онтологии, именованных сущностей, осмысленных словосочетаний и контекстно значимых существительных. Осмысленность словосочетаний оценивается комбинацией извлекающего правила на основе контекстно-свободной грамматики и меры коллокации. Контекстная важность существительного определяется исходя из статистической оценки его употребления в рамках соответствующего экземпляра структуры коммуникации. После извлечения всех терминов-кандидатов (все термины, кроме терминов онтологии, считаются кандидатами на этом этапе), происходит их фильтрация по шагу извлечения, стоп-словам, длине и частоте употребления. Данный процесс представлен на рисунке 3 и более подробно описан в работе [18].

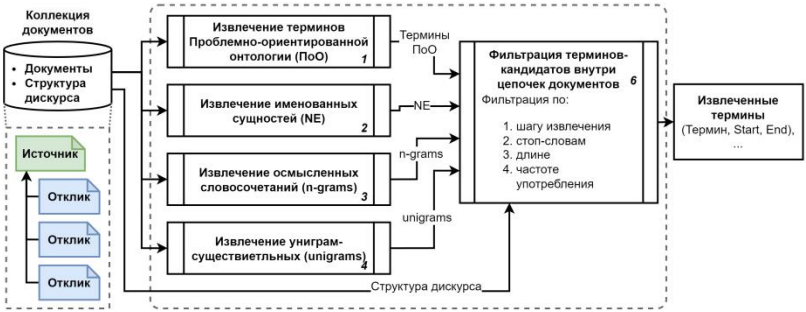


Рис. 3. Схема работы комбинированного метода извлечения терминов

4.3. Динамический многослойный ГЗ. В данной работе используется контекстно-зависимый подход к разбиению ГЗ на слои, основанный на тематической кластеризации документов, полученной в результате тематического моделирования [19]. Такой подход позволяет интерпретировать данные с помощью скрытых тем, идентифицированных в исходном наборе документов. Полученный граф представляет концепты как вершины со взвешенными направленными ребрами, захватывающими статистически значимые паттерны их совместного употребления, улучшенные выходными данными тематического моделирования. Реализуя разбиение на основе контекста, рассматриваемое как отдельные слои графа или взаимосвязанные подграфы, подход облегчает выявление значимых связей даже для тем с низкой распространенностью. Отметим, что эффективность предлагаемого подхода будет зависеть, в том числе от свойств и качества используемой тематической модели.

Формально разработанную технологию, использующую тематическую модель в качестве основы для разбиения на контексты, можно представить следующим образом.
Дано:
$MDDC$ – мультидоменная динамическая коллекция документов.
Найти:

$$G \coloneqq (V, E, K, W_K, w_E, s, t, r),$$

где V – множество вершин – уникальные концепты из коллекции документов $Concept^{MDDC} \subset Concept^{LD}$; E – множество ребер – взвешенные направленные ассоциативные отношения между концептами; K – множество меток контекстов, заданное темами тематической модели; W_K – множество весов ребер, заданное для каждого контекста; $w_E: E \to W_K$ – функция, присваивающая вес каждому ребру; $s: E \to V$ – функция, присваивающая каждому ребру начальную вершину; $t: E \to V$ – функция, присваивающая каждому ребру конечную вершину; $r: E \to K$ – функция, присваивающая каждому ребру метку контекста.

Процесс построения ГЗ включает в себя три основных этапа:

1. Получение размеченных экземпляров структур коммуникаций.
2. Обучение тематической модели.
3. Построение многослойного ГЗ (Рис. 4).

На первом этапе исходная коллекция документов подвергается предварительной обработке, в рамках которой из нее исключаются несодержательные документы и мусорные конструкции. После чего коллекция, в соответствии со структурой, преобразуется в набор экземпляров структур коммуникаций CS, где каждый экземпляр задает полное дерево документов (например, публикация и все связанные с ней комментарии). Затем, с помощью авторского комбинированного метода извлечения терминов, каждый экземпляр размечается терминами:

$$TermExtrMethod(CS) \to CS^{term\ marked}.$$

В результате исходный набор экземпляров структур коммуникаций дополняется размеченными терминами. После этого, с помощью лексической БД, данный набор размечается концептами:

$$annotate(CS^{term\ marked}) \to CS^{concept\ marked},$$

что обеспечивает устранение лексической многозначности. Затем концепты в каждом экземпляре фильтруются для выявления контекстно важных. На данном этапе работы контекстная важность термина определяется исходя из статистической оценки его употребления в рамках соответствующего экземпляра структуры коммуникации.

На втором этапе отфильтрованные экземпляры структур коммуникаций используются при формировании датасета для тематического моделирования, и производится само тематическое моделирование, результатом которого являются тематические распределения этих экземпляров и концептов:

$$TopicModelConst(CS, K) \rightarrow \{\Theta, \Phi\},$$

где Θ – матрица распределения экземпляров структур коммуникаций по темам, Φ – матрица распределения концептов по темам.

На третьем этапе производится подсчет статистики совместного употребления концептов с учетом полученных тематических распределений и заполнение соответствующих тематизированных матриц: $f(CS, \Theta, \Phi) \rightarrow CM$ – набор матриц совместного употребления, где элемент матрицы совместного употребления по k-ой теме вычисляется по формуле:

$$cm_{ij}^{k} = \sum_{CS^P=1}^{M} \theta_{CS^P}^{k} \cdot CS_{ij}^{P} \cdot \left(\frac{\phi_i^k}{\max(\phi_i)} + \frac{\phi_j^k}{\max(\phi_j)} \right), \quad (1)$$

где $\theta_{CS^P}^{k}$ – значение принадлежности экземпляра CS^P теме k; CS_{ij}^{P} – значение совместного употребления концептов i и j в рамках CS, получаемое функцией подсчета совместного употребления в экземпляре структуры коммуникации: $fp(CS^T) \rightarrow CS^P$ вида $< i, j, c_{ij} >$, где i, j – концепты, c_{ij} – значение их совместного употребления в конкретном экземпляре CS; ϕ_i^k – значение принадлежности концепта i теме k; $\max(\phi_i)$ – максимальное значение принадлежности концепта i темам.

Далее выделяются значимые отношения из полученных матриц следующим образом:

1. Для каждого слоя, соответствующего теме из тематической модели, выбирается процент самых сильных связей, соответствующий распределению концептов по темам в тематической модели;

2. Удаляются незначимые значения в соответствии с эвристически заданным порогом:

$$CS_k^P(i,j) > \overline{x}\left(\overline{x}\left(CS_k^P(i,)\right), \overline{x}\left(CS_k^P(,j)\right), \overline{x}\left(CS_k^P\right)\right), \quad (2)$$

где $\overline{x}\left(CS_k^P(i,)\right)$ – среднее значение совместного употребления для i, $\overline{x}\left(CS_k^P(,j)\right)$ – среднее значение совместного употребления для j, $\overline{x}\left(CS_k^P\right)$ – среднее значение совместного употребления для всех концептов в соответствующей матрице (теме) k.

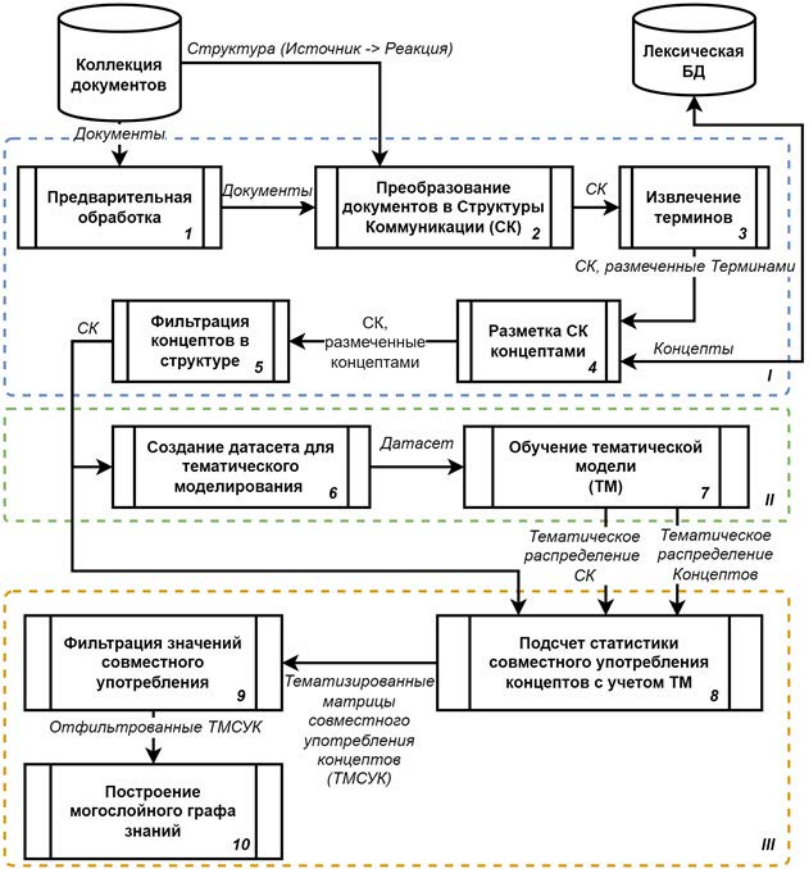


Рис. 4. Технология формирования многослойного ГЗ с применением тематического моделирования

В результате получается набор матриц со значениями совместной встречаемости концептов, соответствующий тематической модели. Набор уникальных концептов интерпретируется как вершины графа, а ненулевые значения матриц совместной встречаемости – как взвешенные ассоциативные отношения между ними.

4.4. Проблемно-ориентированная онтология. Как было отмечено ранее, данная онтология задает некоторую отправную точку зрения на решаемую проблему. Соответственно, в ней содержатся базовая система понятий (классы), их свойства, отношения, экземпляры и аксиомы.

Проблемно-ориентированная онтология задается некоторым внешним источником знаний EKS или экспертными знаниями EK:

$$f: EKS \cup EK \rightarrow PoO$$
$$PoO = \{C, P, R, I, A\},$$

где: C – множество классов; P – множество свойств классов; R – множество отношений между классами; I – множество экземпляров. При этом $\forall r \in R: I(r) \subseteq I \times I,$ где $I(r)$ – множество пар экземпляров, связанных отношением r; A – множество аксиом.

В рамках разрабатываемой системы знаний онтология описывает предметную область поиска или, иначе говоря, контекста, в котором находится пользователь системы. Для работы системы в целом наибольшую роль играет наполнение онтологии экземплярами, что повышает точность и полноту извлечения заведомо релевантных терминов с помощью метода извлечения терминов. Далее эти термины встраиваются в ГЗ, где играют важную роль при определении релевантности документов на его уровне. Отношения, заложенные в онтологию, потенциально могут быть использованы для расширения запроса или документов или для определения/уточнения релевантности на её уровне на основе заложенных в нее аксиом. Однако на данном этапе работы определение релевантности на основе аксиом онтологии не производится и является возможным направлением дальнейших исследований.

На текущий момент в онтологии описана базовая структура классов, представляющих собой основные категории именованных сущностей (Локации, Организации, Личности), включающих подклассы, определены отношения между ними, некоторые свойства классов, а также заданы наиболее значимые экземпляры.

5. Технология определения релевантности публикаций с помощью многослойного графа знаний. Для определения

релевантности с помощью полученного многослойного ГЗ предлагается использовать метрику расстояния между терминами запроса и терминами публикации в наиболее релевантных слоях графа. Релевантность слоя оценивается исходя из тематического распределения запроса, полученного на основе используемой тематической модели. Таким образом, наиболее релевантными считаются слои, тематически соответствующие запросу (Рис. 5). Подобную технологию можно интерпретировать как расширение запроса при помощи ГЗ, содержащего объективные представления о стереотипах пользователей.

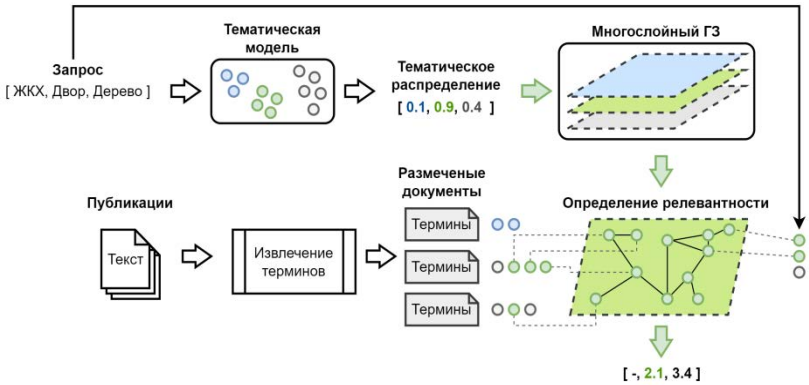


Рис. 5. Определение релевантности с помощью расстояния на графе

Перед определением расстояния на графе имеющиеся взвешенные ассоциативные связи преобразуются в расстояния, обратно пропорциональные весу и нормализованные в диапазоне значений (0, 1). Данное преобразование проводится на матрицах, на основе которых строятся графы, в соответствии с формулой:

$$N_{ij}^{k} = \begin{cases} 1 - \dfrac{M_{ij}^{k} - M_{adjmin}^{k}}{M_{adjmax}^{k} - M_{adjmin}^{k}}, if\ M_{ij}^{k} \neq 0 \\ 0\ , if\ M_{ij}^{k} = 0 \end{cases},$$

где M_k – матрица весов слоя k; $M_{adjmin}^{k} = M_{min}^{k} \cdot 0{,}99999$ – скорректированный минимум матрицы M для слоя k; $M_{adjmax}^{k} = M_{max}^{k} \cdot 1{,}00001$ – скорректированный максимум матрицы M для слоя k.

Соответственно, для оценки релевантности публикации запросу в рамках одного слоя предлагается использовать среднее значение минимальных расстояний от терминов публикации P^t до терминов запроса Q^t:

$$D(P^t, Q^t) = \frac{1}{m}\sum_{i=1}^{m} \min_{j=1,\dots,n} d(p_i, q_j), \quad (3)$$

где $d(p_i, q_j)$ – расстояние (вес кратчайшего пути) от $p_i \in P^t$ до $q_j \in Q^t$.

6. Эксперимент (определение релевантности публикаций с помощью многослойного графа знаний). Рассматриваемые в данной работе модели и методы ориентированы, прежде всего, на задачи мониторинга потоков сообщений, в рамках которых требуется в потоковом режиме определять релевантность появляющихся сообщений. При этом задача осложняется изменчивостью лексикализации семантики предметной области, а также тем, что критерии отбора могут формироваться еще до появления первых релевантных сообщений, что ограничивает возможность использования подходов на базе “эталонных” сообщений, в том числе представленных в виде эмбеддингов. Кроме того, в практических задачах ИП запрос формируется, как правило, неподготовленным пользователем. В таких условиях фактически единственным способом задания критериев релевантности информационной выдачи являются ключевые слова. При этом опыт эксплуатации систем мониторинга сообщений показывает, что начальный поисковый запрос в виде набора ключевых слов претерпевает постепенное расширение, обусловленное уточнением ожиданий пользователя в отношении релевантных результатов.

Исходя из вышеизложенного, в рамках эксперимента проверялась гипотеза о том, что путем формирования многослойного ГЗ из данных социальных медиа возможно решить задачу определения релевантности публикаций не хуже технологии, использующей только ключевые слова.

В рамках решаемой задачи мониторинга социально-политической обстановки набор ключевых слов, по существу, является запросом для поиска релевантных публикаций. Получение и поддержание такого набора в актуальном состоянии является достаточно трудоемкой задачей в силу того, что необходимо обладать не только знанием темы, в рамках которой осуществляется поиск, но иногда и интернет-жаргоном, связанным с ней. С одной стороны,

такой набор может позволить достаточно точно определить релевантность некоторых документов, но, с другой стороны, он требует постоянной модификации для включения новых терминов и их лексикализации или исключения более нерелевантных терминов. Однако стоит отметить, что расширение данного набора терминов может приводить к получению большого числа ложноположительных результатов, т.е. нерелевантных публикаций. Так, например, набор ключевых слов из реального проекта (по проблемно-ориентированному мониторингу публикаций), используемый в рамках данного эксперимента, в начале содержал около 15 терминов, затем пополнялся в процессе использования, в течение нескольких лет. В итоге, расширился до 133 терминов (включая словосочетания и аббревиатуры, которые покрывают ряд различных тематик) на момент проведения эксперимента.

Суть эксперимента заключалась в оценке эффективности предлагаемой технологии определения релевантности с помощью многослойного ГЗ посредством сравнения с технологией, основанной на ключевых словах (точное совпадение и основы слов) (Рис. 6).

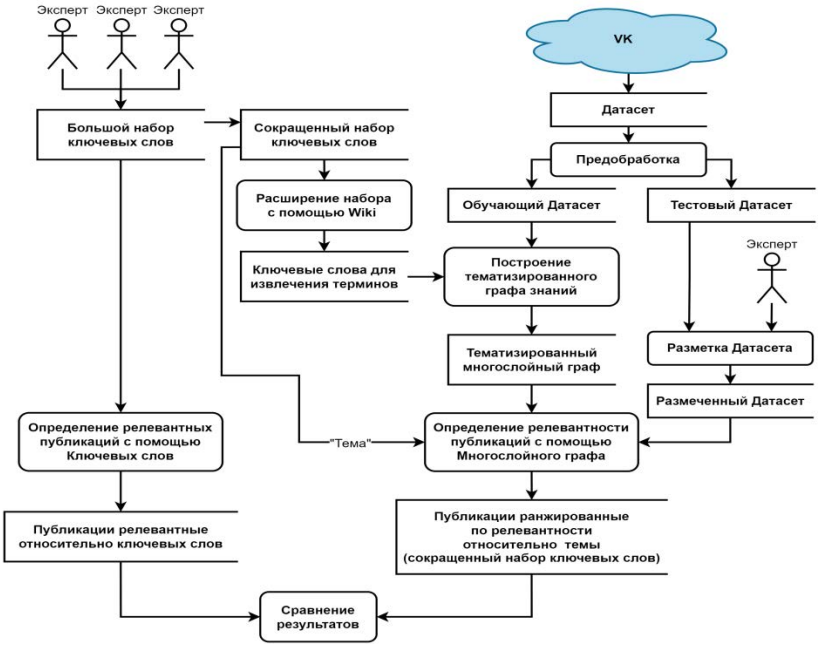


Рис. 6. Схема эксперимента по оценке эффективности предлагаемой технологии определения релевантности с помощью многослойного ГЗ

6.1. Формирование многослойного ГЗ. На первом этапе из упомянутого ранее набора ключевых слов (133 термина) было экспертно отобрано 11 терминов, наиболее полно отражающих общую тему запроса.

Для проведения эксперимента использовался датасет публикаций из 25 групп социальной сети “ВКонтакте” за период примерно в один год (начало 2024 – начало 2025), содержащим более 73 тыс. публикаций и 107 тыс. комментариев к ним. Данный датасет был очищен от повторяющихся и коротких публикаций, содержащих менее 76 символов. Как показали экспертные наблюдения, такие публикации крайне редко были содержательными или имели какой-то смысл для проводимого эксперимента. Данное число символов было выбрано исходя из средней длины слов для разговорной речи и средней длины предложения в русском языке. В результате осталось 35469 публикаций, которые использовались для построения ГЗ. Для оценки технологии использовался размеченный на основе исходного набора ключевых слов датасет из 617 публикаций, 100 из которых считались релевантными и 517 нерелевантными.

Для формирования ГЗ, обучающий датасет был преобразован в набор структур коммуникаций, где каждая структура является представлением публикации и её откликов (комментариев). Вместе с этим проводилось выделение контекстно важных терминов в этих структурах с помощью авторского комбинированного метода извлечения терминов. Термины для словаря, являющегося частью метода, были получены на основе малого набора ключевых слов, который был расширен экспертом с помощью инструмента для поиска связанных терминов, полагающегося на информацию из Википедии [20], и в итоге содержал 26 терминов. В результате работы метода в документах были выделены термины, входящие в заранее заданный словарь, именованные сущности, осмысленные биграммы и триграммы, а также контекстно важные униграммы-существительные. Полученные структуры дополнительно были очищены от редко встречающихся терминов. В результате был получен набор из 32883 структур, содержащих хотя бы два уникальных термина.

На основе полученных структур был обучен ряд тематических моделей, количество тем которых варьировалось от 5 до 20, однако в дальнейшем было решено использовать модель с 11 темами исходя из её показателя согласованности тем. Далее на основе полученных структур и выбранной тематической модели был произведен подсчет статистики совместного употребления терминов с учетом тематического

распределения структур (формула (1)). В результате были получены 11 матриц (по количеству тем выбранной тематической модели) значений совместного употребления терминов. Каждая такая матрица отражает некоторую меру совместного употребления терминов в конкретной теме. Полученные значения дополнительно были отфильтрованы от незначительных в соответствии с формулой (2). Таким образом, был получен ГЗ, состоящий из 11 слоев, вершинами которого стали уникальные термины, имеющие хотя бы одну связь в любом из тематических слоев, а значения совместного употребления, соответственно, определили наличие и силу связи между терминами в каждом слое графа.

6.2. Определение релевантности публикаций. Ранее полученный малый набор из 11 ключевых слов использовался как запрос пользователя для получения оценок релевантности на основе ГЗ. При этом, используя ранее полученную тематическую модель, были получены векторы тематического распределения терминов из этого набора, что позволяет понять, какие темы (они же слои в графе), являются наиболее релевантными запросу. Соответствующее тематическое распределение запроса, полученное как сумма векторов его терминов, приведено в таблице 2 (для удобства значения были нормализованы). Соответственно, релевантными слоями графа считались слои 1, 5 и 9.

Таблица 2. Тематическое распределение запроса, заданного малым набором ключевых слов

1	2	3	4	5	6	7	8	9	10	11
0,9751	0,0003	0,0958	0,0016	**0,7463**	0,0001	0,0003	0,4756	**0,9999**	0,3014	0,0505

Релевантность публикаций в каждом слое графа определялась в соответствии с формулой (3). Для сравнения с методами на основе ключевых слов, которые дают оценку вида да/нет, был выбран порог значения релевантности 0,9, чтобы представить результаты в аналогичном виде. Стоит отметить, что выбор наиболее подходящего порогового значения осложняется из-за несбалансированности имеющегося тестового набора. Соответствующее значение было выбрано исходя из баланса значений метрик точности и полноты с учетом истинно положительных и истинно отрицательных оценок (Рис. 7). Таким образом, публикация считалась релевантной, если она получала минимальное значение метрики в одном из релевантных слоев менее или равное 0,9,

и нерелевантной, если полученное минимальное значение метрики превышало порог или было в любом другом слое графа.

В рамках данного эксперимента проводилось сравнение предлагаемой технологии с технологиями на основе точного (ExactMatch) и частичного (StemmerMatch) совпадений. При определении релевантности с помощью этих технологий публикация считалась релевантной, если содержала хотя бы одно точное совпадение с ключевым словом или вхождение основы слова соответственно. В случае словосочетаний такая проверка проводилась отдельно для каждого слова. В таблице 3 и на рисунках 8-9 представлены результаты эксперимента.

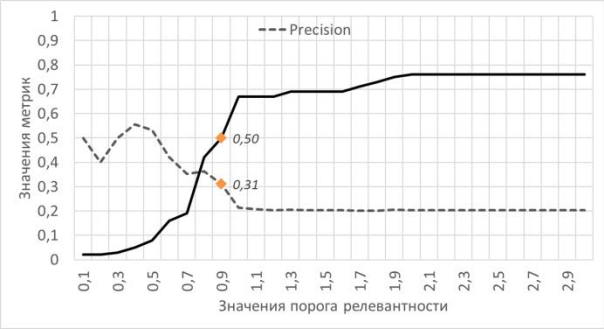


Рис. 7. Зависимость значений метрик Precision и Recall от порога релевантности

Таблица 3. Результаты проверки работы метода на тестовом наборе

	Малый набор ключевых слов (11)			**Большой набор ключевых слов (133)**	
	Многослойный ГЗ	**Exact-Match**	**Stemmer-Match**	**Exact-Match**	**Stemmer-Match**
TP	50	1	42	72	97
TN	406	517	471	402	339
FP	111	0	46	115	178
FN	50	99	58	28	3
Precision	**0,31**	1,00	0,48	**0,39**	**0,35**
Recall	0,50	0,01	0,42	0,72	0,97

Предлагаемая технология практически не уступает по показателям точности и полноты технологиям на основе ключевых слов, использующих полный набор таковых, и превосходит эти же технологии, работающие на неполном наборе ключевых слов. Если смотреть с позиции необходимости идентификации релевантных

публикаций с наименьшим числом ложноположительных результатов, то в целом подтверждается гипотеза о том, что предлагаемая технология может работать не хуже методов на основе ключевых слов, использующих значительно больший набор таковых. Иными словами, с помощью полученного ГЗ нам удалось добиться эффекта расширения набора ключевых слов экспертом с учетом социальных стереотипов. Дополнительно стоит заметить, что предлагаемая технология дает не просто оценку вида, в формате да/нет, а некоторую количественную оценку релевантности, что, в свою очередь, позволяет ранжировать полученные результаты.

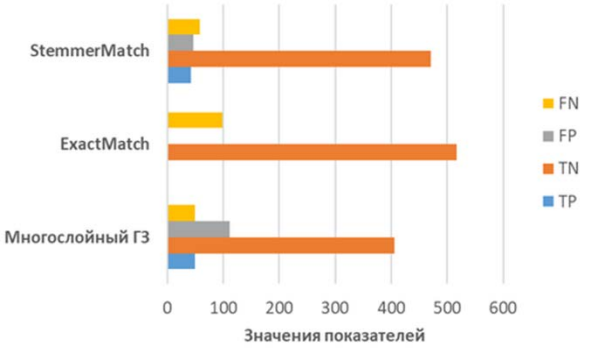


Рис. 8. Результаты эксперимента на малом наборе ключевых слов

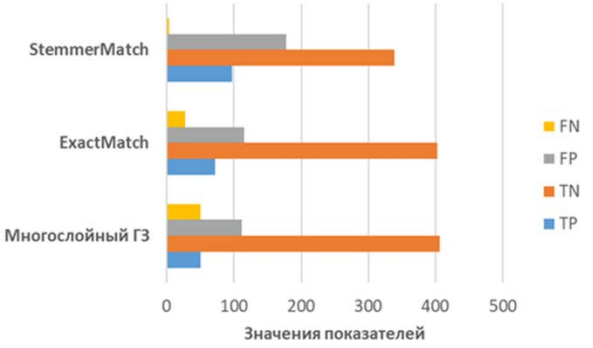


Рис. 9. Результаты сравнения многослойного ГЗ с технологиями, использующими большой набор ключевых слов

На рисунке 10 приведен анализ нескольких примеров публикаций с ложноположительными (публикации №1 и №2) и ложноотрицательными оценками (публикации №3 и №4). Сверху приведен текст публикаций и их оценки экспертами и предлагаемой технологией. Снизу приведены полученные метрики расстояний по слоям, красным выделены минимальные значения в релевантных слоях (1, 5 и 9), которые оказались ошибочными по тем или иным причинам. Для публикаций №1 и №2 наименьшая оценка расстояния находится в одной из релевантных тем, и её значение ниже заданного порога, но сами публикации не являются релевантными. При этом публикация №1 содержит крайне мало текстовой информации в целом. Для публикаций №3 и №4 наименьшая оценка релевантности находится все еще в релевантных слоях, но превышает заданный порог. Значительно превышает (более чем в 2 раза) для №3 и незначительно превышает для №4. Данные примеры демонстрируют выявленные слабые места технологии, связанные с её реализацией или выбранным пороговым значением. Таким образом, необходимо дальнейшее исследование и модернизация выбранного подхода к определению релевантности с помощью метрики расстояния на графе, чтобы избежать или минимизировать подобные ошибки.

	Текст публикации	Метка эксперта	Многослойный ГЗ
№1	Новогодние свободные катания в Ледовом дворце Апатит Арена Сегодня в 18 00 и 20 00.	0	1
№2	Кировчанин Артем Андреев стал бронзовым призером чемпионата Мурманской области по лыжным гонкам в дисциплине классический спринт Андрей Малетин и Владелинка Архипова впервые приняли участие первый в Чемпионате, получили опыт и мотивацию работать дальше.	0	1
№3	Минтранс опроверг слухи о повышении стоимости проезда до 50 рублей. Рост цен на энергоресурсы и другие услуги делает возможным увеличение тарифов, однако конкретных решений пока нет.	1	0
№4	Изменились цены на посещение природных парков в Мурманской области С 1 января 2025 года вступают в силу новые цены на посещение особо охраняемых природных территорий, таких как природный парк Териберка, Полуострова Рыбачий и Средний и Сейдъяввръ. Териберка Граждане РФ 440 рублей, иностранные граждане 880 рублей. Рыбачий и Средний 370 рублей. Сейдъяввръ 740 рублей. Для жителей Мурманской области, школьников, дошкольников, многодетных родителей и других льготных категорий вход по-прежнему бесплатный.	1	0

	Тематический слой графа										
	1	2	3	4	5	6	7	8	9	10	11
№1	**1,48**	1,96	1,47	1,87	**2,14**	1,95	1,79	1,84	**0,78**	-	1,93
№2	**0,76**	1,88	1,75	3,29	**1,69**	1,96	1,85	1,76	**0,96**	-	0,99
№3	**1,83**	2,27	2,71	2,51	**2,19**	1,99	2,27	1,88	-	-	1,98
№4	**1,73**	1,52	2,42	2,96	**1,80**	2,37	1,89	2,08	**0,96**	-	1,49

Рис. 10. Пример анализа публикаций предлагаемой технологией

6.3. Сравнение предлагаемой технологии определения релевантности с технологиями на основе векторных представлений. Определение релевантности текстовых документов с использованием эмбеддингов, как правило, производится на основе вычисления меры косинусного сходства (cosine similarity) между эмбеддингом запроса и эмбеддингами документов, которая определяет косинус угла между ними (что показывает, насколько похожи векторы независимо от их величины).

Для сравнения с предлагаемой технологией использовались три модели контекстных векторных представлений: bge-m3 [21], embeddinggemma [22], nomic-embed-text [23]. Выбор обусловлен богатой практикой использования данных моделей для решения задач ИП и их доступностью. Модели использовались локально без дообучения через платформу Ollama.

Сравнение проводилось на датасете, описанном в разделе 6.1. В начале для запроса и всех документов в датасете была получена оценка их косинусного сходства. При этом для каждой модели, используемой в эксперименте, характерен свой собственный разброс значений метрики на тестовых данных, что продемонстрировано на рисунке 11. Затем для каждой модели был, подобрал подходящий порог релевантности, аналогично предшествующему эксперименту. Зависимости значений метрик точности и полноты от применяемого порога релевантности продемонстрированы на рисунке 12. Таким образом, были выбраны следующие значения порогов: 0,35 (bge-m3), 0,2 (embeddinggemma) и 0,75 (nomic-embed-text). Выбор основан на балансе между количеством истинно положительных и истинно отрицательных оценок. Соответствующие значения метрик приведены в таблице 4.

Модели bge-m3 и nomic-embed-text продемонстрировали результат хуже предлагаемой технологии, поскольку при схожих значениях истинно положительных оценок количество истинно отрицательных оценок у этих моделей значительно ниже. Модель embeddinggemma продемонстрировала схожие, но все еще уступающие результаты. Потенциально модели эмбеддингов могут быть применены для решения данной задачи, но остаются открытыми вопросы выбора подходящей модели, её дообучения (что требует наличия соответствующих ресурсов – как языковых, так и технических) и выбора пороговых значений для выбранной меры оценки релевантности. Также стоит заметить, что дообучение, вероятно, будет производиться не только в начале применения модели, но и в процессе её использования для включения в нее новых данных,

отражающих изменение и развитие социального дискурса в наблюдаемых источниках. Кроме того, как и с любыми методами определения релевантности, основанными на машинном обучении, становится затруднительным объяснение полученных результатов по сравнению с методами, полагающимися на термины и системы знаний.



Рис. 11. Распределение показателей косинусного сходства для тестового датасета по трем моделям

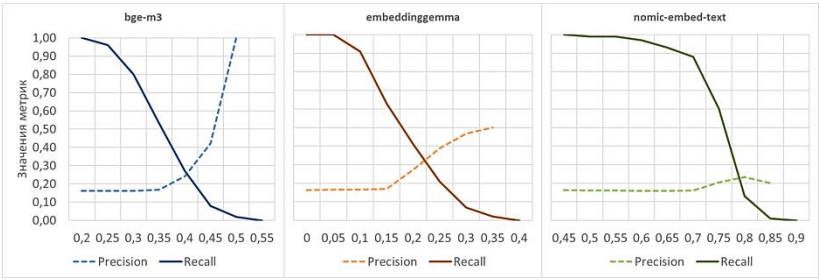


Рис. 12. Зависимость значений метрик точности и полноты от порога релевантности для моделей эмбеддингов

Таблица 4. Результаты сравнения предлагаемой технологии с технологиями на основе векторных представлений

	Многослойный ГЗ	**bge-m3**	**embeddinggemma**	**nomic-embed-text**
TP	50	53	41	60
TN	406	254	408	283
FP	111	263	109	234
FN	50	47	59	40
Precision	**0,31**	0,17	0,27	0,20
Recall	0,50	0,53	0,41	0,60

7. Заключение. В статье рассмотрена проблема повышения эффективности ИП в данных социальных медиа с точки зрения соответствия результатов поиска ожиданиям пользователя (пертинентности результата). Ключевыми проблемами ИП в социальных медиа является динамичность и разнородность знаний, представленных содержанием пользовательской коммуникации с одной стороны и метазадачей поиска, определяющей оценку качества результата – с другой. В качестве решения этих проблем, которые не способны разрешить рассмотренные в разделе 2 подходы (таблица 5), в работе предлагается МДСЗ, интегрирующая в своем составе динамичные разнородные знания и позволяющая приблизить поисковую выдачу к пертинентному результату. В частности, в статье рассмотрены следующие научные и практические результаты:

1. Представлена формализация задачи ИП в социальных медиа, включающая определение концептуальной модели предметной области и коллекции документов, и отражающая свойства мульдоменности и динамичности семантики данных.

2. Разработана модель МДСЗ, одновременно интегрирующая проблемно-ориентированную семантику с помощью онтологии, динамическую доменно-специфическую семантику с помощью многослойного ГЗ и доменно-независимую семантику с помощью лексической БД и большой языковой модели.

3. Предложена технология определения релевантности публикаций с использованием метрики расстояния между терминами пользовательского запроса и терминами публикации в многослойном ГЗ, слои которого соответствуют тематическому распределению коллекции документов.

4. Проведен эксперимент по определению релевантных публикаций с помощью предложенной технологии и её сравнения с технологиями, основанными на полном и частичном поиске вхождения ключевого слова запроса в публикацию и на эмбеддинговых моделях. Сравнительный анализ результатов эксперимента подтверждает эффективность применения разработанной технологии.

В продолжении данного исследования планируется сформировать лексическую БД, расширить функционал использования онтологии в рамках МДСЗ, а также рассмотреть возможность использования не просто меры совместного употребления при построении ГЗ, а точечной совместной информации (pointwise mutual information, PMI) как более точного способа отражения семантических зависимостей ментальной картины пользователей [24]. Кроме того,

в рамках задачи определения релевантности планируется изучить возможность применения алгоритмов распространения и векторных представлений графов [25], таких как node2vec [26].

Таблица 5. Сравнение подходов к определению релевантных документов в рамках ИП

Подход	**Возможность работы с мультидомен-ными данными**	**Возможность работы в условиях динамичности данных**	**Возможность использования дополнительной информации из социальных медиа для повышения точности**
Логический поиск	+	-	-
TF-IDF	+	-	-
Расширение запроса	± (зависит от метода)	-	±
Латентно-семантический анализ	+	-	-
Тематическое моделирование	+	-	-
Глубокое обучение (модели эмбеддингов)	+	-	+
МДСЗ	**+**	**+**	**+**

Литература

1. Hambarde K.A., Proenca H. Information Retrieval: Recent Advances and Beyond // IEEE Access. 2023. vol. 11. pp. 76581–76604. DOI: 10.1109/ACCESS.2023.3295776.1–26.
2. Carpineto C., Romano G. A Survey of Automatic Query Expansion in Information Retrieval // ACM Computing Surveys. 2012. vol. 44. no. 1. pp. 1–50. DOI: 10.1145/2071389.2071390.
3. Azad H.K., Deepak A. Query Expansion Techniques for Information Retrieval: a survey // Information Processing and Management. 2019. vol. 56. no. 5. pp. 1698–1735. DOI: 10.1016/j.ipm.2019.05.009.
4. Russell-Rose T., Gooch P., Kruschwitz U. Interactive query expansion for professional search applications. arXiv:2106.13528. 2021.
5. Zou X. A Survey on Application of Knowledge Graph // Journal of Physics Conference Series. 2020. vol. 1487. no. 1. pp. 1–11. DOI: 10.1088/1742-6596/1487/1/012016.
6. Reinanda R., Meij E., de Rijke M. Knowledge Graphs: An Information Retrieval Perspective. 2020. pp. 1–153. DOI: 10.1561/9781680837292.
7. Janowicz K., et al. Know, Know Where, Knowwheregraph: A Densely Connected, Cross-Domain Knowledge Graph and Geo-Enrichment Service Stack for Applications in Environmental Intelligence // AI Magazine. 2022. vol. 43. no. 1. pp. 30–39. DOI: 10.1609/aimag.v43i1.19120.
8. Bouadjenek M.R., Hacid H., Bouzeghoub M. Social networks and information retrieval, how are they converging? A survey, a taxonomy and an analysis of social

information retrieval approaches and platforms // Information Systems. 2016. vol. 56. pp. 1–18. DOI: 10.1016/j.is.2015.07.008.
9. Khalifi H., Dahir S., El Qadi A., et al Enhancing information retrieval performance by using social analysis // Social Network Analysis and Mining. 2020. vol. 10. no. 1. pp. 1–7. DOI: 10.1007/s13278-020-00635-w.
10. Hua Y., et al. Streamlining social media information retrieval for public health research with deep learning // J Am Med Inform Assoc. 2024. vol. 31. no. 7. pp. 1569–1577. DOI: 10.1093/jamia/ocae118.
11. Fellbaum C. WordNet: An Electronic Lexical Database / With a preface by George Miller. Cambridge, MA, USA: MIT Press, 1998. 422 p.
12. Baayen R.H., Piepenbrock R, Gulikers L. CELEX2 LDC96L14. Web Download. Philadelphia: Linguistic Data Consortium, 1995. DOI: 10.35111/gs6s-gm48.
13. Mahajan Y., Freestone M., Aakur S., Karmaker S. Revisiting Word Embeddings in the LLM Era: arXiv:2502.19607. 2025.
14. Park K., et al. The Geometry of Categorical and Hierarchical Concepts in Large Language Models: arXiv:2406.01506. 2025.
15. Nie Z., et al. When Text Embedding Meets Large Language Model: A Comprehensive Survey: arXiv:2412.09165. 2025.
16. Chroma. URL: https://github.com/chroma-core/chroma (дата обращения: 12.09.2025).
17. Wang J., et al. Milvus: A Purpose-Built Vector Data Management System // Proceedings of the 2021 International Conference on Management of Data. Virtual Event China: ACM. 2021. pp. 2614–2627. DOI: 10.1145/3448016.3457550.
18. Пимешков В.К., Никонорова М.Л., Шишаев М.Г. Комбинированный метод извлечения терминов для задачи мониторинга тематических обсуждений в социальных медиа // Информатика и автоматизация. 2024. Т. 23. №4. С. 1110–1138. DOI: 10.15622/ia.23.4.7.
19. Pimeshkov V., Nikonorova M., Shishaev M. Technology for Forming a Multilayer Knowledge Graph to Determine the Relevance of Documents // Digital and Information Technologies in Economics and Management / In: Gibadullin, A. (eds). Cham: Springer Nature Switzerland. 2025. vol. 1422. pp. 134–152. DOI: 10.1007/978-3-031-94273-0_11.
20. Шишаев М.Г., Пимешков В.К., Никонорова М.Л. Утилита формирования структур данных из документов Википедии // Роспатент: Свидетельство о государственной регистрации программы для ЭВМ №2024661261 от 16.05.2024.
21. Chen J., et al. BGE M3-Embedding: Multi-Lingual, Multi-Functionality, Multi-Granularity Text Embeddings Through Self-Knowledge Distillation: arXiv:2402.03216. 2024.
22. Vera H. S., et al. EmbeddingGemma: Powerful and Lightweight Text Representations: arXiv:2509.20354. 2025.
23. Nussbaum Z., et al. Nomic Embed: Training a Reproducible Long Context Text Embedder: arXiv:2402.01613. 2025.
24. Zhu C., Liang Y., Liang X., Zhong L., Xie F. Fairness modeling for topics with different scales in short texts // PeerJ Computer Science. 2025. vol. 11. pp. 1–20. DOI: 10.7717/peerj-cs.2936.
25. Choudhary S., Luthra T., Mittal A., Singh R. A Survey of Knowledge Graph Embedding and Their Applications: arXiv:2107.07842. 2021.
26. Grover A., Leskovec J. node2vec: Scalable Feature Learning for Networks: arXiv:1607.00653. 2016.

Пимешков Вадим Константинович — стажер-исследователь, лаборатория информационных технологий управления региональным развитием, Институт

информатики и математического моделирования им. В.А. Путилова — обособленное подразделение Федерального государственного бюджетного учреждения науки Федерального исследовательского центра «Кольский научный центр Российской академии наук» (ИИММ КНЦ РАН). Область научных интересов: обработка естественного языка, извлечение знаний из естественно-языковых текстов. Число научных публикаций — 11. v.pimeshkov@ksc.ru; ул. Ферсмана, 24А, 184209, Апатиты, Россия; р.т.: +7(815)557-9689.

Никонорова Марина Леонидовна — инженер-исследователь, лаборатория информационных технологий управления региональным развитием, Институт информатики и математического моделирования им. В.А. Путилова — обособленное подразделение Федерального государственного бюджетного учреждения науки Федерального исследовательского центра «Кольский научный центр Российской академии наук» (ИИММ КНЦ РАН). Область научных интересов: анализ естественного языка, извлечение знаний, онтологическое моделирование, машинное обучение. Число научных публикаций — 15. m.nikonorova@ksc.ru; ул. Ферсмана, 24А, 184209, Апатиты, Россия; р.т.: +7(815)557-9689.

Шишаев Максим Геннадьевич — д-р техн. наук, доцент, главный научный сотрудник, руководитель лаборатории, лаборатория информационных технологий управления региональным развитием, Институт информатики и математического моделирования им. В.А. Путилова — обособленное подразделение Федерального государственного бюджетного учреждения науки Федерального исследовательского центра «Кольский научный центр Российской академии наук» (ИИММ КНЦ РАН). Область научных интересов: информационные системы, региональное развитие, инженерия знаний, искусственный интеллект, машинное обучение, интеллектуальный анализ текстов. Число научных публикаций — 169. m.shishaev@ksc.ru; ул. Ферсмана, 24А, 184209, Апатиты, Россия; р.т.: +7(815)557-9248.

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 неопределенности и риска» (FMEZ-2025-0053). Авторы выражают благодарность Вишнякову Ивану Геннадьевичу, аспиранту и системному администратору ИИММ КНЦ РАН, за выполнение тематического моделирования.

DOI 10.15622/ia.25.3.6

V. PIMESHKOV, M. NIKONOROVA, M. SHISHAEV

INFORMATION RETRIEVAL IN SOCIAL MEDIA BASED ON A MULTI-DOMAIN DYNAMIC KNOWLEDGE SYSTEM

Pimeshkov V., Nikonorova M., Shishaev M. **Information Retrieval in Social Media Based on a Multi-Domain Dynamic Knowledge System.**

Abstract. The task of information retrieval is to find information that best satisfies the user's information needs. In the context of social media, information retrieval is complicated by the high dynamism of content, thematic heterogeneity, and the diversity of users' mental models. This paper proposes an approach to solving the problem of information retrieval under such conditions by constructing a multi-domain dynamic knowledge system. Its novelty lies in the combination of three levels of semantics: problem-oriented, represented by the ontology of the metatask (describing the search objectives); domain-specific, implemented through a dynamic multi-layer knowledge graph built on the basis of user content of social media; and domain-independent, based on a lexical database and a large language model. The knowledge graph allows us to reflect various contexts of concept usage corresponding to thematic clusters in the document collection. Such integration allows us to take into account the evolution of concepts, discourse features, and mental stereotypes of communication participants. To evaluate the effectiveness of the proposed system, an experiment was conducted using a dataset of publications from the VKontakte social network for problem-oriented monitoring of publications, where the selection of relevant publications from non-thematic sources is required. To solve this problem, a technology based on the use of the distance metric between query terms and publication terms in a multi-layer knowledge graph was proposed. The results of the experiment using this technology confirm the effectiveness of the proposed model for information retrieval tasks compared to standard keyword search and embedding models. In continuation of this study, it is planned to create a lexical database and also to consider the possibility of expanding the model by using a measure of pointwise mutual information and graph embedding methods.

Keywords: information retrieval, social media, knowledge systems, multi-layer knowledge graph, relevance.

References

1. Hambarde K.A., Proenca H. Information Retrieval: Recent Advances and Beyond. IEEE Access. 2023. vol. 11. pp. 76581–76604. DOI: 10.1109/ACCESS.2023.3295776.
2. Carpineto C., Romano G. A Survey of Automatic Query Expansion in Information Retrieval. ACM Computing Surveys. 2012. vol. 44. no. 1. pp. 1–50. DOI: 10.1145/2071389.2071390.
3. Azad H.K., Deepak A. Query Expansion Techniques for Information Retrieval: a survey. Information Processing and Management. 2019. vol. 56. no. 5. pp. 1698–1735. DOI: 10.1016/j.ipm.2019.05.009.
4. Russell-Rose T., Gooch P., Kruschwitz U. Interactive query expansion for professional search applications. arXiv:2106.13528. 2021.
5. Zou X. A Survey on Application of Knowledge Graph. Journal of Physics Conference Series. 2020. vol. 1487. no. 1. pp. 1–11. DOI: 10.1088/1742-6596/1487/1/012016.
6. Reinanda R., Meij E., de Rijke M. Knowledge Graphs: An Information Retrieval Perspective. 2020. pp. 1–153. DOI: 10.1561/9781680837292.
7. Janowicz K., et al. Know, Know Where, Knowwheregraph: A Densely Connected, Cross-Domain Knowledge Graph and Geo-Enrichment Service Stack for Applications

in Environmental Intelligence. AI Magazine. 2022. vol. 43. no. 1. pp. 30–39. DOI: 10.1609/aimag.v43i1.19120.
8. Bouadjenek M.R., Hacid H., Bouzeghoub M. Social networks and information retrieval, how are they converging? A survey, a taxonomy and an analysis of social information retrieval approaches and platforms. Information Systems. 2016. vol. 56. pp. 1–18. DOI: 10.1016/j.is.2015.07.008.
9. Khalifi H., Dahir S., El Qadi A., et al Enhancing information retrieval performance by using social analysis. Social Network Analysis and Mining. 2020. vol. 10. no. 1. pp. 1–7. DOI: 10.1007/s13278-020-00635-w.
10. Hua Y., et al. Streamlining social media information retrieval for public health research with deep learning. J Am Med Inform Assoc. 2024. vol. 31. no. 7. pp. 1569–1577. DOI: 10.1093/jamia/ocae118.
11. Fellbaum C. WordNet: An Electronic Lexical Database. With a preface by George Miller. Cambridge, MA, USA: MIT Press, 1998. 422 p.
12. Baayen R.H., Piepenbrock R, Gulikers L. CELEX2 LDC96L14. Web Download. Philadelphia: Linguistic Data Consortium, 1995. DOI: 10.35111/gs6s-gm48.
13. Mahajan Y., Freestone M., Aakur S., Karmaker S. Revisiting Word Embeddings in the LLM Era: arXiv:2502.19607. 2025.
14. Park K., et al. The Geometry of Categorical and Hierarchical Concepts in Large Language Models: arXiv:2406.01506. 2025.
15. Nie Z., et al. When Text Embedding Meets Large Language Model: A Comprehensive Survey: arXiv:2412.09165. 2025.
16. Chroma. Available at: https://github.com/chroma-core/chroma (accessed 12.09.2025).
17. Wang J., et al. Milvus: A Purpose-Built Vector Data Management System. Proceedings of the 2021 International Conference on Management of Data. Virtual Event China: ACM. 2021. pp. 2614–2627. DOI: 10.1145/3448016.3457550.
18. Pimeshkov V., Nikonorova M., Shishaev M. [A Combined Term Extraction Method for the Problem of Monitoring Thematic Discussions in Social Media]. Informatika i Avtomatizatsiya – Informatics and Automation. 2024. vol. 23. no. 4. pp. 1110–1138. DOI: 10.15622/ia.23.4.7. (In Russ.)
19. Pimeshkov V., Nikonorova M., Shishaev M. Technology for Forming a Multilayer Knowledge Graph to Determine the Relevance of Documents. Digital and Information Technologies in Economics and Management. In: Gibadullin, A. (eds). Cham: Springer Nature Switzerland. 2025. vol. 1422. pp. 134–152. DOI: 10.1007/978-3-031-94273-0_11.
20. Shishaev M.G., Pimeshkov V.K., Nikonorova M.L. [Utility for generating data structures from Wikipedia documents]. Rospatent: Svidetel'stvo o gosudarstvennoy registratsii programmy dlya EVM №2024661261 ot 16.05.2024 [Rospatent: Certificate of state registration of a computer program no. 2024661261 dated 05/16/2024]. (In Russ.).
21. Chen J., et al. BGE M3-Embedding: Multi-Lingual, Multi-Functionality, Multi-Granularity Text Embeddings Through Self-Knowledge Distillation: arXiv:2402.03216. 2024.
22. Vera H. S., et al. EmbeddingGemma: Powerful and Lightweight Text Representations: arXiv:2509.20354. 2025.
23. Nussbaum Z., et al. Nomic Embed: Training a Reproducible Long Context Text Embedder: arXiv:2402.01613. 2025.
24. Zhu C., Liang Y., Liang X., Zhong L., Xie F. Fairness modeling for topics with different scales in short texts. PeerJ Computer Science. 2025. vol. 11. pp. 1–20. DOI: 10.7717/peerj-cs.2936.
25. Choudhary S., Luthra T., Mittal A., Singh R. A Survey of Knowledge Graph Embedding and Their Applications: arXiv:2107.07842. 2021.

26. Grover A., Leskovec J. node2vec: Scalable Feature Learning for Networks: arXiv:1607.00653. 2016.

Pimeshkov Vadim — Research intern, Laboratory of Information Technologies for Regional Development Management, V.A. Putilov Institute for Informatics and Mathematical Modeling, a separate subdivision of the Federal Research Center "Kola Scientific Center of the Russian Academy of Sciences" (IIMM KSC RAS). Research interests: natural language processing, knowledge extraction from natural language texts. The number of publications — 11. v.pimeshkov@ksc.ru; 24A, Fersman St., 184209, Apatites, Russia; office phone: +7(815)557-9689.

Nikonorova Marina — Research engineer, Laboratory of Information Technologies for Regional Development Management, V.A. Putilov Institute for Informatics and Mathematical Modeling, a separate subdivision of the Federal Research Center "Kola Scientific Center of the Russian Academy of Sciences" (IIMM KSC RAS). Research interests: natural language analysis, knowledge extraction, ontological modeling, machine learning. The number of publications — 15. m.nikonorova@ksc.ru; 24A, Fersman St., 184209, Apatites, Russia; office phone: +7(815)557-9689.

Shishaev Maksim — Ph.D., Dr. Sci., Associate professor, Chief scientific officer, Head of the Laboratory, Laboratory of Information Technologies for Regional Development Management, V.A. Putilov Institute for Informatics and Mathematical Modeling, a separate subdivision of the Federal Research Center "Kola Scientific Center of the Russian Academy of Sciences" (IIMM KSC RAS). Research interests: information systems, regional development, knowledge engineering, artificial intelligence, machine learning, text mining. The number of publications — 169. m.shishaev@ksc.ru; 24A, Fersman St., 184209, Apatites, Russia; office phone: +7(815)557-9248.

Acknowledgements. The study was carried out within the framework of the state assignment of IIMM KSC RAS from the Ministry of Science and Higher Education of the Russian Federation, research topic "Methods and technologies for creating intelligent information systems to support the development of complex dynamic systems with regional specifics in conditions of uncertainty and risk" (FMEZ-2025-0053). The authors express their gratitude to Ivan Vishnyakov, postgraduate student and system administrator of IIMM KSC RAS, for performing the topic modeling.

УДК 004.89; 004.021; 004.272 **DOI** 10.15622/ia.25.3.7

В.В. Курейчик, В.И. Данильченко, Е.В. Данильченко

АДАПТИВНАЯ ПАРАМЕТРИЗАЦИЯ МЕТАЭВРИСТИЧЕСКОГО АЛГОРИТМА ПРИ РАЗМЕЩЕНИИ КОМПОНЕНТОВ СБИС

Курейчик В.В., Данильченко В.И., Данильченко Е.В. **Адаптивная параметризация метаэвристического алгоритма при размещении компонентов СБИС.**

Аннотация. В работе рассматривается задача параметризации метаэвристического алгоритма размещения компонентов сверхбольших интегральных схем (СБИС) в условиях усложняющихся топологий и ограничений современных технологических норм. Актуальность исследования обусловлена ростом размерности проектируемых схем, увеличением плотности межсоединений и усложнением структуры пространства проектных решений, что ограничивает эффективность применяемых алгоритмов со статическими параметрами. Постановка задачи включает формирование комплекса проектных метрик, таких как длина соединений, плотность, площадь размещения, индекс пересечений и другие. Фиксированная параметризация метаэвристических алгоритмов не в полной мере учитывает динамику пространства поиска, что ограничивает качество принимаемых проектных решений. На основе анализа комплекса проектных метрик предложена динамическая модель, отражающая текущее состояние поискового процесса, включая уровень структурного разнообразия популяции и признаки стагнации. С учетом предложенной модели разработана структура адаптивного метаэвристического алгоритма с замкнутым контуром параметрического управления, где выполняется корректировка параметров для согласования интенсивности глобального и локального поиска. Научная новизна заключается в оригинальной конфигурации набора метрик (скорость улучшения целевой функции δt, разнообразие популяции Dt, локальная неоднородность ландшафта поиска Ht, индикатор стагнации St) и способе их интеграции в контур параметрического управления. В отличие от ресурсоёмких нейросетевых архитектур, требующих значительных вычислительных затрат и обучающих выборок, предложенная аналитическая модель обеспечивает интерпретируемость процесса адаптации и высокую вычислительную эффективность. Для подтверждения эффективности подхода построена программная подсистема, реализующая процесс моделирования и вычислительного эксперимента. Результаты экспериментов на тестовых топологиях различной размерности и плотности связей показали, что по сравнению с алгоритмами с фиксированной многоэтапной настройкой применение разработанного алгоритма позволило повысить интегральный показатель качества размещения на 8-15 % для структур с низкой плотностью связей и на 12-25 % для структур с высокой плотностью связей, уменьшить дисперсию результатов при множественных итерациях в 1,3-2,2 раза, сократить число циклов стагнации и снизить чувствительность к начальному состоянию популяции. Полученные результаты подтверждают повышение качества и эффективности проектных решений, устойчивость разработанного алгоритма размещения и свидетельствуют о практической применимости предложенного подхода. Статья будет полезна специалистам, занимающимся автоматизацией проектирования и интеллектуальным анализом при разработке.

Ключевые слова: метаэвристические алгоритмы, адаптивная параметризация, динамическая модель, размещение компонентов СБИС, комбинаторная оптимизация, метрики поиска, топологический синтез, многомерное дискретное пространство, интеллектуальные системы проектирования.

1. Введение. Современное проектирование сверхбольших интегральных схем (СБИС) характеризуется непрерывным ростом функциональной сложности, плотности компоновки и числа межсоединений, а также ужесточением технологических, временных и энергетических ограничений. Это приводит к существенному повышению сложности проектных процедур, требуя перехода от детерминированных методов к интеллектуальным адаптивным алгоритмам, способным исследовать высокомерное и нелинейное пространство решений [1 – 3]. При этом компонентом функциональной модели СБИС является элемент или блоком элементов, описываемых уникальным именем и функциональным типом, а сама схема задаётся списком таких компонентов и списком соединений между ними [2, 4 – 6]. Задача размещения этих компонентов относится к классу NP-трудных комбинаторных задач, а пространство возможных конфигураций растёт экспоненциально с увеличением числа компонентов СБИС [2, 4]. Комбинаторная сложность задачи размещения компонентов СБИС обусловлена тем, что каждый из n компонентов схемы может быть размещен в любом из m доступных узлов координатной сетки на плоскости кристалла, где $m \geq n$. В предположении об отсутствии симметрии и ограничений, процесс размещения эквивалентен выбору n различных позиций из m доступных и последующему присвоению их n компонентам, поэтому общее число вариантов решения определяется числом размещений и вычисляется по следующей формуле $A_m^n = \frac{m!}{(m-n)!}$. В случае, когда число позиций строго равно числу компонентов $(m = n)$, это выражение сводится к числу перестановок $A_m^n = n!$. При $n = 100$ значение $n!$ превышает 10^{158}. Это исключает использование алгоритмов полного перебора и снижает эффективность классических детерминированных алгоритмов. При многокритериальной постановке задачи традиционные алгоритмы характеризуются ограниченной устойчивостью и высокой чувствительность к выбору управляющих параметров. Также использование фиксированной параметризации в метаэвристических алгоритмах приводит к преждевременной сходимости и снижению полноты исследования пространства поиска. В результате получаемые наборы решений содержат локальные перегрузки, топологические конфликты и увеличенную длину межсоединений, что снижает качество получаемых проектных решений. Также топологическая модель размещения осложняется необходимостью учета и соблюдения технологических норм, минимальных расстояний между компонентами, ограничений плотности маршрутов и требований к распределению задержек сигналов. Включение этих ограничений в целевую функцию

приводит к появлению большого числа локальных минимумов и разрывов в поверхности поиска, что затрудняет применение градиентных методов [1 – 3, 7]. Кроме того, структура связей между компонентами определяет неоднородность пространства решений, а именно области с высокой плотностью соединений требуют точной настройки параметров поиска, тогда как слабосвязанные сегменты допускают выбор из широкого диапазон значений. С учетом этих факторов, требуется применения методов, обладающих высокой степенью гибкости, способностью работать в условиях разреженных и нелинейных пространствах и адаптироваться к изменяющимся характеристикам задачи.

Одними из них являются метаэвристические методы, широко используемые в задачах размещения благодаря способности эффективно исследовать сложные дискретные пространства. Их производительность зависит от настройки параметров, включая вероятности применения операторов мутации, кроссинговера и других стохастических операторов, а также глубину локального поиска [1 – 4]. При этом недостатком статической параметризации является отсутствие обратной связи между текущим состоянием поискового процесса, например уровнем разнообразия популяции или признаками стагнации и значениями управляющих параметров. Это приводит к снижению эффективности работы алгоритма, что проявляется в его преждевременной сходимости или в замедлении процесса поиска. Кроме того, шум, технологические ограничения и конфликты размещения исключают использование универсального набора параметров, обеспечивающего стабильную эффективность. В связи с этим возникает необходимость в адаптивной настройке параметров поиска.

Настоящая статья продолжает цикл исследований авторов, посвящённых разработке метаэвристических алгоритмов размещения компонентов СБИС [4, 5]. Отметим, что отдельные метрики, используемые в предлагаемой динамической модели, применяются и в иных современных подходах к параметрической оптимизации, в частности, в методах глубокого обучения с подкреплением [6]. Принципиальное отличие и научная новизна настоящей работы заключаются не в самом факте использования перечисленных показателей, а в оригинальной конфигурации набора метрик, включающего скорость улучшения целевой функции (δt), разнообразие популяции (Dt), локальную неоднородность ландшафта поиска (Ht) и индикатор стагнации (St), и способе их интеграции в замкнутый контур параметрического управления метаэвристическим алгоритмом.

В отличие от ресурсоёмких нейросетевых архитектур, требующих значительных вычислительных затрат и обучающих выборок, предложенная аналитическая модель обеспечивает интерпретируемость процесса адаптации и высокую вычислительную эффективность, что подтверждается улучшением интегрального показателя качества размещения на 8-25% в ходе вычислительного эксперимента.

Идея адаптивной параметризации заключается в динамической настройке параметров метаэвристических методов в зависимости от текущего этапа оптимизации и локального пространства решений. Это позволяет изменять вероятности применения различных операторов поиска на основе использования задаваемых метрик, таких как скорость улучшения целевой функции, признаков стагнации и разнообразия популяции решений. Данные метрики позволяют поддерживать баланс между диверсификации и интенсификации поиска и снижают риск попадания решений в локальные оптимумы.

Проведённый анализ показывает, что задача размещения компонентов СБИС отличается высокой комбинаторной сложностью и неоднородностью пространства поиска. В условиях технологических ограничений метаэвристические методы с фиксированными параметрами работают неустойчиво. Оптимальные значения параметров невозможно спрогнозировать заранее, так как они зависят от текущей динамики процесса поиска. Для решения данной проблемы требуется применения адаптивного управления, направленного на согласование процессов глобального поиска и локального уточнения решений в пространстве состояний, что обеспечит повышение устойчивости алгоритма и качество получаемых проектных решений. Поэтому разработка динамической модели адаптивной параметризации является актуальной и научно значимой задачей.

2. Анализ существующих решений. Методы размещения компонентов СБИС подразделяются на аналитические, иерархические, кластерные и эвристические, они отличаются принципами формализации задачи, механизмами декомпозиции пространства поиска и стратегиями редукции вычислительной сложности [1 – 4]. При этом аналитические методы [1, 3, 4] основаны на представлении схемы межсоединений в виде математической модели, в которой минимизируется суммарная длина соединений. Несмотря на высокую точность локального поиска, такие методы чувствительны к начальному приближению и плохо масштабируются при увеличении числа компонентов СБИС. Иерархических методы реализуют поэтапную декомпозицию задачи, последовательно уплотняя конфигурацию размещения, однако данные методы приводят

к накоплению ошибок на промежуточных уровнях, тем самым снижая качество получаемого решения.

Классические алгоритмы оптимизации [1, 3, 4], такие как градиентный спуск, методы наискорейшего спуска, релаксационные и сетевые алгоритмы, демонстрируют эффективность в задачах с гладкой или частично выпуклой поверхностью целевой функции. Однако для задачи размещения, характеризующейся множеством локальных минимумов и дискретной природой переменных, эти методы применимы лишь на отдельных этапах локальной оптимизации. При этом комбинаторные методы, включая жадные алгоритмы и алгоритмы поиска в глубину и ширину, обеспечивают приемлемое качество решения на малых наборах данных, что не позволяет обрабатывать все пространство решений экспоненциального размера, возникающее при увеличении количества компонентов и их связей.

Для расширения поисковых возможностей применяются методы имитации отжига, ориентированные на преодоление локальных минимумов за счёт стохастического управления температурой [1, 3, 6]. Однако при высокой размерности задачи они требуют значительных вычислительных затрат и существенно зависят от выбора температурного режима и кривой охлаждения. Таким образом, классические методы оптимизации демонстрируют ограниченную применимость, что обусловливает переход к эвристическим и метаэвристическим методам, способным эффективно исследовать сложные многомерные пространства поиска.

Эвристические и метаэвристические методы широко применяются в задачах размещения СБИС благодаря способности эффективно исследовать дискретные и многомерные пространства с множеством локальных минимумов [1 – 3]. Наиболее широко используемыми из них являются генетические алгоритмы (ГА) [1, 3, 6], реализующие эвристический поиск за счёт выполнения операторов селекции, кроссинговера, мутации и др. Они позволяют получать наборы квазиоптимальных решений при высокой комбинаторной сложности задачи, однако качество результатов существенно зависит от настройки параметров алгоритма, вероятности применения различных операторов поиска и стратегии формирования популяции.

Другим важным направлением метаэвристических методов является алгоритмы роевого интеллекта, включая метод роя частиц (PSO) [1 – 3, 7, 8]. Данный метод ориентирован на коллективное поведение агентов, использующих индивидуальный опыт и глобальную информацию для поиска оптимальных конфигураций. Он способен быстро адаптироваться к изменению структуры пространства поиска,

но подвержен преждевременной сходимости при недостаточной диверсификации частиц. В рамках данного направления также рассматриваются муравьиные алгоритмы (ACO) [1 – 3, 6], основанные на моделировании поведения колоний муравьёв с использованием механизма феромонных следов. Эти алгоритмы эффективно решают задачи, математическая модель которых представляется в виде графа, включая его матричные и списковые эквиваленты. Их преимущество заключается в способности формировать устойчивые маршруты поиска, однако их эффективность зависит от корректности настройки ключевых параметров (таких как начальный уровень феромона, скорость его испарения, коэффициент влияния эвристической информации и д.р.) и выбора адекватных эвристических правил.

Помимо рассмотренных подходов, в задачах размещения компонентов СБИС применяются различные гибридные и модифицированные методы эволюционного моделирования, сочетающие эволюционные принципы с локальным поиском и применением стохастических операторов [1, 3, 5 – 10]. Однако эффективность этих методов требует настройкой многих параметров и соблюдения баланса между глобальным и локальным поиском. При этом задаваемые статистических параметров не позволяет учитывать динамику всего процесса поиска. Данные проблемы приводят к снижению качества получаемых решений, росту вариативности результатов и увеличению вероятности наступления преждевременной сходимости алгоритма.

В литературе [1, 3, 5 – 10] представлены различные подходы к адаптивной и самоадаптивной параметризации эвристических и метаэвристических алгоритмов, направленные на повышение устойчивости поиска и снижение риска преждевременной сходимости. Наиболее распространённые методы основаны на регулировании вероятностей реализации поисковых операторов в зависимости от показателей качества наборов текущей популяции решения. Например, в генетических алгоритмах применяются схемы адаптивного изменения вероятностей применения операторов мутации, кроссинговера и д.р., где их параметры увеличиваются при снижении разнообразия популяции и уменьшаются при наличии стабильного улучшения целевой функции [1, 3, 11]. Внедрение таких механизмов позволяет поддерживать баланс между глобальным и локальным поиском, однако их эффективность ограничена выбором правил обновления, которые не всегда отражают реальное пространство решений. Также в методах роевого интеллекта [1, 7] адаптация реализуется через динамическое изменение инерционного

веса и коэффициентов когнитивной и социальной составляющих, что позволяет контролировать степень «инерции» частиц и их способность к исследованию новых областей. При всей структурной прозрачности, данные методы чувствительны к выбору параметров и часто требуют предварительного их подбора для решения конкретного класса задач. Муравьиные алгоритмы используют адаптивные правила обновления параметров в феромонной матрице, в том числе нелинейное испарение, усиление феромонных следов для выявления перспективных маршрутов, что позволяет снизить стагнацию процесса поиска. Однако влияние учета параметров феромона является сложным для прогнозирования, а их усиление приводит к диверсификации процесса поиска. Наряду с этим применяются самоадаптивные подходы [1, 9, 10], в которых параметры эволюционируют совместно с полученными решениями, что снижает необходимость ручной настройки и повышает гибкость процесса поиска. Вместе с тем такие методы требуют значительного числа итераций и приводят к росту вычислительных затрат. В целом адаптивная параметризация обладает потенциалом для повышения качества получаемых проектных решений, однако в рассматриваемой задаче её применение ограничено отсутствием универсальных метрик, адекватно описывающих состояние процесса поиска в многокритериальном и высокоразмерном пространстве. Поэтому для оценки применимости различных методов поиска к задаче размещения компонентов СБИС необходимо учитывать их вычислительную сложность, масштабируемость и устойчивость при работе в многомерных дискретных пространствах. В таблице 1 представлены асимптотические оценки сложности наиболее распространённых методов [1, 3 – 11], используемых при решении задач размещения, а также краткие характеристики их особенностей.

Таким образом известные подходы недостаточно эффективны в условиях высокой размерности, а эвристические и метаэвристические методы существенно зависят от настройки параметров и используемых метрик. Существующие механизмы адаптивной и самоадаптивной параметризации не в полной мере позволяют обеспечить устойчивую работу при изменении структуры целевой функции, плотности связей и технологических требований, поскольку опираются на ограниченный набор метрик и требуют эмпирической подстройки всех учитываемых параметров. В связи с этим необходима разработка динамической модели управления набором параметров, обеспечивающих адаптацию к текущему состоянию процесса и структуры пространства поиска.

Таблица 1. Вычислительная сложность алгоритмов

Алгоритм	**Вычислительная сложность**	**Примечание**
Градиентные методы	$O(n^2)$-$O(n^3)$	Эффективны для сглаженных задач, чувствительны к локальным минимумам
Метод имитации отжига (SA)	$O(k \cdot n)$	Требует подбора температурного режима; высокая трудоёмкость при больших n
Иерархические методы размещения	$O(n \cdot \log n)$	Эффективны для грубого размещения, снижают точность при детализации
Генетический алгоритм (GA)	$O(n \cdot m)$	Качество зависит от параметров; склонен к преждевременной сходимости
Метод роя частиц (PSO)	$O(n \cdot m)$	Быстрое сближение, высокая вероятность попадания в локальные минимумы
Муравьиный алгоритм (ACO)	$O(n \cdot m)$	Эффективен для графовых моделей; чувствителен к параметрам феромона
Гибридные эволюционные алгоритмы	$O(n \cdot m + n \cdot \log n)$	Объединяют глобальный и локальный поиск; улучшают стабильность результатов
Кластерные методы размещения	$O(n^2)$	Эффективны при высокой связности; хуже работают при неоднородных структурах
Аналитические методы	$O(n^3)$	Точны, но плохо масштабируются при увеличении числа компонентов

3. Постановка задачи. Задача размещения компонентов СБИС заключается в определении оптимального размещения компонентов на плоскости кристалла с учётом топологических, метрических и технологических ограничений [2, 4, 11]. Она представляет собой NP-трудную комбинаторную задачу многокритериальной оптимизации, направленную на определение оптимального расположения компонентов на плоскости или в решётке с учётом метрических критериев (длина проводников, число изгибов), топологических параметров (взаимное расположение элементов, число пересечений, количество слоёв), а также критериев надёжности,

включающих тепловую и электромагнитную совместимость, временные задержки сигналов и плотность упаковки, при этом основным критерием является учет эффективности будущей трассировки [12]. В настоящее время наибольшее распространение в алгоритмах размещения получил критерий суммарной длины соединений, который позволяет косвенно учитывать другие критерии, способствуя уменьшению длин соединений, улучшению электрических характеристик и повышению надёжности СБИС [11]. Решение задачи размещения требует предварительной формализации, основанной на представлении схемы в виде взвешенного графа $G = (V, E)$, где каждое ребро $(i, j) \in E$ описывает электрическое соединение между компонентами v_i и v_j. В соответствии с введённой взвешенной графовой моделью структуры СБИС множество компонентов определяется следующим выражением:

$$V = \{v_1, v_2, \dots, v_n\}, \quad (1)$$

где каждому компоненту v_i необходимо присвоить позицию на плоскости кристалла. Размещение выполняется в узлах координатной решётки, заданной на прямоугольной конструкции декартовой системой координат (s, t). Цель задачи состоит в отображении заданного графа G в решётку G_r таким образом, чтобы все вершины из V заняли конечное множество узлов решётки, а также соблюдался выбранный критерий оптимизации. Таким образом, каждому компоненту v_i необходимо сопоставить координаты (s_i, t_i) на плоскости кристалла, что определяет окончательное расположение элементов в узлах координатной сетки.

Для формализации взаимодействий между компонентами вводится следующая матрица весов:

$$W = \{w_{ij}\}, \quad (2)$$

где w_{ij} – это критические соединения, которые напрямую зависят от длины межсоединений. Также на основе матрицы весов определяется метрика оценки качества проектных решений – суммарная длина межсоединений [11], которая рассчитывается по следующей формуле:

$$L = \sum_{(i,j)\in E} w_{ij} \cdot d(v_i, v_j), \quad (3)$$

где $d(v_i, v_j)$ – расстояние между позициями компонентов v_i и v_j, как правило, рассчитываемое по манхэттенской метрике с учётом ограничений топологии проводников [1, 11 – 13].

Также дополнительно учитываются технологические и топологические ограничения, включая плотность размещения в локальных областях и ограничение на число линейных сегментов проводников [11 – 13], которые формализуются в виде штрафных функций и включаются в общую целевую функцию. В результате формируется многокритериальная модель размещения, объединяющая метрические показатели качества и технологические нормы, определяющие корректность конфигурации. Сложность трассировки $LS(P)$, оцениваемая числом линейных сегментов, учитывается в целевой функции с учётом критичности соединений и вычисляется по следующей формуле:

$$LS(P) = \sum_{(i,j)\in E} w_{ij} \cdot LS_{ij}. \tag{4}$$

В том числе к технологическим ограничениям относят минимальные расстояния между компонентами, плотность размещения в локальных областях и допустимость конфигураций с точки зрения технологических норм [11, 12]. Нарушения этих требований выражаются в виде штрафных функций, рассчитываемых на основе следующего выражения:

$$LT(P) = \sum_{i<j} w_{ij} \cdot \max(0, d_{\min} - d(v_i, v_j)), \tag{5}$$

это позволит учесть конфликты размещения и исключать недопустимые конфигурации.

Таким образом, учитывая все описанные метрики [11, 12], целевая функция рассчитывается по следующей формуле:

$$F(P) = \lambda_1 \cdot L(P) + \lambda_2 \cdot LS(P) + \lambda_3 \cdot T(P), \tag{6}$$

где $L(P)$ – суммарная длина межсоединений, $T(P)$ – штраф за нарушение технологических ограничений, а на основе коэффициентов λ_k определяются относительные веса каждого критерия, что позволяет обеспечивать баланс между длиной

межсоединений, сложностью трассировки и соблюдением технологических ограничений при оценке проектных решений.

Для систематизации взаимодействий между компонентами, связями и критериями качества вводится концептуальная модель, представленная на рисунке 1. Модель построена для двумерного случая размещения компонентов на плоскости кристалла, что соответствует базовой постановке задачи [1, 8]. При этом заложенные в модель метрики поиска и принципы адаптивного управления параметрами не зависят от размерности пространства. Данное свойство позволяет рассматривать предложенную концептуальную модель как основу для последующего развития методов трёхмерного размещения компонентов, где потребуется модификация системы координат и учёт дополнительных критериев, таких как объёмная плотность и тепловые характеристики, без изменения ядра адаптивной параметризации.

Модель содержит требования к размещению компонентов и особенностям задачи, требующая применения таких алгоритмов, которые позволяют учитывать изменения процесса поиска. Данная модель является основой при разработке адаптивного механизма параметризации.

Представленная концептуальная модель позволяет рассматривать пространство поиска в данной задаче как разреженное и нелинейное, что определяет высокую сложность поиска решений и особенности поведения эвристических и метаэвритических алгоритмов. Каждая конфигурация размещения задаётся дискретными координатами компонентов на плоскости кристалла, вследствие чего общее число возможных состояний носит факториальный характер. При этом множество допустимых решений ограничено, так как значительная часть конфигураций нарушает технологические ограничения, такие как минимальные расстояния, плотность размещения и топологические нормы. Таким образом только четверть пространства поиска может содержать валидные или близкие к валидным решения.

В условиях разреженного пространства поиска целевая функция характеризуется высокой нелинейностью, которая обусловлена сочетанием влияния заданных метрик оценки качества проектных решений. Перемещение отдельного компонента неоднородно влияет на структуру межсоединений, вызывая разрывы в функции и локальные неровности ее поверхности. Даже незначительные локальные перестановки приводят к резким изменениям значений целевой функции, что затрудняет применение методов локального поиска.

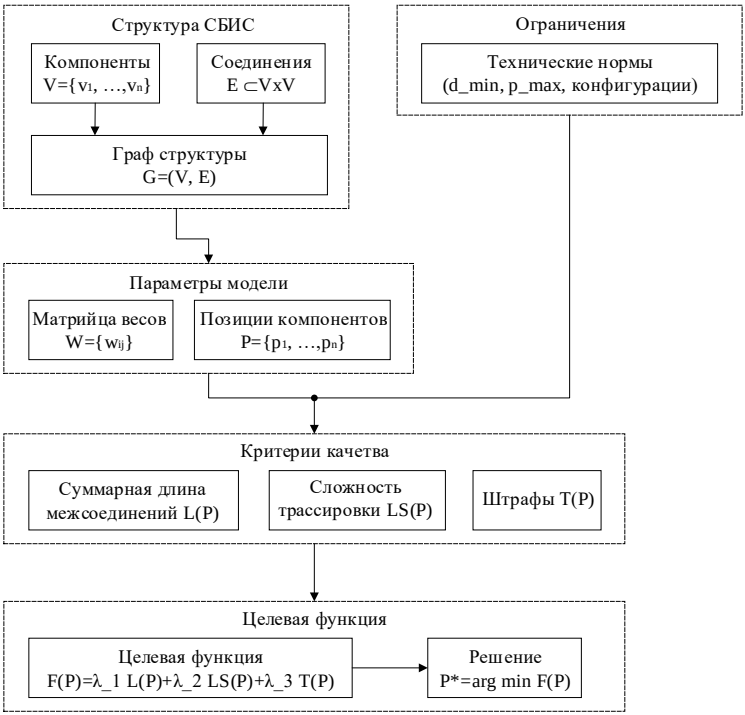


Рис. 1. Концептуальная модель критериев и ограничений задачи размещения компонентов СБИС

Кроме локальных неровностей и резких изменений значений целевой функции, так же учитывается сложность, создаваемая локальными минимумами. Это обусловлено плотной связностью схем, неоднородным распределением критически значимых соединений и конфликтами между заданными метриками. Наличие локальных минимумов с протяжёнными областями притяжения повышает вероятность преждевременной сходимости метаэвристик при фиксированных параметрах. В сочетании с разреженностью допустимого множества решений это усиливает чувствительность к начальным условиям и снижает устойчивость процесса поиска.

Таким образом, из формализованного представления задачи размещения компонентов СБИС следует, что пространство поиска характеризуется разрывной, многомодальной и существенно разреженной структурой, а многокритериальный характер целевой функции ограничивает применимость известных методов. Выявленные

свойства пространства решений и требования к устойчивости алгоритмов свидетельствуют о необходимости разработки адаптивного механизма параметризации, ориентированного на учёт текущего состояния процесса поиска.

4. Динамическая модель адаптивной параметризации. Эффективность эвристических и метаэвристических алгоритмов в задаче размещения СБИС зависит от их адаптивности к процессу поиска. При этом статическая параметризация не в полной мере отражает текущее состояние популяции решений и не обеспечивает согласование между глобальным и локальным поиском в условиях разрывной и многомодальной целевой функции. В этой связи целесообразно применение динамической модели адаптивной параметризации, ориентированной на устойчивое управление процессом поиском за счёт корректировки ключевых параметров на основе метрик состояния популяции, признаков стагнации и локальных характеристик пространства решений.

В предлагаемой динамической модели корректировка осуществляется на основе количественных метрик, отражающих текущее состояние популяции решений и динамику целевой функции в процессе оптимизации. Для формального описания этих процессов введём следующие величины, где P_t – множество решений на итерации t, ΔF – целевая функция. В частности, одной из ключевых метрик [11, 12, 14] является скорость улучшения целевой функции, которая характеризует степень приближение алгоритма к лучшим решениям F_{best} за фиксированный интервал Δt . При этом её количественная оценка вычисляется на основе следующего выражения:

$$\Delta F_t = \frac{F_{\text{best}} \cdot (t - \Delta t) - F_{\text{best}}(t)}{\Delta t}. \quad (7)$$

Тогда для оценки скорости улучшения целевой функции введём показатель относительного прогресса поиска [11, 15]. Этот показатель позволяет нормировать изменение целевой функции и оценивать эффективность приближения алгоритма к локальному оптимуму независимо от абсолютных параметров метрики. Для этого используется относительное улучшение δ_t, которое рассчитываются на основе следующего выражения:

$$\delta_t = \frac{F_{\text{best}} \cdot (t - \Delta t) - F_{\text{best}}(t)}{F_{\text{best}} \cdot (t - \Delta t) + \varepsilon}, \quad (8)$$

где $\varepsilon > 0$ – положительная константа, обеспечивающая численную стабильность вычислений, которая выбирается лицом, принимающим ращение (ЛПР) из интервала [0; 1]. Здесь интенсивность параметров поиска увеличивается при минимальных значениях δ_t, что приводит к стагнации и уменьшается при высоких значениях, соответствующих прогрессу в улучшении значений целевой функции.

При расчёте введённых метрик необходимо учитывать показатель разнообразия популяции решений, ее диверсификацию. В качестве оценки количественной меры диверсификации рассчитывается дисперсия расстояний D_t между решениями в популяции [11, 14, 16]. Для формального описания предложенной метрики введём следующие обозначения. Позиция компонента в решении i на итерации t обозначается как $p_i^{(t)}$, тогда дисперсия рассчитывается по следующей формуле:

$$D_t = \frac{1}{|\ P_t\ |} \sum_{x \in P_t} \frac{1}{n} \sum_{i=1}^{n} d(p_i^{(x)}, \bar{p}_i), \tag{9}$$

где $\bar{p}_i$ – среднее положение компонента i в популяции.

Наряду с геометрической мерой разнообразия также учитывается их функциональная форма σ_t^2 [12, 14, 16], основанная на распределении значений целевой функции в популяции, которая выражается через функцию дисперсии:

$$\sigma_t^2 = \frac{1}{|\ P_t\ |} \sum_{x \in P_t} (F(x) - \bar{F}_t)^2. \tag{10}$$

Снижение значений D_t или σ_t^2 указывает на снижение структурного разнообразия популяции и её концентрации в ограниченной области пространства поиска, это связано с повышенной вероятностью преждевременной сходимости алгоритма.

Для отслеживания момента прекращается положительная динамика улучшения качества решений вводится бинарный индикатор S_t, рассчитываемый по следующей формуле:

$$S_t = \begin{cases} 1, & \text{если } \delta_t < \tau_1 \text{ и } D_t < \tau_2, \\ 0, & \text{иначе.} \end{cases} \tag{11}$$

где τ_1, τ_2 – пороговые значения, которые задаются в процессе поиска ЛПР. Значение $S_t = 1$ свидетельствует о состоянии стагнации, которое возникает при отсутствии улучшении значения целевой функции (низкая относительная скорость улучшения δ_t) и недостаточном структурном разнообразии популяции (низкое значение метрики D_t). В противном случае, при $S_t = 0$, процесс считается активным, это означает, что либо сохраняется устойчивая положительная динамика улучшения значений целевой функции, либо популяция обладает достаточным разнообразием для продолжения процесса поиска проектных решений.

Для оценки сложности поиска при локальном окружении решений применяется градиентная вариация текущего значения целевой функции [9, 13], которая рассчитываются по следующей формуле:

$$H_t(x) = \frac{1}{k} \sum_{y \in N_k(x)} | F(y) - F(x) |, \tag{12}$$

где $N_k(x)$ – множество k ближайших решений x, измеряемых по Манхэттенскому расстоянию [1, 8, 11]. Этот показатель отражает локальные изменения целевой функции и используется для количественной оценки «неровности» или сложности локального окружения каждого решения. При этом высокие значения локальной характеристики $H_t(x)$ указывают на дискретность и неоднородность области целевой функции в окрестности точки x. Для интегральной оценки используется сводный индикатор M_t состояния популяции, который рассчитывается по следующей формуле:

$$M_t = \alpha_1 \cdot \delta_t + \alpha_2 \cdot D_t + \alpha_3 \cdot H_t - \alpha_4 \cdot S_t, \tag{13}$$

где H_t – значение локальной неоднородности пространства поиска, а α_i – весовые коэффициенты, определяющие интегральную оценку каждой метрики. Для определения весовых коэффициентов $\alpha_1, \alpha_2, \alpha_3, \alpha_4$ в интегральном индикаторе M_t использовалась процедура предварительной калибровки на основе вычислительного эксперимента. С этой целью сформирован набор тестовых топологий, не участвовавших в дальнейших сравнительных испытаниях (валидационный набор), включающий 10 схем различной размерности и связности. Для каждой комбинации весов из дискретного множества значений $\alpha_i \in \{0.1, 0.2, \ldots, 0.9\}$ с шагом 0.1, удовлетворяющих следующему условию $\sum_{i=1}^{4} \alpha_i = 1$, выполнялось по 20 независимых

итераций алгоритма поиска. Выбранная таким образом комбинация ($\alpha_1 = 0{,}3, \alpha_2 = 0{,}3, \alpha_3 = 0{,}2, \alpha_4 = 0{,}2$) позволила достичь требуемого баланса между скоростью сходимости и устойчивостью поиска при тестировании на валидационных схемах.

На основе сводного индикатора состояний популяций решений сформируем математическую модель для адаптивного изменения параметров алгоритма. Обозначим через $\Theta(t)$ множество изменяемых параметров метаэвристического алгоритма на итерации t. В зависимости от используемого метода в состав $\Theta(t)$ входит вероятность мутации $\mu(t)$, коэффициенты кроссинговера $\chi(t)$, параметры локального поиска, интенсивности стохастических операторов, а также коэффициенты применяемых роевых алгоритмов. Тогда задача динамической модели состоит в определении функциональной зависимости, которая рассчитывается по следующей формуле:

$$\Theta(t+1) = \Phi(\Theta(t), M_t), \tag{14}$$

где Ф – оператор адаптации, преобразующий текущие параметры в их новые значения.

В данной модели выполнение оператора мутации реализует механизм глобального исследования пространства решений. Поэтому при снижении разнообразия решений в популяции и стагнации целевой функции интенсивность выполнения оператора мутации возрастает [12, 14, 16], что рассчитывается на основе следующего выражения:

$$\mu(t+1) = \mu(t) + \eta_\mu \cdot (1 - D_t) \cdot (1 - \delta_t), \tag{15}$$

где η_μ – коэффициент регулирования, определяющий величину корректировки параметра в зависимости от состояния популяции, а $\mu(t)$ – значение вероятности мутации на текущей итерации t.

Для обеспечения допустимого диапазона значений интенсивности выполнение оператора мутации рассчитывается на основе следующего выражения [1, 3, 11]:

$$\mu(t+1) = \min(\mu_{\max}, \max(\mu_{\min}, \mu(t+1))). \tag{16}$$

Таким же образом учитывается выполнения оператора кроссинговера. При устойчивом улучшении получаемых решений интенсивность оператора кроссинговера применения увеличивается, а при уменьшении динамики улучшений его корректировка

ослабляется. Этот процесс рассчитывается на основе следующего выражения [1, 11]:

$$\chi(t+1) = \chi(t) + \eta_\chi \cdot (\delta_t - \tau_\delta), \tag{17}$$

где τ_δ – порог, определяющий режим «стабильного улучшения», а η_χ – коэффициент скорости адаптации выполнения оператора кроссинговера. Заметим, что в этом выражении дополнительно учитывается локальная структура пространства решений. При превышении допустимого уровня локальной неоднородности концентрируется направленный локальный поиск, интенсивность которого корректируется на основе следующего выражения [8, 11]:

$$\rho(t+1) = \rho(t) + \eta_\rho \cdot H_t, \tag{18}$$

где $\rho(t)$ – интенсивность локального поиска на итерации t, H_t – локальная вариация ландшафта поиска, а η_ρ – коэффициент скорости адаптации для оператора локального поиска. Отметим, что каждый коэффициент адаптации η $(\eta_\mu, \eta_\chi, \eta_\rho)$ является индивидуальным параметром скорости поиска, определяющим величину коррекции соответствующего оператора (мутации μ, кроссинговера χ, локального поиска ρ) в зависимости от текущих значений метрик состояния популяции D_t, δ_t и H_t соответственно.

В дополнение к базовым правилам адаптации при фиксации стагнации процесса поиска применяется компенсирующий механизм корректировки параметров. При $(S_t = 1)$ обновление параметров выполняется по следующей формуле [11]:

$$\Theta(t+1) = \Theta(t) + \Delta\Theta_{\text{stag}}, \tag{19}$$

где $\Delta\Theta_{\text{stag}} = (\delta_\mu, -\delta_\chi, \delta_\rho)$ – вектор направленных приращений. Его переменные, задаваемые в следующих диапазонах $\Delta_\mu \in (0, \Delta_\mu^{\max}]$, $\Delta_\rho \in (0, \Delta_\rho^{\max}]$ и $\Delta_\chi \in (0, \Delta_\chi^{\max}]$, позволяют выявлять увеличение интенсивности операторов поиска. Такая процедура корректировки реализует антистагнационный механизм, направленный на восстановление поисковой активности и предотвращение преждевременной сходимости алгоритма.

Рассмотренные процедуры и механизмы адаптации представляются в виде единой формализованной модели. При этом для

каждого параметра $\theta_i(t) \in \Theta(t)$ обновление рассчитывается на основе следующего выражения:

$$\theta_i(t+1) = \theta_i(t) + \eta_i \cdot g_i(M_t), \tag{20}$$

где η_i – коэффициент адаптации, а $g_i(M_t)$ функция отклика, определяющаяся типом используемого параметра. Вероятность применения операторов мутации $\rho_{\text{ом}}$ и кроссинговера $\rho_{\text{ок}}$. Она пропорциональна произведению следующих метрик $\rho_{\text{ом}} = g_\mu(M_t) \propto (1 - D_t) \cdot (1 - \delta_t)$, $\rho_{\text{ок}} = g_\chi(M_t) \propto (\delta_t - \tau_\delta)$, а интенсивность применения локального поиска $\rho_{\text{лп}}$ рассчитывается на основе произведением $g_\rho(M_t) \propto H_t$.

Для обеспечения устойчивости процесса адаптации и исключения колебательных режимов при динамическом обновлении параметров применяется процедура сглаживания [9, 11, 16]. Их обновление выполняется с учётом предыдущих значений на основе следующего выражения:

$$\theta_i(t+1) = \alpha \cdot \theta_i(t+1) + (1 - \alpha) \cdot \theta_i(t), \tag{21}$$

где $0 < \alpha < 1$ – коэффициент сглаживания. Данный механизм позволяет ограничивать влияние кратковременных флуктуаций метрик состояния и способствует стабильному процессу поиска.

Для практической реализации предложенной процедуры адаптации требуется формализовать условия переключения режимов и ограничений динамического обновления параметров. С этой целью вводится система пороговых значений, обеспечивающая корректную интерпретацию метрик состояния и стабилизацию реакции процесса поиска на изменения. Пусть δ_t, D_t, H_t и S_t – вычисляемые метрики процесса поиска. Тогда система порогов выражается следующим образом:

$$\Theta_{\text{thr}} = \{\tau_\delta, \tau_D, \tau_H\}, \tag{22}$$

где каждый порог определяет переход между режимами поиска.

На основе введённой системы пороговых значений формируются критерии перехода между режимами работы алгоритма. Порог улучшения τ_δ выделяет фазы прогресса и стагнации: при $\delta_t < \tau_\delta$ – режим стагнации, при $\delta_t \geq \tau_\delta$ – режим локального улучшения. Порог разнообразия τ_D фиксирует уменьшения области поиска: при $D_t < \tau_D$ реализуется глобальный поиск. Порог

неоднородности τ_H отражает сложность ландшафта: при $H_t > \tau_H$ реализуется локальный поиск. Совокупность этих порогов реализует логические условия для адаптивного перехода между различными режимами работы алгоритма в процессе поиска. Однако, в условиях разрывной структуры целевой функции и скачкообразного изменения метрик, прямое применение данных процедур и механизмов приведет к нестабильному поведению алгоритма. Поэтому для устранения этой проблемы и обеспечения устойчивости процедуры адаптации вводятся стабилизирующие механизмы, предотвращающие резкие колебания, учитываемых параметров.

Рассмотрим данные механизмы более подробно [12 – 15].

1. Механизм сглаживания параметров применяется для предотвращения осцилляций. Здесь параметр $\theta_i(t)$ обновляется с использованием экспоненциального фильтра, что рассчитывается по следующей формуле:

$$\theta_i(t+1) = \alpha \cdot \theta_i(t+1) + (1-\alpha) \cdot \theta_i(t), \tag{23}$$

где $0 < \alpha < 1$ – коэффициент сглаживания.

2. Механизм нормирования представляет собой процедуру приведения обновлённых значений параметров к допустимому диапазону за счет заданных граничных условий, что рассчитывается по следующей формуле:

$$\theta_i(t+1) = \min(\theta_i^{\max}, \max(\theta_i^{\min}, \theta_i(t+1))), \tag{24}$$

где $\theta_i^{\min}$ и $\theta_i^{\max}$ – соответственно нижняя и верхняя границы для параметра θ_i . Это исключает выход значений за область, определённую физической постановкой задачи или алгоритмическими ограничениями.

3. Компенсирующий механизм используется для преодоления локальных оптимумов. Если индикатор стагнации принимает значение $S_t = 1$, что соответствует состоянию стагнации поиска, то выполняется дополнительная корректировка параметров по следующей формуле:

$$\theta_i(t+1) \leftarrow \theta_i(t+1) + \Delta\theta_i^{\text{stag}}, \tag{25}$$

где амплитуда $\Delta\theta_i^{\text{stag}}$ определяется уровнем стагнации.

Взаимосвязь между пороговыми критериями и стабилизирующей процедурой поиска представлена в виде следующей схемы (Рис. 2).

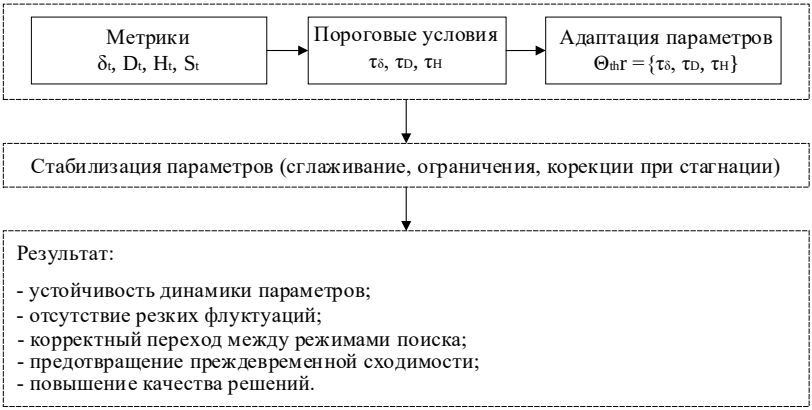


Рис. 2. Схема работы пороговых критериев и стабилизирующих процедур

На основе проведённых исследований установлено, что применение приведенных метрик, пороговых критериев и стабилизирующих механизмов позволяет построить устойчивый алгоритм управления поисковым процессом. Интеграция данной адаптивной модели в структуру эвристического или метаэвристического алгоритма позволяет реализовать двухконтурную организацию поискового процесса, в которой параметры операторов динамически корректируются в соответствии с текущим состоянием популяции решений и локальными особенностями пространства поиска. Такая организация поиска позволяет обеспечивать высокую чувствительность алгоритма к изменяемой динамике процесса, устойчивой сходимости, а также повысить способность алгоритмов к преодолению локальных оптимумов. Таким образом, системная адаптивная параметризация способствует преодолению ключевых ограничений методов со статическими параметрами, при этом обеспечивая повышение качества решений в условиях высокой разрывности и многомодальности пространства конфигураций.

5. Интеграция адаптивной модели параметризации в метаэвристический алгоритм. Адаптивная модель реализуется за счет механизма корректировки параметров, который анализирует состояние популяции и изменяет характеристики операторов поиска на каждой итерации. В результате алгоритм работает в двухконтурном

режиме, в котором классические операторы поддержки процесса поиска решений функционируют совместно с адаптивным механизмом регулирования параметров, обеспечивающей согласование поисковой стратегии с локальными особенностями пространства конфигураций.

На рисунке 3 представлена схема интеграции адаптивной модели параметризации в структуру метаэвристического алгоритма.

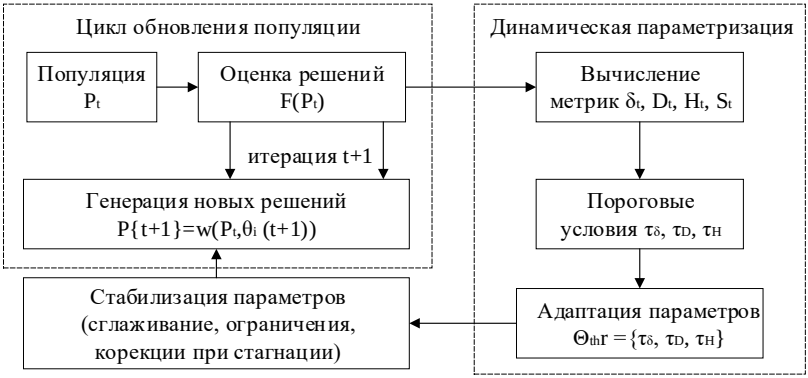


Рис. 3. Схема интеграция адаптивной модели параметризации в метаэвристический алгоритм

В данной схеме отражается взаимодействие основных модулей системы: базового цикла метаэвристического поиска, блока вычисления метрик и механизма адаптивного обновления параметров. Что позволяет идентифицировать двухконтурную организацию поискового процесса, где операторы поиска работают в рамках внешнего адаптивного контура, который управляет их параметрами на основе обратной связи. Структурно алгоритм включает четыре ключевых модуля: генератор новых решений, блок оценки, адаптивный модуль корректировки параметров и стабилизирующий механизм, которые взаимодействуют последовательно через встроенный контур обратной связи. В такой структуре параметры $\Theta(t)$ перестают быть фиксированными величинами и определяются динамически изменяющимися метриками поискового процесса.

Для формализации представленной адаптивной модели введём следующие обозначения: пусть P_t – популяция на итерации t, а $\Theta(t)$ – множество параметров алгоритма. Тогда интеграция адаптивной модели в процесс поиска выражается следующей системой уравнений [8 – 11]:

$$P_{t+1} = \Psi(P_t, \Theta(t)), \Theta(t+1) = \Omega(\Theta(t), M_t), \quad (26)$$

где Ψ – оператор генерации новых решений,Ω – адаптивный механизм корректировки параметров, M_t – интегральный показатель состояния популяции.

Далее рассмотрим более подробно работу адаптивного механизма обновление параметров Ω на каждой итерации. Данный механизм включает следующие этапы.

1. Оценка состояния популяции. На основе текущей популяции P_t вычисляются базовые метрики, характеризующие динамику поиска: скорость улучшения δ_t, разнообразие D_t, неоднородность ландшафта H_t, дисперсию целевой функции σ_t^2 и индикатор стагнации S_t.

2. Пороговая интерпретация состояния. Полученные метрики сопоставляются с заданными пороговыми значениями $\tau_\delta, \tau_D, \tau_H$. На основе этого сопоставления определяется текущий режим работы алгоритма (прогресс, стагнация, локальная сложность, сжатие популяции), который задаёт стратегию адаптации.

3. Обновление параметров. В соответствии с установленным режимом выполняется корректировка множества параметров $\Theta(t)$, которая рассчитывается по следующей формуле:

$$\Theta(t+1) = \Theta(t) + \eta \cdot g(M_t), \tag{27}$$

где η – коэффициент скорости адаптации, а $g(M_t)$ – функция отклика, определяющая направление и величину изменения параметров на основе интегрального показателя состояния M_t , агрегирующего оценки предыдущих этапов.

4. Нормирование параметров. Чтобы исключить выход значений за допустимые пределы, полученные параметры ограничиваются заданными границами, что выражается следующей формулой:

$$\Theta(t+1) = \min(\Theta_{\max}, \max(\Theta_{\min}, \Theta(t+1))). \tag{28}$$

где $\Theta_{\min}$ и $\Theta_{\max}$ – векторы минимально и максимально допустимых значений для каждого параметра.

5. Стабилизация параметров. Для обеспечения плавности изменений и подавления шума применяется экспоненциальное сглаживание, которое рассчитывается по следующей формуле:

$$\Theta(t+1) = \alpha \cdot \Theta(t+1) + (1-\alpha) \cdot \Theta(t), \tag{29}$$

где $\alpha \in (0,1]$ – коэффициент сглаживания, подбирается ЛПР на основе эмпирических данных. Данная процедура позволяет предотвратить резкие колебания параметров и обеспечить устойчивость динамической настройки алгоритма.

6. Генерация новой популяции. На основе скорректированных и стабилизированных параметров $\Theta(t+1)$ выполняются базовые операторы поиска, которые позволяют сформировать следующее поколение решений, рассчитываемое по следующей формуле:

$$P_{t+1} = \Psi(P_t, \Theta(t+1)). \quad (30)$$

Таким образом, интеграция адаптивной модели позволяет создать саморегулируемую систему с обратной связью в метаэвристическом алгоритме, где ключевые параметры динамически корректируются под структуру пространства поиска и текущие характеристики динамически изменяемого процесса. Это позволяет обеспечивать повышение устойчивости работы алгоритма, снижать вероятность преждевременной сходимости и способствовать улучшению качества получаемых проектных решений.

6. Вычислительный эксперимент. Для подтверждения эффективности предложенного подхода и проведения вычислительного эксперимента разработана программная подсистема, включающая адаптивную модель параметризации, обеспечивающая цикл формирования и анализа конфигураций при размещении компонентов СБИС. Цель вычислительного эксперимента заключается в анализе влияния адаптивного механизма на качество формируемых решений, устойчивость алгоритма и характер сходимости в условиях многокритериальной и многомодальной структуры пространства поиска. Эксперимент выполнен на тестовых топологиях различной размерности и плотности связей (таблица 2). Используемые в таблице значения 100, 300, 600 компонентов соответствуют типовым размерам тестовых примеров из известных бенчмарков (MCNC, IWLS, ISPD) [17, 18], которые широко применяются для верификации алгоритмов размещения. Такой диапазон позволяет оценить масштабируемость предложенного подхода и выявить закономерности его работы при переходе от разреженных к плотным структурам без потери общности получаемых выводов. При этом оцениваются показатели качества размещения, вариативность результатов при множественных итерациях, динамика изменения параметров $\Theta(t)$ и заданных метрик, а также особенности сходимости

адаптивного алгоритма по сравнению с известными аналогами. Для количественной характеристики тестовых топологий введён коэффициент сложности топологии C_t , рассчитываемый как отношение числа рёбер в графе схемы к числу рёбер в полном графе с тем же числом вершин: $C_t = 2|E|/\big(n(n-1)\big)$, где n – число компонентов (вершин), |E| – число соединений (рёбер) в схеме. Данный коэффициент является безразмерной нормированной величиной, принимающей значения в интервале [0, 1], и отражает связность схемы, то есть чем ближе C_t к единице, тем сложнее топология и тем выше требования к адаптивным свойствам алгоритма. А также оцениваются показатели качества размещения, вариативность результатов при множественных итерациях, динамика изменения параметров *Θ(t)*, заданных метрик и особенности сходимости адаптивного алгоритма по сравнению с известными аналогами.

Тестовые сценарии включали три класса топологий: разреженные схемы с низкой связностью и небольшой плотностью активных соединений; плотные схемы, характеризующиеся высокой степенью взаимодействия и критических путей; а также смешанные топологии, в которых плотные локальные области сочетаются с разрежёнными сегментами. Такой набор тестовых топологий позволяет оценить поведение разработанного метаэвристическиого алгоритма, включавшего модель адаптивной параметризации в условиях неоднородной структуры пространства поиска, определить чувствительность к топологическим особенностям, устойчивость решений и исследовать влияние связности на динамику сходимости.

Таблица 2. Параметры тестовых топологий и режимы параметризации

Тип топологии	**Число компонентов n, (шт.)**	**Плотность связей (среднее число связей на компонент), у.е.**	**Коэффициент сложности топологии (нормированный) в интервале [0, 1]**	**Режимы параметризации**
Разреженная	100	3,2	0,28	Статические, фиксированная, адаптивная
Смешанная	300	6,8	0,51	Статические, фиксированная, адаптивная
Плотная	600	11,5	0,84	Статические, фиксированная, адаптивная

Вычислительный эксперимент проводился в унифицированной программной среде [19], включающей модуль генерации топологий, вычислительный модуль эвристических и метаэвристических

алгоритмов и подсистему различных метрик. Реализация выполнена на языке Python с использованием специализированных библиотек, включая NumPy и SciPy для численных вычислений, NetworkX для работы со схемами, а также DEAP для реализации эвристик и метаэвристик [20]. Вычисления выполнялись на компьютере с операционной системой Windows 11, процессором AMD Ryzen 5, 32 ГБ оперативной памяти и видеокартой NVIDIA RTX 3060, что исключает влияние аппаратных ограничений на результаты эксперимента.

Для обеспечения корректного сопоставления алгоритмов использовались единые настройки вычислительного процесса. Размер популяции составлял 40-80 альтернативных решений в зависимости от размерности задачи, число итераций фиксировано и равно 500. Как известно из литературы [1] увеличение числа итераций приводит к улучшению значения целевой функции до определенного предела. Выбор числа итераций (500) был определён на основе анализа сходимости, представленного на рисунке 4. На графике видно, что после 350-400 итераций интегральный показатель качества F(P) достигает асимптотического значения, а относительное улучшение δt (рассчитываемое по формуле (8)) становится меньше порогового значения $\tau\delta$=0,01, что интерпретируется как состояние практической сходимости. Дальнейшее увеличение числа итераций приводит к незначительному улучшению (менее 1% при удвоении до 1000 итераций) при значительном росте вычислительных затрат. Для адаптивного механизма параметризации заданы допустимые интервалы параметров Θmin и Θmax, включающие вероятность применения операторов мутации, кроссинговера и интенсивность локального поиска. Обновление параметров выполнялось с учетом коэффициентов скорости η и сглаживания α. В эксперименте рассмотрены три режима работы алгоритма: со статическими параметрами, с фиксированной многоэтапной настройкой и с адаптивной параметризацией, при которой параметры $\Theta(t)$ обновляются на основе заданных метрик, технологических ограничений и граничных условий.

Оценка эффективности алгоритмов проводилась по совокупности метрик, включающих суммарную длину межсоединений $L(P)$, сложность трассировки $LS(P)$ и показатель технологических ограничений $T(P)$. Для анализа работы адаптивного механизма использовались динамически изменяемые метрики поискового процесса, характеризующие скорость улучшения, разнообразие популяции, структурные различия, дисперсию решений

и наличие фаз стагнации. Устойчивость алгоритмов оценивалась по вариативности результатов, дисперсии при множественных итерациях и диапазону значений интегрального показателя качества.

Результаты вычислительного эксперимента сведены в таблице 3. Анализ таблицы подтверждает влияние адаптивной параметризации на качество и устойчивость вычислительного процесса. Для всех классов тестовых топологий получены численные оценки интегрального показателя качества $F(P)$.

Анализ значений из таблицы 3 показывает, что адаптивный алгоритм формирует решения с меньшими значениями L(P) и F(P) по сравнению со статическими и фиксированными методами, что подтверждает эффективность механизма динамической параметризации. При этом временные затраты адаптивного алгоритма лишь незначительно превышают время работы методов со статической параметризацией (на 5-10% в зависимости от размерности), что объясняется расчётом диагностических метрик δt, Dt, Ht, St и обновлением параметров на каждой итерации. Полученные значения времени (от 124 с для 100 компонентов до 441 с для 600 компонентов) позволяют сделать вывод о практической применимости предложенного подхода в рамках интерактивного проектирования СБИС умеренной сложности, а для более крупных схем целесообразно применение иерархических стратегий размещения.

Таблица 3. Численные результаты качества размещения

Класс топологии	Режим параметризации	L(P) длина межсоединений, у.е.	LS(P) трассировка, у.е.	T(P) технологические ограничений (нормированный)	F(P) интегральный показатель, у.е.	$\sigma^2(F)$ дисперсия F, у.е.^{2}	Время работы, с
Разреженная	Статические	8 610	1 420	0.18	8 420	1.25×10^5	112
	Фиксированная	8 120	1 310	0.12	7 920	8.6×10^4	118
	Адаптивная	7 410	1 180	0.07	7 340	4.2×10^4	124
Смешанная	Статические	28 900	4 870	0.31	31 200	3.9×10^5	215
	Фиксированная	26 100	4 420	0.21	27 400	2.6×10^5	223
	Адаптивная	22 300	3 980	0.13	24 100	1.1×10^5	234
Плотная	Статические	78 200	11 600	0.52	86 500	8.2×10^5	418
	Фиксированная	69 300	10 200	0.37	76 400	5.4×10^5	427
	Адаптивная	59 200	8 900	0.24	65 100	2.7×10^5	441

На рисунке 4 представлен график сходимости интегрального показателя качества $F(P)$ для различных режимов параметризации, отражающий особенности поведения алгоритма на начальном этапе поиска и выход в область устойчивых значений.

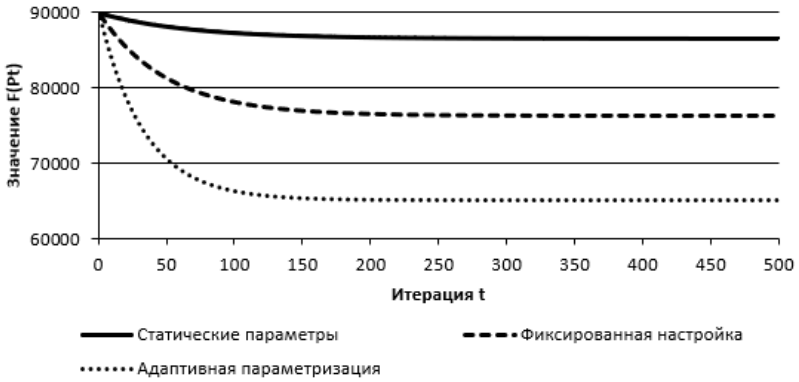


Рис. 4. График сходимости интегрального показателя качества $F(P)$ при различных режимах параметризации

На рисунке 5 представлен график сравнения значений *L(P), LS(P), T(P)* и *F(P)* при различных режимах параметризации, отражающий влияние отдельных показателей на формирование интегрального критерия качества.

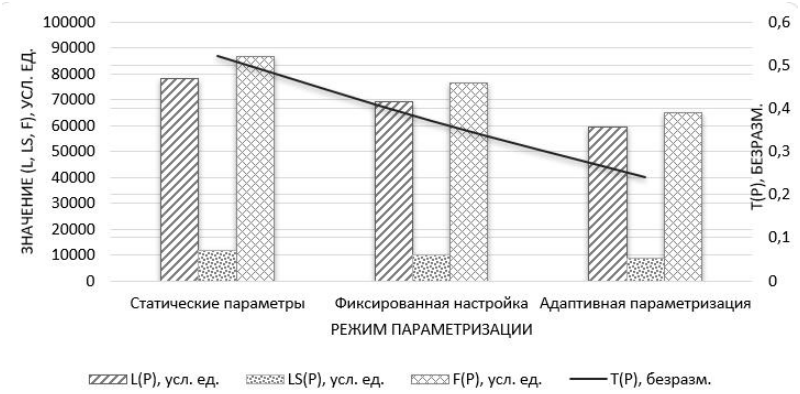


Рис. 5. График сравнения значений *L(P), LS(P), T(P) и F(P)* при различных режимах параметризации

Проведенный анализ представленных таблиц и графиков $F(P_t)$ показал, что разработанный метаэвристический алгоритм, включающий динамическую модель параметризации, характеризуется стабильным снижением целевой функции и меньшим количеством фаз стагнации. При этом траектория сходимости отличается сглаженным профилем благодаря стабилизации параметров, что снижает вероятность резких отклонений и преждевременной остановки в локальных оптимумах. В свою очередь, оценка устойчивости подтверждает снижение вариативности результатов предложенного алгоритма, проявляющаяся в уменьшении диапазона значений $F(P)$ и дисперсии решений в 1.3-2.2 раза (в зависимости от набора данных), а также в смещении медианы решений в область наилучших конфигураций задаваемых параметров. Отметим, что для методов со статическими задаваемыми параметрами характерно расширенное распределение результатов и высокая чувствительность к начальному набору данных.

На рисунке 6 представлен график изменения параметров $\Theta(t)$, отражающий адаптацию алгоритма к структуре задачи и согласованность механизма параметрического управления с показателями процесса поиска.



Рис. 6. График изменения параметров $\Theta(t)$

Совокупность полученных результатов подтверждает, что адаптивный механизм параметризации обеспечивает стабильное поведение предложенного алгоритма и повышает качество решений, снижая вероятность преждевременной сходимости. Сравнительный анализ показал, что данный алгоритм по заданным метрикам качества

размещения демонстрирует лучшие результаты по сравнению с методами, использующими статические параметры и ручную многоэтапную настройку. При этом улучшение интегрального показателя качества $F(P)$ в среднем составляет 8-15 % для разреженных структур и 12-25 % для структур с высокой плотностью связей.

7. Заключение. В работе исследована проблема повышения эффективности применения эвристических и метаэвристических алгоритмов при решении задачи размещения компонентов СБИС, характеризующейся высокой комбинаторной сложностью, разрывностью поверхности поиска и технологическими ограничениями топологического синтеза. В статье приведена многокритериальная постановка задачи, учитывающая длину межсоединений, сложность трассировки и технологические ограничения. Анализ поведения метаэвристик со статическими параметрами показал, что их эффективность существенно ограничивается отсутствием адаптации поискового процесса, что проявляется в преждевременной сходимости, высокой вариативности результатов и снижении устойчивости поиска.

В работе предложена динамическая модель адаптивной параметризации, включающая набор метрик (скорость улучшения, разнообразие популяции, локальная неоднородность ландшафта поиска, индикатор стагнации), пороговых механизмов и стабилизирующих операторов, которые позволяют обеспечить согласование параметров поиска с текущим состоянием популяции альтернативных решений. Разработана процедура обновления параметров поиска, основанная на интегральном показателе состояния популяции и обеспечивающая регулирование баланса между глобальным и локальным поиском. Интеграция адаптивной модели в эвристические и метаэвристические алгоритмы позволила сформировать двухконтурную организацию поискового процесса с замкнутым контуром параметрического управления.

Результаты вычислительного эксперимента подтвердили эффективность предложенного подхода: интегральный показатель качества размещения улучшен на 8-15% для структур с низкой плотностью связей и на 12-25% для структур с высокой плотностью связей; дисперсия результатов при множественных итерациях снижена в 1,3-2,2 раза; число циклов стагнации сократилось в среднем в 2,5 раза; чувствительность к начальному состоянию популяции снижена (разброс финальных значений F(P) при разных начальных популяциях уменьшился на 30-40%). Временные затраты адаптивного алгоритма превысили время работы статических методов лишь на 5-10% (для

плотных топологий – 441 с и 418 с у статического алгоритма), что обусловлено применением предложенного набора метрик и динамическим обновлением параметров поиска. Полученное время (от 124 с для 100 компонентов до 441 с для 600 компонентов) позволяет использовать разработанный алгоритм в интерактивных средах проектирования СБИС умеренной сложности, а для более крупных схем целесообразно применение иерархических стратегий поиска.

Предложенный механизм адаптивной параметризации формирует основу для дальнейшего развития интеллектуальных методов проектирования СБИС. Перспективы исследований включают интеграцию динамической параметризации с процедурами трассировки, разработку гибридных архитектур с обучением параметров на основе неструктурированных данных, расширение набора метрик состояния популяции, а также применение рассмотренной модели в задачах трёхмерного размещения, оптимизации критических маршрутов и автоматизации процесса многокритериального топологического синтеза.

Литература

1. Гладков Л.А., Кравченко Ю.А., Курейчик В.В., Родзин С.И. Интеллектуальные системы: модели и методы метаэвристической оптимизации: Монография // Чебоксары: ИД «Среда». 2024. 228 с. DOI: 10.31483/a-10639.
2. Qiu Y., Xing Y., Zheng X., Gao P., Cai S., Xiong X. Progress of Placement Optimization for Accelerating VLSI. Physical Design // Electronics. 2023. vol. 12. no. 2. pp. 1–23. DOI: 10.3390/electronics12020337.
3. Курейчик В.В. Вычислительные модели эволюционных и роевых биоэвристик (обзор) / В.В. Курейчик, С.И. Родзин // Информационные технологии. 2021. Т. 27. №10. С. 507–520. DOI: 10.17587/it.27.507-520.
4. Курейчик В.М., Данильченко В.И. Генетический алгоритм планирования размещения СБИС / В.М. Курейчик, В.И. Данильченко // Известия ЮФУ. Технические науки. 2019. №2. С. 26–34.
5. Данильченко В.И., Курейчик В.В. Оценка эффективности и качества проектных решений при размещении фрагментов сверхбольших интегральных схем // Программные продукты и системы. 2025. Т. 38. №4. С. 617–629. DOI: 10.15827/0236-235X.152.617-629.
6. Agnesina A., Chang K., Lim S.K. VLSI Placement Parameter Optimization using Deep Reinforcement Learning // Proceedings of the 2020 International Conference on Computer-Aided Design (ICCAD '20). 2020. pp. 1–9. DOI: 10.1145/3400302.3415690.
7. Карпенко А.П. Современные алгоритмы поисковой оптимизации. Алгоритмы, вдохновленные природой / 3-е изд., испр. // М.: Издательство МГТУ им. Н. Э. Баумана. 2021. 448 с.
8. Курейчик В.В. Вычислительные модели биоэвристик, основанных на физических и когнитивных процессах (обзор) / В.В. Курейчик, С.И. Родзин // Информационные технологии. 2021. Т. 27. №11. С. 563–574. DOI: 10.17587/it.27.563-574.
9. Абас В.М.А. Приближенные методы оптимального размещения элементов электрических и электронных цепей // Известия высших учебных заведений.

Северо-Кавказский регион. Технические науки. 2021. №2. С. 18–26. DOI: 10.17213/1560-3644-2021-2-18-26.
10. Родзин С.И. Биоэвристики: теория, алгоритмы и приложения: Монография / С.И. Родзин, Ю.А. Скобцов, С.А. Эль-Хатиб / Чебоксары: ИД «Среда». 2019. 224 с. DOI: 10.31483/a-54.
11. Курейчик В.М. Математическое обеспечение конструкторского и технологического проектирования с применением САПР / Учеб. для вузов по спец. "Конструирование и технология радиоэлектрон. средств", "Конструирование и технология электрон. вычисл. средств" // Москва: Радио и связь. 1990. 52 с.
12. Sherwani N.A. Algorithms for VLSI Physical Design Automation / New York: Springer Science & Business Media. 2013. 488 p.
13. Амирханов А.В., Волков С.И., Кизиев С.А., Макарчук В.В., Сухов А.Г., Шахнов В.А. Методы оценки качества конструктивных решений, применяемых в СБИС / Международный форум "Микроэлектроника-2016". Тезисы докладов 2-ой научной конференции форума // М.: Рекламно-издательский центр "ТЕХНОСФЕРА". 2016. С. 321–323.
14. Иванов В.К., Думина Д.С., Семенов Н.А. Определение весовых коэффициентов для аддитивной фитнес-функции генетического алгоритма // Программные продукты и системы. 2020. Т. 33. №1. С. 47–53.
15. Plummer J.D., Griffin, P.B. Integrated Circuit Fabrication: Science and Technology // Cambridge University Press. 2023 678 p.
16. Bykovsky N.V., Harutyunyan R.V., Nasedkin A.V. Numerical modeling and optimization results for the placement of irregularly shaped elements on a multidimensional switching field with complex topology // T-Comm. 2024. vol. 18. no. 9. pp. 48–54.
17. de Souza L.A.M., da Silva J.E.H., Chaves L.J., Bernardino H.S. A benchmark suite for designing combinational logic circuits via metaheuristics // Applied Soft Computing. 2020. vol. 91. 106246 p.
18. Govia L.C.G., Jurcevic P., Wood C.J., et al. A randomized benchmarking suite for mid-circuit measurements // New Journal of Physics. 2023. vol. 25. no. 12. pp. 1-28.
19. Гладков Л.А., Лейба С.Н., Тарасов В.Б. Разработка и программная реализация гибридного алгоритма решения оптимизационных задач автоматизированного проектирования // Программные продукты и системы. 2018. Т. 31. №3. С. 569–580. DOI: 10.15827/0236-235X.031.3.569-580.
20. Лобанов А.В., Егоров А.А., Викторова В.В. Использование пакета DEAP для решения задач оптимизации / Актуальные вопросы современных научных исследований: сборник статей IX Международной научно-практической конференции // Пенза: Наука и Просвещение. 2024. С. 63–66.

Курейчик Владимир Викторович — д-р техн. наук, профессор, лауреат премии Правительства РФ в области образования, заведующий кафедры, кафедра систем автоматизированного проектирования (САПР) им. В. М. Курейчика, Южный федеральный университет (ЮФУ). Область научных интересов: искусственный интеллект, автоматизация проектирования, принятие решений, эволюционное моделирование, генетические алгоритмы, биоинспирированный поиск.. Число научных публикаций — 424. vkur@sfedu.ru; ул. Энгельса, 1, 347928, Таганрог, Россия; р.т.: +7(863)438-3451.

Данильченко Владислав Иванович — канд. техн. наук, доцент, доцент, кафедра систем автоматизированного проектирования (САПР) им. В. М. Курейчика, Южный федеральный университет (ЮФУ). Область научных интересов: биоинспирированные

алгоритмы, метаэвристические методы оптимизации, эволюционное моделирование, автоматизация проектирования электронных средств и СБИС, применение методов искусственного интеллекта для решения задач моделирования сложных технических систем. Число научных публикаций — 70. vdanilchenko@sfedu.ru; ул. Энгельса, 1, 347928, Таганрог, Россия; р.т.: +7(863)437-1651.

Данильченко Евгения Владимировна — ассистент, кафедра систем автоматизированного проектирования (САПР) им. В. М. Курейчика, Южный федеральный университет (ЮФУ). Область научных интересов: искусственный интеллект, автоматизация проектирования, принятие решений. Число научных публикаций — 30. lipkina@sfedu.ru; ул. Энгельса, 1, 347928, Таганрог, Россия; р.т.: +7(863)437-1651.

Поддержка исследований. Исследование выполнено за счет гранта Российского научного фонда No 24-71-00035, https://rscf.ru/project/24-71-00035/ в Южном федеральном университете.

DOI 10.15622/ia.25.3.7

V. KUREYCHIK, V. DANILCHENKO, E. DANILCHENKO

ADAPTIVE PARAMETERIZATION OF A METAHEURISTIC ALGORITHM IN VLSI COMPONENT PLACEMENT

Kureychik V., Danilchenko V., Danilchenko E. **Adaptive Parameterization of a Metaheuristic Algorithm in VLSI Component Placement.**

Abstract. The paper considers the problem of parameterization of a metaheuristic algorithm for placing components of very large-scale integrated circuits (VLSI) under conditions of increasingly complex topologies and constraints of modern technological standards. The relevance of the study is due to the growth in the dimensionality of the designed circuits, an increase in the density of interconnections, and the increasing complexity of the design solution space, all of which limit the effectiveness of applied algorithms with static parameters. The problem statement includes the formation of a set of design metrics, such as wire length, density, placement area, crossing index, and others. Fixed parameterization of metaheuristic algorithms does not fully account for the dynamics of the search space, which limits the quality of the design solutions made. Based on the analysis of the set of design metrics, a dynamic model is proposed that reflects the current state of the search process, including the level of structural diversity of the population and signs of stagnation. Taking into account the proposed model, the structure of an adaptive metaheuristic algorithm with a closed-loop parameter control loop is developed, where parameters are adjusted to coordinate the intensity of global and local search. The scientific novelty lies in the original configuration of a set of metrics (objective function improvement rate δt, population diversity Dt, local landscape heterogeneity Ht, stagnation indicator St) and the method of their integration into the parameter control loop. In contrast to resource-intensive neural network architectures that require significant computational costs and training samples, the proposed analytical model ensures the interpretability of the adaptation process and high computational efficiency. To confirm the effectiveness of the approach, a software subsystem was built that implements the simulation and computational experiment process. The results of the experiments on test topologies of various dimensions and connection densities showed that, compared to algorithms with static parameterization and fixed multi-stage tuning, the application of the developed algorithm made it possible to increase the integral placement quality metric by 8-15% for structures with low connection density and by 12-25% for structures with high connection density, reduce the variance of results over multiple runs by a factor of 1.3-2.2, reduce the number of stagnation cycles, and decrease sensitivity to the initial state of the population. The obtained results confirm the improvement in the quality and efficiency of design solutions, the robustness of the developed placement algorithm, and indicate the practical applicability of the proposed approach. The article will be useful for specialists engaged in design automation and intelligent analysis in VLSI development.

Keywords: metaheuristic algorithms, adaptive parameterization, dynamic model, VLSI component placement, combinatorial optimization, search metrics, topological synthesis, multidimensional discrete space, intelligent design systems.

References

1. Gladkov L.A., Kravchenko Yu.A., Kureichik V.V., Rodzin S.I. [Intelligent Systems: Models and Methods of Metaheuristic Optimization: Monograph]. Cheboksary: ID "Sreda", 2024. 228 p. DOI: 10.31483/a-10639. (In Russ.).

2. Qiu Y., Xing Y., Zheng X., Gao P., Cai S., Xiong X. Progress of Placement Optimization for Accelerating VLSI. Physical Design. Electronics. 2023. vol. 12. no. 2. pp. 1–23. DOI: 10.3390/electronics12020337.
3. Kureichik V.V. [Computational Models of Evolutionary and Swarm Bio-Heuristics (Review)]. V.V. Kureichik, S.I. Rodzin. Informatsionnyye tekhnologii – Information Technologies. 2021. vol. 27. no. 10. pp. 507–520. DOI: 10.17587/it.27.507-520. (In Russ.).
4. Kureichik V.M., Danilchenko V.I. [Genetic Algorithm for Planning the Placement of VLSI]. V.M. Kureichik, V.I. Danilchenko. Izvestiya YuFU. Tekhnicheskiye nauki – Izvestiya SFU. Technical Sciences. 2019. no. 2. pp. 26–34. (In Russ.).
5. Danilchenko V.I., Kureichik V.V. [Evaluation of the Effectiveness and Quality of Design Solutions for Placing Fragments of Ultra-Large-Scale Integrated Circuits]. Programmnyye produkty i sistemy – Software Products and Systems. 2025. vol. 38. no. 4. pp. 617–629. DOI: 10.15827/0236-235X.152.617-629. (In Russ.).
6. Agnesina A., Chang K., Lim S.K. VLSI Placement Parameter Optimization using Deep Reinforcement Learning. Proceedings of the 2020 International Conference on Computer-Aided Design (ICCAD '20). 2020. pp. 1–9. DOI: 10.1145/3400302.3415690.
7. Karpenko A.P. [Modern Search Optimization Algorithms. Nature-Inspired Algorithms]. 3rd ed., revised. Moscow: Izdatel'stvo MGTU im. N. E. Baumana, 2021. 448 p. (In Russ.).
8. Kureichik V.V. [Computational Models of Bio-Heuristics Based on Physical and Cognitive Processes (Review)]. V.V. Kureichik, S.I. Rodzin. Informatsionnyye tekhnologii – Information Technologies. 2021. vol. 27. no. 11. pp. 563–574. DOI: 10.17587/it.27.563-574. (In Russ.).
9. Abas V.M.A. [Approximate Methods of Optimal Placement of Elements in Electrical and Electronic Circuits]. Izvestiya vysshikh uchebnykh zavedeniy. Severo-Kavkazskiy region. Tekhnicheskiye nauki – Proceedings of Higher Educational Institutions. North Caucasus Region. Technical Sciences. 2021. no. 2. pp. 18–26. DOI: 10.17213/1560-3644-2021-2-18-26. (In Russ.).
10. Rodzin S.I. [Bio-Heuristics: Theory, Algorithms, and Applications: Monograph]. S.I. Rodzin, Yu.A. Skobtsov, S.A. El-Khatib. Cheboksary: ID "Sreda", 2019. 224 p. DOI: 10.31483/a-54. (In Russ.).
11. Kureichik V.M. [Mathematical support for design and technological design using CAD]. Ucheb. dlya vuzov po spets. "Konstruirovaniye i tekhnologiya radioelektron. sredstv", "Konstruirovaniye i tekhnologiya elektron. vychisl. sredstv". Moscow: Radio i svyaz', 1990. 52 p. (In Russ.).
12. Sherwani N.A. Algorithms for VLSI Physical Design Automation. New York: Springer Science & Business Media. 2013. 488 p.
13. Amirkhanov A.V., Volkov S.I., Kiziev S.A., Makarchuk V.V., Sukhov A.G., Shakhnov V.A. [Methods for Assessing the Quality of Design Solutions Used in Vlsi]. Mezhdunarodnyy forum «Mikroelektronika-2016»: Tezisy dokladov 2-oy nauchnoy konferentsii foruma* [International Forum «Microelectronics-2016»: Abstracts of the 2nd Scientific Conference]. Moscow: Reklamno-izdatel'skiy tsentr «TEKHNOSFERA». 2016. pp. 321–323. (In Russ.)
14. Ivanov V.K., Dumina D.S., and Semenov N.A. [Determining Weight Coefficients for the Additive Fitness Function of a Genetic Algorithm]. Programmnyye produkty i sistemy – Software Products and Systems. 2020. vol. 33. no. 1. pp. 47–53. (In Russ.).
15. Plummer J.D., Griffin, P.B. Integrated Circuit Fabrication: Science and Technology. Cambridge University Press. 2023 678 p.
16. Bykovsky N.V., Harutyunyan R.V., Nasedkin A.V. Numerical modeling and optimization results for the placement of irregularly shaped elements on a

multidimensional switching field with complex topology. T-Comm. 2024. vol. 18. no. 9. pp. 48–54.
17. de Souza L.A.M., da Silva J.E.H., Chaves L.J., Bernardino H.S. A benchmark suite for designing combinational logic circuits via metaheuristics. Applied Soft Computing. 2020. vol. 91. 106246 p.
18. Govia L.C.G., Jurcevic P., Wood C.J., et al. A randomized benchmarking suite for mid-circuit measurements. New Journal of Physics. 2023. vol. 25. no. 12. pp. 1-28.
19. Gladkov L.A., Leiba S.N., Tarasov V.B. [Development and Software Implementation of a Hybrid Algorithm for Solving Optimization Problems in Automated Design]. Programmnyye produkty i sistemy – Software Products and Systems. 2018. vol. 31. no. 3. pp. 569–580. DOI: 10.15827/0236-235X.031.3.569-580. (In Russ.).
20. Lobanov A.V., Egorov A.A., Viktorova V.V. [Using the DEAP package to solve optimization problems]. Aktual'nyye voprosy sovremennykh nauchnykh issledovaniy: sbornik statey IX Mezhdunarodnoy nauchno-prakticheskoy konferentsii [Actual Issues of Modern Scientific Research: Collection of Articles from the IX International Scientific and Practical Conference]. Penza: Nauka i Prosveshcheniye. 2024. pp. 63–66. (In Russ.).

Kureychik Vladimir — Ph.D., Dr. Sci., Professor, Laureate of the Government of the Russian Federation Award in the Field of Education, Head of the Department, Department of Computer-Aided Design Systems (CAD) named after V. M. Kureichik, Southern Federal University (SFU). Research interests: artificial intelligence, design automation, decision-making, evolutionary modeling, genetic algorithms, and bio-inspired search. The number of publications — 424. vkur@sfedu.ru; 1, Engels St., 347928, Taganrog, Russia; office phone: +7(863)438-3451.

Danilchenko Vladislav — Ph.D., Associate professor, Department of Computer-Aided Design Systems (CAD) named after V. M. Kureichik, Southern Federal University (SFU). Research interests: bio-inspired algorithms, metaheuristic optimization methods, evolutionary modeling, automation of electronic and VLSI design, and the application of artificial intelligence methods for solving complex technical systems modeling problems. The number of publications — 70. vdanilchenko@sfedu.ru; 1, Engels St., 347928, Taganrog, Russia; office phone: +7(863)437-1651.

Danilchenko Evgeniya — Assistant, Department of Computer-Aided Design Systems (CAD) named after V. M. Kureichik, Southern Federal University (SFU). Research interests: artificial intelligence, design automation, and decision-making. The number of publications — 30. lipkina@sfedu.ru; 1, Engels St., 347928, Taganrog, Russia; office phone: +7(863)437-1651.

Acknowledgements. The research was funded by the Russian Science Foundation project No. 24-71-00035, https://rscf.ru/project/24-71-00035/ implemented by the Southern Federal University.

УДК 004.056 **DOI** 10.15622/ia.25.3.8

И.В. Котенко, И.Б. Саенко, Е.С. Митяков

ИСПОЛЬЗОВАНИЕ ЦИФРОВОГО ДВОЙНИКА ДЛЯ ОБНАРУЖЕНИЯ КИБЕРУГРОЗ В СИСТЕМАХ УПРАВЛЕНИЯ ИНТЕЛЛЕКТУАЛЬНЫМИ ЭНЕРГОСЕТЯМИ

Котенко И.В., Саенко И.Б., Митяков Е.С. **Использование цифрового двойника для обнаружения киберугроз в системах управления интеллектуальными энергосетями.**

Аннотация. Цифровая трансформация энергетического сектора, сопровождающаяся повсеместным внедрением интеллектуальных сетей, расширяет поверхность кибератак и повышает уязвимость критически важной инфраструктуры. Традиционные системы безопасности, основанные на сигнатурах, статических правилах, а также на текущих реализациях методов машинного обучения, демонстрируют ограниченную эффективность против новых типов угроз, а их верификация на реальных объектах сопряжена с неприемлемыми рисками. В статье представлен и экспериментально апробирован интегрированный адаптивный фреймворк на основе цифрового двойника (ЦД), направленный на повышение результативности обнаружения киберугроз в системах управления интеллектуальными энергосетями. Фреймворк включает формализованную математическую постановку задачи обнаружения с учетом трех ключевых показателей: полноты (Recall), доли ложных срабатываний (FPR) и совокупного времени реакции (TTR), а также целевую функцию для их совместной оптимизации. Ключевым элементом фреймворка является контур адаптации с обратной связью по метрикам качества, управляемый трехуровневым механизмом (оперативный, тактический, стратегический режимы): результаты работы системы обнаружения угроз на реальном объекте используются для автоматического запуска процедур дообучения, оптимизации или генерации новых сценариев в безопасной виртуальной среде ЦД. Цифровой двойник формализован как кортеж взаимосвязанных моделей физических процессов, коммуникационной инфраструктуры, логики управления и базы знаний об угрозах, что позволяет учитывать многоуровневую архитектуру интеллектуальных энергосетей и специфические протоколы (МЭК 61850, Modbus). Экспериментальная валидация на синтетическом наборе данных, имитирующем работу АСУ ТП, подтвердила применимость концепции: использование синтетических данных, генерируемых в среде ЦД, позволило снизить частоту ложных срабатываний на 6,4% по сравнению с обучением на статичных данных, а активация адаптационного механизма способствовала сокращению совокупного модельного времени реакции на инциденты на 3,82%. Расчет комплексной целевой функции показал снижение обобщенного показателя качества на 16,8% после адаптации. Предложенный подход вносит вклад в решение проблем дефицита представительных данных и ограничений безопасного тестирования и может рассматриваться как шаг к реализации парадигмы «безопасность через проектирование» в энергетике. Полученные результаты представляют собой доказательство реализуемости концепции и обосновывают необходимость дальнейших исследований для промышленного внедрения, включая проверку на реальных промышленных сценариях.

Ключевые слова: интеллектуальные энергосети, цифровой двойник, адаптивное обнаружение угроз, киберустойчивость, системы управления.

1. Введение. Энергетический сектор, являясь фундаментом национальной безопасности и экономического благополучия,

переживает беспрецедентную цифровую трансформацию. На смену традиционным энергосистемам приходят интеллектуальные сети – сложные киберфизические системы (КФС), интегрирующие физические процессы генерации, передачи и распределения электроэнергии с цифровыми технологиями управления, мониторинга и связи [1]. Эта интеграция открывает огромные возможности для повышения эффективности, надежности и устойчивости энергоснабжения. Однако она же кардинально расширяет векторы киберугроз, превращая энергетическую инфраструктуру в привлекательную и критически важную мишень для злоумышленников [2, 3].

Объекты критической информационной инфраструктуры (КИИ), к которым относятся интеллектуальные энергосети, сталкиваются с эволюционирующими, целенаправленными и высоко скрытными атаками [4, 5]. Традиционные системы информационной безопасности, основанные на сигнатурном анализе и статических правилах, демонстрируют ограниченную эффективность против таких угроз, особенно в части своевременного выявления ранних признаков компрометации [6, 7]. Эти индикаторы часто статистически неотличимы от редких, но легитимных событий или технических сбоев, что приводит к высокому уровню ложных срабатываний или, что еще хуже, к пропуску реальных атак [8].

Системы управления интеллектуальными энергосетями представляют собой многоуровневые киберфизические комплексы, включающие полевой уровень (датчики, исполнительные механизмы, PMU), уровень контроллеров (ПЛК, RTU) и диспетчерский уровень (SCADA, EMS). На каждом уровне используются специфические протоколы (Modbus, IEC 61850, DNP3) и существуют характерные векторы кибератак (инъекции ложных данных, подмена управляющих команд на ПЛК, компрометация систем сбора и отображения данных и т.д.). Указанные особенности предъявляют повышенные требования к средствам обнаружения – необходимо учитывать временные ограничения реального времени и каскадный характер распространения атак и аномальных возмущений (сбоев, отказов и др.) [9].

Разработка и валидация современных методов обнаружения угроз ИБ для энергосистем сталкивается с рядом фундаментальных проблем [10, 11]:

1. Дефицит репрезентативных данных. Получение реальных, размеченных данных о кибератаках на действующих объектах КИИ крайне затруднено из-за конфиденциальности, этических и юридических ограничений, а также самой природы атак, которые стремятся остаться незамеченными.

2. Ограничения безопасного тестирования. Проведение экспериментов по моделированию атак на реальных энергетических объектах связано с неприемлемо высокими операционными и экономическими рисками, включая потенциальные сбои в энергоснабжении.

3. Динамичность угроз. Тактики, техники и процедуры (TTP) злоумышленников постоянно эволюционируют, требуя от систем защиты не статичности, а способности к непрерывной адаптации и самообучению.

Для операторов КИИ эффективность систем обнаружения угроз ИБ традиционно оценивается по трем ключевым показателям [12]:

1. Точность прогнозирования, т.е. способность системы корректно выявлять все реальные угрозы, минимизируя количество пропущенных атак (ложноотрицательных срабатываний). В терминах метрик классификации данному критерию соответствует полнота (Recall).

2. Снижение частоты ложных срабатываний, поскольку высокий уровень ложных тревог приводит к «усталости от оповещений» у персонала центров мониторинга безопасности, отвлекая их от реальных инцидентов и увеличивая операционные издержки. В количественном выражении это показатель доли ложных срабатываний (FPR).

3. Сокращение времени реагирования. Чем быстрее система позволяет обнаружить и классифицировать угрозу, тем быстрее могут быть инициированы контрмеры, что напрямую снижает потенциальный ущерб. В работе используется совокупное время реагирования (TTR), зависящее от количества истинных и ложных срабатываний.

Указанные показатели образуют компромиссную систему: стремление к максимальному Recall может увеличить FPR, а снижение FPR иногда достигается за счет потери Recall. TTR зависит от обоих. Поэтому в экспериментальной части работы оценивается изменение всех трех показателей совместно, что позволяет судить о комплексной эффективности адаптации.

В этом контексте концепция цифрового двойника (ЦД) – виртуальной, динамически обновляемой реплики физического объекта или системы – приобретает особую значимость как потенциальное решение данных проблем [13, 14]. Исследования последних лет подтверждают, что интеграция ЦД с технологиями искусственного интеллекта и машинного обучения позволяет создавать безопасные среды для генерации синтетических данных, моделирования сложных сценариев кибератак, обучения и тестирования моделей обнаружения аномалий без риска для реальной инфраструктуры [15 – 17]. Такой

подход способствует реализации парадигмы «безопасность через проектирование» (security-by-design) и обеспечивает возможность выявления даже ранее неизвестных (zero-day) угроз [18].

Несмотря на значительный прогресс, существующие решения на основе ЦД для энергетики часто фокусируются либо на моделировании физических процессов и атак [19, 20], либо на обнаружении аномалий [21, 22], но редко предлагают замкнутый, саморегулирующийся цикл, где результаты работы системы на реальном объекте автоматически используются для ее совершенствования в виртуальной среде. Особенно остро стоит проблема создания адаптивных механизмов, способных не только реагировать на новые угрозы, но и проактивно обновлять защитные модели на основе анализа их эффективности в реальном времени.

Цель данной статьи – представить и экспериментально валидировать интегрированный адаптивный фреймворк на основе ЦД для обнаружения киберугроз в системах управления интеллектуальными энергосетями, целенаправленно оптимизирующий три ключевых показателя результативности: полноту обнаружения (Recall), частоту ложных срабатываний (FPR) и совокупное время реагирования на инциденты (TTR). Этот фреймворк синергетически объединяет методологию адаптивного обнаружения угроз ИБ с архитектурой прототипа ЦД для АСУ энергосистемы.

Основной научный вклад работы заключается в следующем:

1. Предложение интегрированного фреймворка, объединяющего архитектуру ЦД для энергосистемы и метод адаптивного обнаружения угроз в единую систему, ориентированную на повышение Recall, снижение FPR и сокращение совокупного времени реагирования на инциденты.
2. Формализация и реализация трехуровневого адаптационного механизма (оперативный, тактический, стратегический режимы) как ключевого элемента фреймворка, обеспечивающего непрерывное поддержание актуальности и эффективности моделей защиты по всем трем критериям.
3. Экспериментальная валидация фреймворка, включающая:

– демонстрацию эффективности адаптационного механизма на модели АСУ ТП энергосистемы.

– подтверждение пользы использования синтетических данных, генерируемых в среде ЦД, для повышения качества обнаружения угроз ИБ, что косвенно способствует сокращению времени реагирования за счет повышения достоверности срабатываний.

– обсуждение практической применимости и ограничений предложенного подхода для промышленного внедрения.

2. Анализ релевантных работ. Научное сообщество активно исследует потенциал ЦД как инструмента для повышения киберустойчивости критически важных инфраструктур, включая энергетические системы. Обзор релевантных работ структурирован вокруг трех взаимосвязанных направлений: применение ЦД в энергетике, их использование для задач кибербезопасности и современные методы адаптивного обнаружения аномалий.

ЦД становятся ключевым инструментом цифровой трансформации энергетического сектора, обеспечивая виртуальное отражение физических активов и процессов в реальном времени. Эта технология позволяет не просто наблюдать за системой, но и прогнозировать ее поведение, оптимизировать работу и интегрировать возобновляемые источники энергии, что напрямую способствует повышению эффективности и устойчивости всей энергетической цепочки [23, 24].

Применение ЦД охватывает все этапы жизненного цикла энергии – от генерации и хранения до передачи, распределения и конечного потребления [25]. В ветро- и солнечной энергетике они повышают производительность и надежность установок [26, 27], в системах хранения – улучшают управление батареями [28], а в традиционных секторах, таких как нефтегаз, – обеспечивают контроль целостности активов и управление жизненным циклом [29]. Основные функции включают мониторинг в реальном времени, предиктивное обслуживание, оптимизацию сетей и прогнозирование отказов [30, 31].

Экономический и экологический эффект от внедрения ЦД подтверждается многочисленными исследованиями: снижение операционных затрат, экономия энергии до 30% и ускорение перехода к «чистой» энергетике за счет более точного и основанного на данных управления [32, 33]. ЦД становятся катализатором устойчивого развития, помогая минимизировать экологический след и двигаться к целям углеродной нейтральности [34].

Однако путь к массовому внедрению не лишен препятствий. Высокая стоимость реализации, сложности интеграции с унаследованными системами, отсутствие единых стандартов и, что особенно актуально для нашей работы, недостаточная проработка вопросов кибербезопасности остаются ключевыми барьерами [35, 36]. Будущее за созданием масштабируемых и совместимых архитектур,

а также за глубокой интеграцией с технологиями ИИ, 5G и IoT для построения автономных и самовосстанавливающихся энергосистем [37, 38].

ЦД все чаще рассматриваются не просто как инструмент оптимизации, но и как мощный актив в арсенале киберзащиты. Виртуальная реплика реальной системы превращается в безопасную «песочницу», где можно моделировать самые изощренные атаки, не рискуя нарушить работу КИИ [39, 40]. Это особенно ценно для энергетики, где любой сбой может иметь каскадные последствия. Ключевая ценность ЦД для ИБ проявляется в трех основных направлениях:

1. Моделирование и тестирование. ЦД позволяют создавать высокоточные модели киберфизических систем, на которых можно безопасно отрабатывать сценарии атак, тестировать защитные меры и оценивать уязвимости. Это позволяет выявлять слабые места до того, как они будут найдены злоумышленниками [41, 42].

2. Проактивная защита. Благодаря зеркальному отображению реальных процессов, системы на основе ЦД способны выявлять аномалии на ранних стадиях, что позволяет операторам оперативно реагировать и минимизировать потенциальный ущерб [43, 44].

3. Повышение эффективности обнаружения. ЦД могут использоваться как платформы для обмана, эффективно отвлекая злоумышленников и собирая ценные данные об их тактиках. В сочетании с методами объяснимого ИИ (XAI) это позволяет не просто обнаруживать угрозы, но и понимать их природу [45, 46].

Вместе с тем сама технология ЦД создает новые векторы атак. Интеграция с IoT и облачными сервисами расширяет поверхность атак, вызывая обеспокоенность по поводу целостности данных и возможности манипуляции системой [47, 48]. Отсутствие единых стандартов безопасности усложняет развертывание защищенных решений, особенно в сложных, гетерогенных средах [49]. Критически важным также является обеспечение точной и безопасной синхронизации между цифровой моделью и ее физическим прототипом, поскольку любые рассогласования могут быть использованы злоумышленниками [40, 45].

Перспективным направлением считается интеграция ЦД с передовыми технологиями. Например, сочетание с ИИ позволяет автоматизировать обнаружение угроз, а использование блокчейна – обеспечить прозрачность и контроль доступа [13, 50]. Не менее важен и принцип «безопасность через проектирование» (security-by-design), который предполагает внедрение механизмов защиты, таких как

модели нулевого доверия и непрерывный мониторинг, уже на этапе проектирования архитектуры ЦД [39, 47].

Таким образом, хотя ЦД открывают новые горизонты для кибербезопасности, их эффективное и безопасное применение требует комплексного подхода, учитывающего как возможности, так и присущие им риски.

Быстрая цифровизация промышленных систем управления и интеллектуальных сетей неизбежно влечет за собой рост уязвимостей. В этих условиях статические системы защиты теряют актуальность, уступая место адаптивным методам обнаружения аномалий, способным учиться на новых данных и подстраиваться под меняющиеся условия работы и тактики злоумышленников. Такие методы становятся критически важными для обеспечения безопасности и надежности энергетической инфраструктуры.

Современный арсенал адаптивных подходов разнообразен и активно развивается. Рассмотрим некоторые из них:

1. Глубокое обучение показывает высокую точность в выявлении сложных, многоэтапных атак. Архитектуры, такие как адаптивные остаточные рекуррентные сети (ARRNN-DGRU) и сверточные LSTM (ConvLSTM), превосходят традиционные модели по скорости и надежности детектирования, особенно в условиях зашумленных данных [51 – 53].

2. Методы распределенного обучения и обучения с сохранением конфиденциальности решают проблему защиты данных в децентрализованных системах. Эти подходы позволяют обучать модели непосредственно на локальных устройствах – например, на «умных» счетчиках или контроллерах без необходимости передачи чувствительной информации на центральный сервер, сохраняя при этом высокую эффективность обнаружения [54 – 56]. Такие решения особенно актуальны для ресурсоограниченных сред и способствуют повышению доверия со стороны пользователей и операторов системы.

3. Модели, устойчивые к изменению в данных и интерпретируемые модели решают проблему ложных срабатываний, вызванных естественными изменениями в работе системы (например, сменой привычек потребления). Используя такие фреймворки, как LSTM или системы на основе правил (AI-BRB), эти методы не только адаптируются к новым условиям, но и объясняют причины своих решений, что критически важно для операторов [57, 58].

4. Статистические и физико-математические методы обеспечивают масштабируемое и быстрое обнаружение в реальном времени. Подходы, основанные на расширенных фильтрах Калмана

или гибридных AI-физических моделях, эффективно различают технические сбои, внешние возмущения и целенаправленные кибератаки, опираясь на фундаментальные законы физики энергосистем [59 – 61].

Таким образом, адаптивные методы обнаружения аномалий формируют важную парадигму для современных энергосистем. Их способность к самообучению и адаптации позволяет операторам не просто реагировать на угрозы, а опережать их. Однако для реализации этого потенциала требуется безопасная и управляемая среда для непрерывного обучения и верификации моделей – именно ту нишу и занимает технология ЦД.

Проведенный анализ подтверждает, что области применения ЦД в энергетике, использования ЦД для кибербезопасности и разработки адаптивных методов обнаружения аномалий развиваются стремительно, однако остаются в значительной степени изолированными. Существующие решения либо используют ЦД как пассивную среду для симуляции, либо применяют адаптивные алгоритмы без механизма непрерывного обновления на основе обратной связи от реальной системы. Ключевая научная новизна и практическая ценность нашей работы заключаются в создании интегрированного, саморегулирующегося фреймворка, в котором ЦД энергосистемы выступает не просто виртуальной копией, а динамической обучающей платформой. Предложенный замкнутый адаптивный цикл, управляемый трехуровневым механизмом (оперативный, тактический, стратегический), обеспечивает непрерывную верификацию и актуализацию моделей обнаружения угроз на основе реальных данных. Именно эта архитектура позволяет напрямую и системно оптимизировать три критических показателя эффективности: точность прогнозирования угроз, снижение частоты ложных срабатываний и сокращение времени реагирования на инциденты, что является конечной целью для операторов критически важной энергетической инфраструктуры.

3. Предлагаемый фреймворк. Фреймворк представляет собой интегрированную систему для повышения результативности обнаружения киберугроз в интеллектуальных энергосетях, основанную на концепции цифрового двойника и механизме адаптивного обнаружения угроз информационной безопасности. Отличительной особенностью фреймворка выступает формализация и автоматизация цикла адаптации, в котором обратная связь от реальной системы служит триггером для запуска конкретных

процедур в виртуальной среде: дообучения модели, оптимизации ее параметров или генерации новых сценариев.

Математическая постановка задачи. Пусть $X = \{x_t\}$ – временной ряд наблюдаемых параметров (например, напряжение, ток). Каждому наблюдению соответствует неизвестное состояние $y_t \in \{\text{норма}, \text{сбой}, \text{атака}\}$. Требуется построить классификатор $f(X, \Theta)$, где Θ – параметры модели, который по наблюдениям определяет состояние.

Качество классификации оценивается метриками *Recall* (полнота), *Precision* (точность), *FPR* (доля ложных срабатываний). Совокупное время реагирования *TTR* зависит от количества истинных и ложных срабатываний. Адаптация заключается в обновлении Θ на основе обратной связи от реальной системы. Цель адаптации – минимизация взвешенной суммы потерь:

$$L = \lambda_1 \cdot (1 - Recall) + \lambda_2 \cdot FPR + \lambda_3 \cdot \frac{TTR}{T_{\max}} \rightarrow \min, \quad (1)$$

где $\lambda_1, \lambda_2, \lambda_3$ – весовые коэффициенты, отражающие приоритеты при защите (например, $\lambda_1 > \lambda_2$ при критичности пропуска атаки), а $T_{\max}$ – максимально допустимое время реагирования.

Для решения сформулированной задачи разработан на основе ЦД, архитектура которого представлена далее.

Архитектура фреймворка построена на взаимодействии двух функциональных контуров: ЦД и реального объекта (АСУ энергосетью), связанных через центральный адаптационный механизм (Рис. 1) [62].

Логика работы фреймворка определяется следующим потоком данных и управления:

1. Инициализация в контуре ЦД. Процесс начинается с генерации синтетических наборов данных, отражающих нормальные, аварийные и атакующие сценарии. Эти данные используются для обучения и верификации моделей обнаружения аномалий.

2. Развертывание на реальном объекте. Верифицированные модели передаются и интегрируются в системы мониторинга и управления на всех уровнях АСУ энергосети.

3. Мониторинг и обнаружение. Модели анализируют потоки операционных данных в реальном времени, выявляя отклонения и классифицируя события.

4. Формирование обратной связи. Агрегированные данные о работе моделей, включая метрики качества (F1-score, *Recall*, *FPR*) и информацию о срабатываниях, передаются обратно в контур ЦД.

5. Адаптация. Адаптационный механизм анализирует поступившую обратную связь. При отклонении метрик за предустановленные пороги инициируется один из трех режимов адаптации (оперативный, тактический, стратегический), что приводит к обновлению моделей или сценариев в ЦД. После этого цикл повторяется.

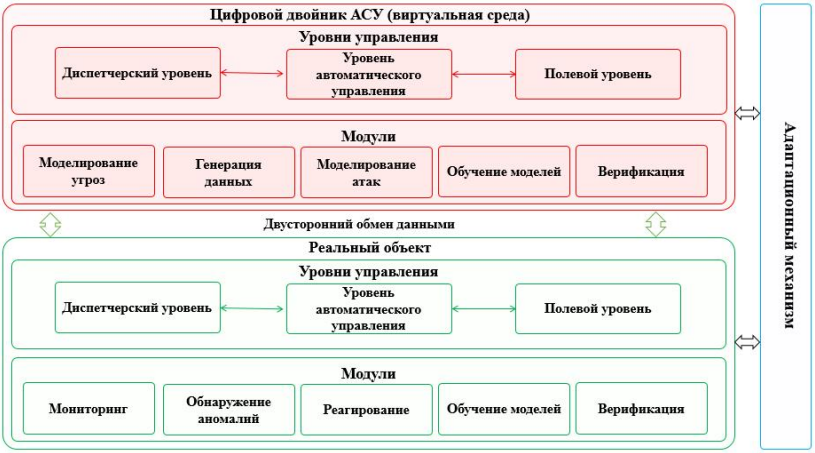


Рис. 1. Общая архитектура адаптивного фреймворка на основе ЦД

Такая архитектура создает предпосылки для поддержания актуальности механизмов защиты без необходимости прямого вмешательства в работу критически важной инфраструктуры.

Архитектура ЦД декомпозирована на восемь специализированных функциональных модулей, каждый из которых играет определенную роль в реализации замкнутого адаптивного цикла (таблица 1). В совокупности эти модули образуют единую платформу, способную не только имитировать работу энергосистемы, но и обеспечивать непрерывное обучение и адаптацию моделей обнаружения угроз, что способствует поддержанию их актуальности в условиях изменяющейся среды.

Для формализации предложенного подхода определим ЦД как кортеж взаимосвязанных моделей:

$$DT = \langle M_{phys}, M_{comm}, M_{cyber}, M_{threat}, Sync \rangle,$$

где M_{phys} – модель физических процессов энергосистемы, параметризуемая вектором настраиваемых параметров p_{phys} (в общем случае может быть задана системой дифференциальных уравнений, описывающих переходные процессы в линиях, генераторах, нагрузках); M_{comm} – модель коммуникационной инфраструктуры, включающая протоколы, топологию сети, задержки и нарушения связи; M_{cyber} – модель логики управления (алгоритмы ПЛК, SCADA, противоаварийной автоматики), представленная конечным автоматом или набором правил; M_{threat} – база знаний об уязвимостях, угрозах и сценариях атак (TTP), структурированная, например, в формате MITRE ATT&CK для промышленных систем; $Sync$ – механизм синхронизации, обеспечивающий согласование состояния ЦД с реальным объектом путем обновления параметров моделей по данным телеметрии.

В рамках экспериментального прототипа, описанного в разделе 4, реализована упрощенная версия ЦД, достаточная для доказательства реализуемости концепции. Модель M_{phys} представлена параметрической генерацией временного ряда напряжения; модели M_{comm} и на данном этапе не задействованы, а M_{threat} применяется на уровне генерации сценариев атак (импульсные искажения, фазовые сдвиги, низкочастотная модуляция). Предложенная архитектура ЦД (таблица 1) является расширяемой: компоненты могут заменяться на более сложные альтернативы (например, M_{phys} на нейросетевую модель или полную систему дифференциальных уравнений, M_{comm} на эмулятор сети) без изменения общей структуры фреймворка. Параметры p_{phys} калибруются по историческим данным реального объекта, их обновление осуществляется при поступлении новой информации в процессе адаптации.

Представленная архитектура ЦД реализована программно в соответствии с модульной структурой таблицы 1. Конкретные алгоритмы генерации синтетических данных (включая параметры сигналов нормального, аварийного и атакующего режимов), методы предобработки (вейвлет-преобразование с базисом Добеши db1 и нормализация Min-Max), а также процедуры обучения и адаптации модели обнаружения на основе Isolation Forest описаны далее. Таким образом, ЦД раскрыт не на уровне общих требований, а как конкретная реализация, параметры которой обеспечивают воспроизводимость экспериментов и позволяют рассматривать его как рабочий прототип.

Таблица 1. Роль модулей ЦД в реализации адаптивного фреймворка

№	Модуль	Описание
1	Модуль киберфизического моделирования	Разворачиваются все процессы моделирования, симуляции и адаптации. Его точность напрямую влияет на релевантность генерируемых сценариев и, следовательно, на качество обучения моделей.
2	Модуль моделирования угроз ИБ	Формирует и управляет базой знаний об уязвимостях, TTP-сценариях (например, на основе MITRE ATT&CK) и последствиях атак. В стратегическом режиме адаптации именно этот модуль обновляется на основе анализа новых угроз, выявленных на реальном объекте.
3	Модуль генерации синтетических данных	Генерирует обучающие и тестовые выборки, включая сложные, ранее не встречавшиеся паттерны, идентифицированные как недостаточно покрытые текущими моделями. Это ключевой компонент для оперативного и стратегического режимов адаптации.
4	Модуль симуляции атак	Внедряет сгенерированные атаки в киберфизическую модель для тестирования их реализуемости и оценки воздействия. Позволяет безопасно "проиграть" новые сценарии до их потенциального появления в реальной системе.
5	Модуль обнаружения аномалий	Выполняет детектирование угроз как в контуре ЦД (на этапе тестирования), так и на реальном объекте. Его выходные данные (метрики Precision, Recall, FPR) являются основными триггерами для запуска адаптационного механизма.
6	Модуль агрегации данных	Обеспечивает единый доступ ко всем данным – синтетическим, историческим и потоковым. Критически важен для процессов дообучения (оперативный режим) и переобучения (стратегический режим).
7	Модуль оценки угроз	Интерпретирует результаты обнаружения, оценивая угрозы по критериям вероятности, критичности и потенциального ущерба. Его выводы могут использоваться для приоритезации сценариев, которые должны быть сгенерированы в стратегическом режиме адаптации.
8	Модуль верификации	Оценивает корректность работы всей системы. Его отчеты о снижении эффективности моделей или недостаточной реалистичности сценариев являются формальным основанием для запуска адаптации, особенно в тактическом и стратегическом режимах.

Метод адаптивного обнаружения. Фреймворк реализует метод адаптивного обнаружения, который формализует процесс обновления моделей на основе анализа их эффективности в реальных условиях эксплуатации. Методика разделяет процессы, происходящие в контуре ЦД и на реальном объекте, и определяет условия для их взаимодействия через адаптационный механизм [63].

Этапы в контуре ЦД:

1. Моделирование сценариев и генерация данных. Создаются синтетические наборы, включающие нормальные, аварийные и атакующие режимы.

2. Предобработка данных. Применяются методы снижения размерности (PCA) и шумоподавления (вейвлет-преобразование, например, с базисом Добеши) для повышения качества входных данных.

3. Обучение моделей. Используются алгоритмы машинного обучения, такие как Isolation Forest, для построения моделей классификации событий на «норму», «сбой» или «угрозу».

4. Тестирование моделей. Оцениваются метрики качества (F1-score, *Recall, Precision, FPR*) на независимых тестовых выборках.

5. Внедрение моделей. Верифицированные модели передаются в контур реального объекта.

Этапы в контуре реального объекта:

1. Развертывание моделей. Алгоритмы интегрируются в программно-аппаратные средства контроля на объекте.

2. Мониторинг операционных данных. Осуществляется непрерывный сбор телеметрии, логов событий и параметров технологических процессов.

3. Обнаружение и классификация в реальном времени. Модели анализируют поступающие данные, выявляя аномалии и классифицируя события.

4. Реагирование. При обнаружении угрозы инициируются заранее определенные меры (оповещение, изоляция сегмента сети и т.д.).

5. Обратная связь и анализ. Агрегированные данные о работе моделей передаются обратно в контур ЦД для анализа.

Адаптация инициируется автоматически, когда качество работы модели на реальном объекте опускается ниже заданного порога. Формально, если модель вычисляет набор показателей качества $Q = \{q_1, q_2, \dots, q_n\}$, то адаптация запускается, если выполняется условие:

$$\exists q_i \in Q: q_i \notin [q_{imin}, q_{imax}]. \quad (2)$$

В зависимости от причины и характера отклонений выделяются три режима адаптации [63]:

1. Оперативный режим. В рамках данного режима осуществляется дообучение моделей на новых данных без изменения их архитектуры и параметров. Формально, при поступлении нового набора данных $\mathrm{D_{new}}$, обновляются параметры модели $\theta \to \theta'$ методом дообучения:

$$\theta' = \arg\min_{\theta} L(\theta;\ D_{train} \cup D_{new}), \quad (3)$$

где L – функция потерь, D_{train} – обучающая выборка, D_{new} – новые данные, θ – обучаемые параметры модели. Например, если модель начала ошибочно срабатывать на новую легитимную активность, ее дообучают на примерах такой активности, чтобы скорректировать ее реакцию.

2. Тактический режим. Включает пересмотр архитектуры модели, параметров и порогов классификации, что формализуется как оптимизация модели в более широком пространстве гиперпараметров:

$$h' = \arg\min_{h}\left[\min_{\theta} L(\theta,\ h;\ D_{train} \cup\ D_{new})\right],$$
$$\theta' = \arg\min_{\theta}(\theta,\ h';\ D_{train} \cup\ D_{new}), \quad (4)$$

где h – гиперпараметры модели (архитектура, пороги и др.). Например, при росте ложных срабатываний на определенный тип аварийных состояний производится подбор новых порогов классификации или увеличение сложности модели (например, глубины или числа деревьев в ансамбле).

3. Стратегический режим. В данном режиме происходит возврат к этапу моделирования сценариев и генерации обучающих данных с последующим повторным обучением моделей:

$$D_{synth} = G(S_{new}), \theta'' = \arg\min_{\theta} L\left(\theta; D_{train} \cup D_{synth}\right), \quad (5)$$

где S_{new} – новые сценарии, G – генератор данных (в ЦД объекта КИИ), D_{synth} – данные, сгенерированные по сценариям угроз. Режим обеспечивает обновление обучающих выборок с учетом новых или измененных сценариев угроз. Например, при появлении информации о новой тактике целевой атаки в ЦД моделируются соответствующие сценарии, и модели обучаются на расширенном наборе данных.

Стратегический режим адаптации предназначен для реагирования на принципиально новые типы угроз, которые не могут быть покрыты простым дообучением или оптимизацией гиперпараметров. Его запуск может быть инициирован автоматически при обнаружении аномалий, не соответствующих известным сценариям, либо по команде оператора после появления информации о новых тактиках атак из внешних источников. В первом случае модуль моделирования угроз (таблица 1) генерирует гипотетические сценарии, соответствующие выявленным отклонениям, и передает их в модуль генерации синтетических данных. Во втором случае новые сценарии разрабатываются экспертом и загружаются в ЦД.

В рамках текущей реализации фреймворка предусмотрена возможность как автоматической, так и полуавтоматической генерации, однако для обеспечения безопасности целесообразно включать этап верификации оператором перед обновлением (повторным обучением) моделей.

Процесс адаптации показан на рисунке 2.



Рис 2. Замкнутый цикл адаптации

Для автоматического выбора режима адаптации введем систему решающих правил, основанную на анализе отклонений метрик качества и характеристик новых данных.

Пусть в момент времени t после обработки очередного блока данных на реальном объекте получен вектор метрик $Q_t = (Recall_t, FPR_t, F1_t)$. Заданы допустимые интервалы: $Recall \in [R_{min}, 1], FPR \in [0, F_{max}]$, а также порог δ_{rec} минимального улучшения *Recall* при оперативной адаптации.

Оперативный режим (O) активируется при выполнении следующих условий: ($Recall_t < R_{min}$) и имеются новые достоверные данные D_{new}. Выполняется дообучение без изменения гиперпараметров.

Тактический режим (T) активируется, если после k последовательных оперативных адаптаций ($k \geq 2$) улучшение *Recall* не превысило δ_{rec}: ($Recall_t - Recall_{t-k} < \delta_{rec}$) и ($FPR_t > F_{max}$). Производится оптимизация гиперпараметров модели (например, поиск по сетке) на накопленных данных.

Стратегический режим (C) активируется при выполнении одного из следующих условий:

– обнаружение аномалии, не соответствующей ни одному из известных классов в M_threat (фиксируется модулем оценки угроз на основе кластеризации или экспертных правил);

– поступление из внешних источников информации о новой тактике атак (TTP), отсутствующей в текущей базе знаний.

Формально:

Значения R_{min}, F_{max}, δ_{rec} и k задаются на этапе настройки фреймворка исходя из требований к безопасности объекта и могут корректироваться в процессе эксплуатации.

4. Экспериментальная валидация и результаты. Целью экспериментальной валидации выступает количественная оценка эффективности предложенного адаптивного фреймворка на основе ЦД для повышения киберустойчивости интеллектуальных энергосистем. Оценка проводится по трем критериям:

– точность прогнозирования угроз ИБ;
– частота ложноположительных срабатываний;
– сокращение времени реагирования на инциденты ИБ.

Для обеспечения научной строгости и воспроизводимости результатов ниже представлено описание **набора данных**, на котором проводилась **проверка**, а также методика проведения серии из трех взаимосвязанных экспериментов.

Для апробации фреймворка был сгенерирован синтетический набор данных, имитирующий работу АСУ интеллектуальной энергосети. Набор данных моделирует поведение одного параметра (напряжения в линии электропередачи) в трех различных режимах функционирования системы:

1. Режим «Норма» (Normal). Основной сигнал – синусоида с частотой 50 Гц и амплитудой 220 В. Добавлена суточная модуляция амплитуды ±10% и гауссов шум со стандартным отклонением σ = 5 В.
2. Режим «Технический сбой» (Failure): Искаженная синусоида с плавным изменением частоты (45-55 Гц) и амплитуды (±25%).
3. Режим «Признаки киберугрозы» (Threat). Синусоида 50 Гц с наложенными атакующими воздействиями: импульсные искажения (падение амплитуды до -40%), фазовые сдвиги (0,1π-0,4π) и низкочастотная модуляция (1-5 Гц, амплитуда 10-20% от основной). Генерация синтетических данных выполняется модулем 3 ЦД. В стратегическом режиме адаптации генератор может формировать новые комбинации атакующих воздействий на основе анализа недостаточно покрытых областей пространства признаков.

Структура данных следующая:

– общее количество записей – 150000 временных отсчетов;
– частота дискретизации – 100 Гц (100 измерений напряжения в секунду);

– продолжительность моделирования – 25 минут реального времени;

– целевая переменная (метка класса). Каждая запись размечена одним из трех классов: Normal (Норма): 100000 записей (66,7%); Failure (Технический сбой): 25000 записей (16,7%); Threat (Признаки киберугрозы): 25000 записей (16,7%).

Разбиение набора данных было выполнено с учетом задачи обучения модели на "нормальном" поведении и последующей адаптации. Обучающая выборка была сформирована из первых 70000 записей класса Normal (70% от общего объема данных этого класса). Оставшиеся 30000 записей класса Normal, а также все записи классов Failure (25000) и Threat (25 000) были объединены в пул для формирования валидационной и тестовой выборок.

Валидационная выборка была сформирована путем случайной выборки 7500 записей из каждого из трех классов (всего 22500 записей, 15% от общего объема набора данных). Аналогично, тестовая выборка была сформирована из следующих 7500 записей каждого класса (всего 22500 записей, 15% от общего объема набора данных). Такое разбиение обеспечило сбалансированность валидационной и тестовой выборок по классам, что критически важно для объективной оценки способности модели различать нормальное поведение, технические сбои и признаки кибератак.

В целях повышения эффективности последующего машинного обучения, к сгенерированным временным рядам применяется двухэтапная предобработка данных: вейвлет-преобразование и нормализация. Вейвлет-преобразование применялось для подавления высокочастотного шума. В рамках эксперимента использовался базис Добеши первого порядка (db1) с уровнем декомпозиции 4. Коэффициенты детализации двух самых высокочастотных уровней были обнулены, что позволило снизить шумовую составляющую сигнала при сохранении информативных признаков аномалий. Нормализация (в частности, Min-Max нормализация) применялась для масштабирования всех признаков к единому диапазону [0; 1]. Нормализация обеспечивает равноправное участие всех признаков в формировании прогноза модели. Последовательное применение вейвлет-преобразования и нормализации позволяет подготовить данные к этапу обучения модели, повысив их информативность и обеспечив корректную работу алгоритмов машинного обучения.

Для обеспечения полной воспроизводимости экспериментов синтетический набор данных был сгенерирован с помощью детерминированного алгоритма, инициализированного фиксированным

случайным сидом (seed = 42). Алгоритм последовательно формирует временной ряд напряжения, имитирующий три режима функционирования энергосистемы – нормальный (Normal), аварийный (Failure) и признаки киберугрозы (Threat). Процедура генерации включает следующие шесть шагов:

1. Создание базового сигнала – синусоиды частотой 50 Гц и амплитудой 220 В.
2. Наложение суточной модуляции амплитуды в пределах ±10%.
3. Добавление гауссова шума со стандартным отклонением $\sigma = 5$ В.
4. Последовательная генерация 150000 временных отсчетов с шагом 10 мс (частота дискретизации 100 Гц).
5. Внедрение в случайные моменты времени заранее определенных шаблонов (паттернов):

– для режима «Технический сбой» – плавное изменение частоты (45-55 Гц) и амплитуды (±25%);

– для режима «Признаки киберугрозы» – импульсные искажения (падение амплитуды до -40%), фазовые сдвиги ($0{,}1\pi$-$0{,}4\pi$) и низкочастотная модуляция (1-5 Гц, амплитуда 10-20% от основной).

6. Присвоение каждому сгенерированному отсчету метки класса в зависимости от активного режима: «Normal», «Failure» или «Threat».

Эксперименты проводились на программно-аппаратной платформе, имитирующей архитектуру АСУ ТП интеллектуальной энергосети. Платформа реализована в виде изолированной виртуальной среды и включает эмуляторы промышленных контроллеров (ПЛК) и систем SCADA, что позволяет воспроизводить реалистичные сценарии функционирования и кибератак в контролируемых условиях, исключая риски для реальной инфраструктуры.

Все вычисления выполнялись на рабочей станции с процессором Intel Core i7, 32 ГБ ОЗУ, под управлением ОС Ubuntu 22.04 LTS. Использовались библиотеки Python: scikit-learn 1.4.2, PyWavelets 1.4.1, pandas 2.2.1, numpy 1.26.4. Для обеспечения воспроизводимости все случайные генераторы были зафиксированы с помощью np.random.seed(42).

Эксперимент 1. Валидация адаптационного механизма – повышение точности прогнозирования. Целью первого эксперимента являлась количественная оценка эффективности предложенного адаптационного механизма в повышении точности прогнозирования угроз информационной безопасности, с акцентом на минимизацию пропущенных атак через увеличение метрики Recall. В качестве тестовой среды использовалась синтетическая модель АСУ ТП

энергосистемы, построенная на основе описанного ранее набора данных. В качестве базового алгоритма обнаружения был выбран метод Isolation Forest [64]. Данный выбор обусловлен задачей доказательства концепции: этот алгоритм устойчив к несбалансированным выборкам, обладает высокой интерпретируемостью и позволяет изолированно оценить вклад адаптационного механизма без усложнения, связанного с глубокими нейросетями. Предложенный фреймворк является алгоритмически независимым и может быть расширен на другие модели (LSTM, графовые нейросети, гибридные физико-информационные модели) [65] без изменения логики адаптации.

Начальные параметры модели были установлены следующим образом: количество деревьев – 100, параметр ‘contamination’ («загрязнение») – 0,1, что соответствует ожидаемой доле аномалий в 10%. Обучение проводилось исключительно на данных класса «Норма» (70000 записей), что имитирует реальную ситуацию дефицита данных об атаках на этапе первоначального развертывания системы. Тестирование осуществлялось на сбалансированной выборке из 22500 записей, включающей по 7500 записей каждого из трех классов событий.

На первом этапе модель была обучена и протестирована в исходной конфигурации. Результаты показали, что метрика *Recall* составила 0,8253, что указывает на высокую базовую способность модели выявлять аномалии, при этом сохраняется потенциал для улучшения. Для повышения эффективности был активирован адаптационный механизм. Были последовательно применены два режима: оперативный (инкрементное дообучение на дополнительных 7500 записях класса Normal из валидационной выборки) и тактический (оптимизация гиперпараметров с помощью поиска по сетке, включавшего варианты n_estimators = [100; 150; 200] и contamination = [0,05; 0,1; 0,15]). Лучшей комбинацией, максимизирующей Recall на валидационной выборке, оказалась n_estimators =100, contamination = 0,15. После проведения этих процедур финальная адаптированная модель была протестирована на том же тестовом наборе.

Результаты, представленные в таблице 2, демонстрируют, что активация адаптационного механизма позволила стабилизировать и незначительно улучшить производительность модели. Метрика *Recall*, отражающая способность системы выявлять истинные угрозы, увеличилась на 0,7%, в то время как FPR снизилась на 0,7 процентных пункта. Это свидетельствует о том, что даже в условиях упрощенной модели, адаптация на основе обратной связи позволяет системе тонко

настраивать свои параметры для повышения общей сбалансированности (F1-score), что является важным шагом в направлении минимизации совокупного ущерба.

Таблица 2. Сравнение метрик модели до и после адаптации

Показатель	До адаптации	Послед адаптации	Изменение
Recall	0,8253	0,8325	+0,0072
F1-score	0,7583	0,7605	+0,0021
Precision	0,7014	0,6999	-0,0015
FPR (%)	17,47%	16,75%	-0,72%

Следует подчеркнуть, что данные результаты получены в контролируемой виртуальной среде на синтетических данных, моделирующих лишь один параметр энергосистемы. Они не могут быть напрямую экстраполированы на работу системы в реальных промышленных условиях, где данные многомерны, а атаки являются многокомпонентными, скрытными и эволюционирующими. Тем не менее, эксперимент подтверждает принципиальную работоспособность предложенного адаптационного механизма: система успешно прошла полный цикл – от получения обратной связи до корректировки модели – и продемонстрировала измеримое улучшение ключевых показателей. Это является важным доказательством реализуемости концепции для дальнейшего масштабирования фреймворка.

Незначительный прирост метрик, зафиксированный в ходе эксперимента, объясняется рядом факторов, заложенных в методику исследования. Основной целью первого эксперимента было не достижение максимальных абсолютных значений метрик, а демонстрация работоспособности замкнутого цикла адаптации. Успешное прохождение данных через все этапы фреймворка (от генерации в ЦД до адаптации модели на основе обратной связи) и фиксация даже незначительного, но статистически устойчивого улучшения является важным результатом, подтверждающим жизнеспособность предложенной архитектуры.

Эксперимент 2. Валидация пользы синтетических данных – снижение частоты ложных срабатываний. Целью второго эксперимента является количественная оценка вклада синтетических данных, генерируемых в среде ЦД, в снижение FPR системы обнаружения угроз информационной безопасности. В отличие от первого эксперимента, где адаптация модели происходила за счет дообучения на новых «нормальных» данных и оптимизации гиперпараметров, здесь фокус смещен на оценку качества

и релевантности аномальных данных, используемых для обучения. Ключевой исследовательский вопрос: обеспечивают ли синтетические аномалии, сгенерированные по запросу адаптационного механизма ЦД, измеримое преимущество по сравнению с использованием эквивалентного объема «внешних» имитационных данных при прочих равных условиях?

Для обеспечения научной строгости и исключения влияния конфаундеров (в частности, объема обучающей выборки), эксперимент реализован в рамках методологии контролируемого сравнения при фиксированном объеме данных. Это означает, что обе сравниваемые модели обучаются на идентичном общем количестве примеров (85000 записей), но с принципиальным различием в происхождении данных для аномальных классов:

1. Контрольная конфигурация (Baseline – «Внешние имитационные данные»); модель обучается на комбинированном наборе, включающем:

– 70000 записей класса «Норма» (исходная обучающая выборка, идентичная используемой в Эксперименте 1);

– 15000 записей аномалий (по 7500 записей классов «Failure» и «Threat»), случайным образом выбранных из валидационной выборки. Эти данные представляют собой статичный, «внешний» пул, сформированный заранее и не адаптируемый под текущие потребности модели.

2. Экспериментальная конфигурация (Proposed – «Синтетические данные ЦД»); модель обучается на комбинированном наборе, включающем:

– 70000 записей класса «Норма» (та же исходная выборка);

– 15000 записей аномалий (по 7500 записей классов «Failure» и «Threat»), сгенерированных заново с помощью детерминированного алгоритма, имитирующего работу модуля генерации синтетических данных ЦД. Эти данные создаются в момент проведения эксперимента, что соответствует сценарию проактивной генерации данных по запросу адаптационного механизма.

Таким образом, единственной независимой переменной в эксперименте выступает источник данных для аномальных классов: статичный внешний пул vs. динамически сгенерированный в среде ЦД. Все остальные параметры (алгоритм обучения, процедура предобработки, валидационная и тестовая выборки) остаются идентичными для обеих конфигураций.

Эксперимент включает следующие этапы:

1. Формирование обучающих наборов. Для контрольной конфигурации из валидационной выборки извлекаются все доступные записи классов «Failure» и «Threat». Для экспериментальной конфигурации с помощью детерминированного алгоритма генерируются новые временные ряды для тех же классов и в том же количестве, с использованием физических и атакующих моделей, описанных в разделе 4.1. Оба обучающих набора формируются путем конкатенации 70 000 «нормальных» записей с соответствующими 15000 аномальных записей и последующего случайного перемешивания.

2. Предобработка данных. Ко всем данным (обучающим, валидационным, тестовым) применяется единый конвейер предобработки: дискретное вейвлет-преобразование с последующей нормализацией Min-Max. Нормализация выполняется с использованием скейлера, обученного исключительно на исходных 70000 записях класса «Норма».

3. Оптимизация гиперпараметров. Для каждой конфигурации независимо выполняется поиск оптимальных гиперпараметров алгоритма Isolation Forest (n_estimators $\in$ {100; 150; 200}, contamination $\in$ {0,05; 0,1; 0,15}) на валидационной выборке. Целевой функцией оптимизации является максимизация метрики Recall для аномалий (объединенные классы «Failure» и «Threat»), что соответствует задаче минимизации пропущенных угроз.

4. Финальное обучение и оценка. Модели с найденными оптимальными гиперпараметрами обучают на своих полных обучающих наборах и оценивают на единой, сбалансированной тестовой выборке (22500 записей: по 7500 на каждый из трех классов). Оценка проводится по ключевым метрикам: *Recall*, *Precision*, F1-score и *FPR*.

Результаты оценки на тестовой выборке представлены в таблице 3. Анализ данных демонстрирует статистически значимое преимущество экспериментальной конфигурации.

Основной результат эксперимента заключается в следующем. Использование синтетических данных, генерируемых в среде ЦД, позволило снизить FPR на 6,4% (или на 3,56 процентных пункта) при одновременном повышении F1-score на 4,7% и Recall на 8,0%. Это свидетельствует о том, что данные, созданные «по запросу» в рамках адаптивного контура ЦД, в данных условиях являются более информативными для задачи обучения модели, чем данные, взятые из статичного «внешнего» пула. Они позволяют модели лучше отделить истинные угрозы (класс «Threat») и технические сбои (класс «Failure») от нормального функционирования, что напрямую приводит к снижению числа ложных тревог.

Таблица 3. Сравнение эффективности моделей, обученных на «внешних» и «синтетических» аномалиях на тестовой выборке

Метрика	Контрольная конфигурация	Экспериментальная конфигурация	Абсолютное изменение	Относительное изменение
Recall	0,4469	0,4825	+0,0356	+8,0%
Precision	0,7155	0,7138	-0,0017	-0,2%
F1-score	0,5501	0,5758	+0,0257	+4,7%
FPR (%)	55,31%	51,75%	-3,56 п.п.	-6,4%

Результаты эксперимента подтверждают гипотезу о том, что ЦД выступает не просто как инструмент симуляции, а как генератор знаний. Способность ЦД проактивно создавать синтетические сценарии атак и сбоев, адаптированные под текущие потребности модели, является ключевым фактором для снижения операционной нагрузки на персонал за счет уменьшения ложных срабатываний.

Эксперимент 3. Количественная оценка сокращения совокупного времени реагирования на инциденты ИБ через оптимизацию операционной нагрузки. Третий эксперимент направлен на количественную оценку влияния адаптационного механизма на моделируемое сокращение совокупного времени реагирования (Total Time to Respond, TTR) на инциденты информационной безопасности.

В отличие от экспериментов 1 и 2, где аномалиями считались как кибератаки, так и технические сбои, в данном эксперименте мы фокусируемся исключительно на классе «Threat». Технические сбои (Failure) рассматриваются как часть нормального функционирования, поскольку они не требуют немедленного реагирования со стороны центра мониторинга безопасности. Таким образом, тестовая выборка содержит 15 000 нормальных записей (7500 класса «Normal» и 7500 класса «Failure») и 7500 записей класса «Threat». Используется та же выборка, что и в эксперименте 1, но с переопределенной целевой переменной.

Совокупное время реагирования определяется как суммарное время, затрачиваемое на обработку всех срабатываний (истинных и ложных) в течение фиксированного периода мониторинга. Оно зависит от количества истинных срабатываний (True Positives, TP) и ложных срабатываний (False Positives, FP):

$$TTR = (TP \times C_{tp}) + (FP \times C_{fp}), \quad (6)$$

где C_{tp} – среднее время (в секундах), затрачиваемое на обработку одного истинного срабатывания (например, верификация угрозы, инициация

контмер, документирование инцидента), C_{fp} – среднее время (в секундах), затрачиваемое на обработку одного ложного срабатывания (например, анализ логов, отклонение тревоги, сброс состояния).

Значения констант C_{tp} и C_{fp} основаны на эмпирических данных, приведенных в исследованиях по эффективности центров мониторинга безопасности. Согласно отраслевым оценкам, среднее время первичной верификации реального инцидента составляет от 30 до 90 минут, а обработка ложного срабатывания – от 5 до 30 минут [66 – 68]. Для целей данного эксперимента выбраны консервативные оценки нижней границы в секундах: C_{tp} = 1800 секунд (30 минут), C_{fp} = 300 секунд (5 минут). Это соответствует минимально возможному времени обработки при высокой степени автоматизации и зрелости процессов центра мониторинга и реагирования на инциденты. В реальных условиях с более длительными процедурами обработки выигрыш от снижения числа ложных срабатываний будет выше.

Процедура проведения эксперимента следующая:

1. Исходные данные. В качестве входных данных используются результаты прогнозирования двух моделей на единой тестовой выборке, полученные в ходе эксперимента 1:

– сценарий «До адаптации»: предсказания базовой модели (обучена только на данных класса «Норма»);

– сценарий «После адаптации»: предсказания финальной адаптированной модели (прошедшей оперативный и тактический режимы адаптации)».

Для расчета *TP* и *FP* метки классов переопределяются. Положительным классом считается «Threat», отрицательным – объединение классов «Normal» и «Failure».

2. Расчет метрик TP и FP. Для каждого сценария на основе вектора предсказаний и истинных меток рассчитывается матрица ошибок. Из нее извлекаются значения TP и FP:

– *TP* – количество записей, где истинный класс – «Threat» и модель предсказала «аномалия» (-1);

– *FP* – количество записей, где истинный класс – «Normal» или «Failure», но модель предсказала «аномалия» (-1).

3. Расчет TTR. По формуле (6) рассчитывается совокупное время реагирования для каждого сценария.

4. Сравнение и оценка сокращения. Рассчитывается абсолютное и относительное сокращение TTR:

$$\Delta TTR = TTR_{before} - TTR_{after}, \quad (7)$$

$$\Delta TTR\% = \left(\frac{\Delta TTR}{TTR_{before}}\right) \times 100\%. \quad (8)$$

Результаты расчета представлены в таблице 4. Расчет в таблице выполнен для бинарной классификации, где положительным классом являются только кибератаки («Threat»), а технические сбои отнесены к нормальному фону. Значения *TP* и *FP* получены непосредственно из матриц ошибок моделей до и после адаптации на тестовой выборке объемом 22500 записей (15000 нормальных и 7500 атак).

Таблица 4. Расчет совокупного времени реагирования (TTR) до и после адаптации.

Сценарий	*TP*	*FP*	*TTR*, сек.	Моделируемое сокращение *TTR*
До адаптации	2231	2621	4802100	-
После адаптации	2147	2513	4618500	-183 600 (-3,82%)

Анализ данных демонстрирует, что активация адаптационного механизма привела к сокращению совокупного времени реагирования на 3,82%. Сокращение обусловлено снижением количества ложных срабатываний на 108 единиц (с 2621 до 2513), что уменьшает нагрузку на персонал центра мониторинга безопасности. Небольшое снижение числа истинных срабатываний (на 84 события, с 2231 до 2147) связано с изменением порогов классификации в процессе тактической адаптации, однако вклад сокращения *FP* оказался доминирующим при выбранных весовых коэффициентах. Полученный результат подтверждает, что даже умеренное улучшение метрик, особенно снижение *FPR*, способно дать измеримый практический эффект в виде экономии времени аналитиков.

Для количественной оценки эффективности адаптации использовалась функция потерь (1). Приняв весовые коэффициенты $\lambda_1 = 0{,}5$ (приоритет полноты), $\lambda_2 = 0{,}3$ (чувствительность к ложным срабатываниям) и $\lambda_3 = 0{,}2$ (время реагирования), а также $T_{max} = 4802100$ с (базовое время до адаптации), получим снижение функции потерь (1) на 16,8%.

5. Обсуждение. Представленная работа предлагает интегрированный адаптивный фреймворк на основе ЦД для повышения результативности обнаружения киберугроз в интеллектуальных энергосистемах. Данный раздел посвящен обсуждению полученных результатов, их места в контексте существующих исследований, а также анализу сильных и слабых сторон предложенного подхода.

Научный вклад настоящего исследования заключается в интеграции архитектуры ЦД для АСУ энергосистемы, метода адаптивного обнаружения аномалий и механизма генерации синтетических данных в единый, управляемый цикл адаптации на основе обратной связи. В отличие от большинства существующих решений, фокусирующихся на одном аспекте (моделирование, обнаружение или генерация данных), предложенный фреймворк создает замкнутую систему, где результаты работы на виртуальном прототипе используются для непрерывного совершенствования моделей обнаружения.

Результаты серии экспериментов позволяют сделать следующие выводы:

1. Активация адаптационного механизма приводит к измеримому, хотя и незначительному, улучшению баланса между чувствительностью и специфичностью модели: *Recall* увеличился на 0,7%, *FPR* снизился на 0,7 п.п. Это подтверждает работоспособность концепции замкнутого цикла, где обратная связь от системы используется для тонкой настройки модели.

2. Использование синтетических данных, генерируемых в среде ЦД, демонстрирует статистически значимое преимущество: снижение *FPR* на 6,4% (3,56 п.п.) и рост F1-score на 4,7% по сравнению с обучением на статичном внешнем пуле данных. Это свидетельствует о том, что адаптивная генерация сценариев позволяет создавать более релевантные обучающие примеры, что напрямую снижает операционную нагрузку за счет уменьшения ложных срабатываний.

3. Моделируемое сокращение совокупного времени реагирования (*TTR*) на 3,82% при снижении числа ложных срабатываний на 108 единиц подчеркивает практическую значимость даже незначительных улучшений метрик. Это измеримый эффект, оказывающий прямое влияние на эффективность работы центров мониторинга безопасности.

В целом, результаты согласуются с гипотезой о том, что ЦД может служить не просто инструментом для симуляции, а динамической платформой для генерации данных и тестирования. Однако важно подчеркнуть: ЦД не является источником автономной функции обнаружения или принятия решений. Его роль строго ограничена предоставлением контролируемой среды для генерации сценариев и верификации моделей. Качество конечного результата полностью определяется двумя факторами:

- качеством алгоритма обнаружения аномалий;
- релевантностью данных.

Эксперименты демонстрируют, что при фиксированном алгоритме обнаружения, улучшение качества входных данных (через адаптивную генерацию в ЦД) приводит к измеримому повышению эффективности. Таким образом, ЦД выступает как инфраструктурный элемент, автоматизирующий процессы, которые вручную были бы крайне затратны или невозможны.

Результаты подтверждают возможность применения подхода, однако не доказывают его эффективность в промышленной эксплуатации.

Основные преимущества фреймворка, вытекающие из его архитектуры и подтвержденные экспериментами, заключаются в следующем:

1. Все этапы, связанные с моделированием атак и обучением моделей, происходят в изолированной виртуальной среде ЦД. Это нивелирует риск нарушения работы реальной энергосистемы.

2. Трехуровневый адаптационный механизм (оперативный, тактический, стратегический) обеспечивает непрерывную обратную связь и корректировку моделей **с минимальным ручным вмешательством** (в стратегическом режиме возможна полуавтоматическая генерация сценариев с последующей верификацией оператором), позволяя системе адаптироваться к меняющимся угрозам.

3. Генерация синтетических данных в среде ЦД позволяет обходить фундаментальную проблему отрасли – дефицит размеченных данных о реальных инцидентах.

4. Интеграция задач кибербезопасности на этапе проектирования и моделирования в ЦД позволяет выявлять и устранять уязвимости еще до развертывания системы в промышленную эксплуатацию.

Несмотря на полученные положительные результаты, необходимо четко обозначить границы их применимости. Проведенное исследование имеет ряд существенных ограничений:

1. Ограниченная размерность данных. Все эксперименты проведены на синтетическом наборе данных, имитирующем поведение единственного параметра (напряжения). Реальные киберфизические атаки на АСУ ТП являются многомерными, затрагивая множество взаимосвязанных параметров (ток, частота, температура, давление, сетевой трафик, логи событий). Архитектура фреймворка поддерживает многомерные данные за счет модульной структуры и возможности интеграции алгоритмов обнаружения, работающих с векторами признаков произвольной размерности. Проверка

эффективности в условиях многомерной корреляции составляет одно из направлений дальнейших исследований.

2. Упрощенность алгоритмической базы. В качестве детектора аномалий в рамках исследования использовался алгоритм Isolation Forest. Выбор обусловлен задачей доказательства осуществимости концепции: алгоритм обладает высокой интерпретируемостью, устойчив к несбалансированным выборкам и требует незначительных вычислительных ресурсов, что позволяет изолировать и оценить вклад именно механизмов адаптации и генерации данных, минимизируя влияние сложности модели детектирования. Следует отметить, что Isolation Forest не предназначен для решения задач обнаружения сложных, многоэтапных или малозаметных атак, требующих учета временной динамики и скрытых зависимостей в данных. Результаты, полученные с его использованием, не могут быть напрямую экстраполированы на более сложные архитектуры, такие как рекуррентные или сверточно-рекуррентные нейронные сети, а также гибридные модели, сочетающие методы машинного обучения с физическими законами функционирования энергосистем, которые в современных исследованиях демонстрируют более высокие показатели точности в аналогичных задачах.

3. Отсутствие верификации на реальных данных. Все результаты получены на синтетических данных, сгенерированных детерминированным алгоритмом. Они не подтверждены на исторических данных реальных инцидентов или на данных, полученных с физических тестовых стендов. Это является критическим ограничением, поскольку синтетические данные, даже при высокой степени детализации, не способны в полной мере воспроизвести стохастическую природу, шум и непредсказуемость реальных технологических процессов и атак.

4. Недостаточная сложность моделирования атак. Сгенерированные атаки представляют собой относительно простые паттерны (импульсные искажения, фазовые сдвиги). Они не моделируют скрытность, адаптивность или многоэтапность реальных APT, которые специально разработаны для того, чтобы оставаться незамеченными в течение длительного времени. Эффективность фреймворка против таких изощренных угроз не оценивалась.

5. Отсутствие анализа операционных издержек. В работе не проводился анализ вычислительной сложности адаптационного механизма и затрат на создание и поддержку ЦД. Внедрение ЦД

требует значительных ресурсов на этапе проектирования, калибровки и синхронизации с реальной системой. Эти затраты должны быть сопоставлены с получаемой выгодой для принятия обоснованных решений о промышленном внедрении.

Несмотря на указанные ограничения, проведенное исследование может представлять практический интерес на текущем этапе развития технологий ЦД. Предложенный подход формирует методологическую основу и предоставляет инструментарий для безопасной генерации обучающих данных и верификации алгоритмов обнаружения аномалий в контролируемой среде. Это может быть полезно на этапах проектирования, предварительной оценки устойчивости и обучения персонала, когда доступ к реальным данным ограничен или их использование сопряжено с рисками. Интеграция подобного фреймворка в процессы жизненного цикла АСУ может способствовать повышению их киберустойчивости, однако окончательная оценка его практической применимости требует дальнейшей экспериментальной проверки в условиях, приближенных к промышленной эксплуатации.

Следует отметить потенциальную уязвимость предложенного механизма к атакам типа «отравление модели» (Model Poisoning) [69]. Если злоумышленник сможет контролировать часть данных обратной связи, используемых для дообучения (например, имитировать легитимные изменения или маскировать атаки под нормальные отклонения), это может привести к постепенному смещению модели и снижению качества обнаружения. Данная проблема характерна для систем с обратной связью и требует дополнительных мер защиты, таких как проверка источников данных, использование робастных методов обучения или периодический аудит обновлений оператором. В рамках данного исследования данный аспект не рассматривался и выступает предметом дальнейших исследований.

Для объективной оценки места предложенного фреймворка в существующем ландшафте, в таблице 5 приведено его сравнение с двумя основными классами альтернативных решений. Сравнительный анализ носит качественный характер. Он основан на результатах, полученных в рамках экспериментов с алгоритмом Isolation Forest. Количественная валидация указанных характеристик – в частности, точности против сложных APT, скорости адаптации и полной стоимости владения (TCO) – в условиях реальной промышленной эксплуатации выходит за рамки данного исследования и определена как направление будущей работы.

Важно подчеркнуть, что предлагаемый фреймворк не позиционируется как замена существующим методам обнаружения

аномалий. Скорее, он предлагает альтернативный подход к организации процессов обучения и верификации моделей через использование виртуальной среды ЦД. Его основное преимущество – возможность безопасного тестирования и генерации данных без зависимости от наличия реальных инцидентов – потенциально позволяет использовать его в качестве вспомогательного инструмента при работе с различными классами алгоритмов обнаружения. Однако заявлять о его универсальной применимости ко всем существующим или будущим архитектурам преждевременно. Эффективность интеграции с конкретными моделями (включая нейросетевые или физико-ориентированные) требует отдельной экспериментальной проверки и не может быть постулирована априори.

Таблица 5. Сравнительный анализ подходов к обнаружению угроз ИБ

Критерий	Традиционные методы (сигнатуры, правила)	Современные методы глубокого обучения	Предлагаемый фреймворк
Точность (против zero-day)	Низкая	Высокая (при наличии данных)	Высокая (адаптируется к новым угрозам)
Адаптивность	Отсутствует	Низкая (требует ручного переобучения)	Высокая (автоматическая адаптация)
Безопасность тестирования	Высокая	Низкая (риск при обучении на live-данных)	Очень высокая (тестирование в ЦД)
Зависимость от реальных данных об атаках	Низкая	Очень высокая	Низкая (используются синтетические данные)
Время и стоимость внедрения	Быстрое и дешевое	Долгое и дорогое (сбор и разметка данных)	Среднее (настройка ЦД и сценариев)

6. Заключение. В статье представлен и экспериментально апробирован прототип интегрированного адаптивного фреймворка на основе ЦД для обнаружения признаков киберугроз в системах управления интеллектуальными энергосетями. Ключевым элементом фреймворка выступает замкнутый цикл адаптации, управляемый метриками качества, в котором результаты работы системы на реальном объекте используются для формализованного совершенствования моделей в виртуальной среде ЦД.

Результаты экспериментов демонстрируют принципиальную работоспособность предложенного подхода. Активация адаптационного механизма позволяет улучшить баланс между

чувствительностью и специфичностью модели (рост *Recall* на 0,7%, снижение *FPR* на 0,7 процентных пункта). Использование синтетических данных, генерируемых в среде ЦД, обеспечивает статистически значимое снижение частоты ложных срабатываний (на 6,4%) и рост F1-score (на 4,7%). Даже небольшое снижение числа ложных срабатываний (на 108 единиц) приводит к моделируемому сокращению совокупного времени реагирования (на 3,82%), что подтверждает практическую значимость оптимизации метрики FPR для повышения результативности работы систем обнаружения.

Следует подчеркнуть, что все выводы, сделанные в настоящей работе, основаны на результатах, полученных в рамках ограниченной экспериментальной модели – синтетического набора данных и виртуальной платформы, имитирующей архитектуру АСУ ТП Smart-сети. Следовательно, представленные результаты следует рассматривать исключительно как доказательство реализуемости концепции, демонстрирующее принципиальную работоспособность и потенциал предложенного фреймворка. Они не являются окончательным подтверждением его эффективности в условиях реальной промышленной эксплуатации, но обосновывают целесообразность проведения дальнейших, более масштабных исследований.

Для трансформации прототипа в промышленно применимое решение необходимо решить следующие задачи:

1. Расширить фреймворк для работы с комплексными, многомерными потоками данных, включая параметры тока, частоты, температуры, сетевого трафика и логов событий.
2. Провести апробацию на физических тестовых стендах и с использованием исторических данных реальных инцидентов для подтверждения эффективности в условиях, близких к промышленным.
3. Провести серию экспериментов с заменой Isolation Forest на другие классы моделей для определения границ применимости и вклада адаптационного механизма при различных типах детекторов.
4. Создать API для бесшовной интеграции фреймворка с существующими корпоративными платформами (SIEM, SOAR).
5. Оценка эффективности фреймворка при интеграции с современными архитектурами глубокого обучения (например, LSTM, ConvLSTM) и физико-ориентированными моделями на одинаковых наборах данных для объективного определения его вклада в повышение точности обнаружения.

Реализация указанных направлений, на наш взгляд, позволит приблизить перевод фреймворка из стадии исследовательского прототипа в стадию зрелого инженерного решения.

В заключение отметим направления дальнейших фундаментальных и прикладных исследований:

1. Исследование многомерной адаптивной детекции в киберфизических системах. Разработка и верификация моделей, способных одновременно анализировать коррелированные потоки данных (напряжение, ток, частота, сетевой трафик, логи событий).

2. Разработка теоретико-игровых и активных методов генерации сценариев в ЦД, исследование механизмов проактивного моделирования адаптивных атакующих стратегий на основе принципов состязательного машинного обучения (adversarial machine learning) и теории игр. Это позволит генерировать синтетические данные, максимально приближенные к тактикам реальных APT.

3. Формализация метрик эффективности замкнутого адаптационного цикла. Создание математической модели, описывающей взаимосвязь между качеством синтетических данных, скоростью адаптации, изменением метрик обнаружения и совокупным временем реагирования.

4. Исследование устойчивости адаптивных моделей к целенаправленным атакам на процесс обучения.

5. Разработка методов защиты цифрового двойника как критического элемента инфраструктуры кибербезопасности.

6. Экспериментальная верификация на физических тестовых стендах и реальных промышленных данных.

Эти направления формируют основу для систематического развития предложенного подхода от исследовательского прототипа к теоретически обоснованной и практически применимой методологии обеспечения киберустойчивости интеллектуальных энергосистем.

Литература

1. Gehrmann C., Gunnarsson M. A Digital Twin Based Industrial Automation and Control System Security Architecture // IEEE Transactions on Industrial Informatics. 2020. vol. 16. no. 1. pp. 669–680. DOI: 10.1109/TII.2019.2938885.
2. Venkatachary S.K., Prasad J., Alagappan A., Andrews L.J.B., Raj R.A., Duraisamy S. Cybersecurity and cyber-terrorism challenges to energy-related infrastructures – Cybersecurity frameworks and economics – Comprehensive review // International Journal of Critical Infrastructure Protection. 2024. vol. 45. 100677 p. DOI: 10.1016/j.ijcip.2024.100677.
3. Diaba S.Y., Shafie-khah M., Elmusrati M. Cyber-physical attack and the future energy systems: A review // Energy Reports. 2024. vol. 12. pp. 2914–2932. DOI: 10.1016/j.egyr.2024.08.060.
4. Kotenko I., Saenko I., Lauta O., Kribel A. An Approach to Detecting Cyber Attacks against Smart Power Grids Based on the Analysis of Network Traffic Self-Similarity // Energies. 2020. vol. 13. no. 19. 5031 p. DOI: 10.3390/en13195031.
5. Izrailov K., Buinevich M., Kotenko I. Intelligent Detection of Cyber Attacks on Electrical Power Systems Based on Simulation and Graph-Based Modeling //

Proceedings of the International Conference Intelligent Systems (INTELS'24). Communications in Computer and Information Science (CCIS). Springer. 2026. vol. 2603. pp. 231–245. DOI: 10.1007/978-3-032-04758-8_18.
6. Lin H. Computer network information security monitoring system // Proceedings of Second International Conference on Big Data, Computational Intelligence, and Applications (BDCIA 2024). 2025. vol. 13550. DOI: 10.1117/12.3059895.
7. Villegas-Ch W., Govea J., Gutierrez R., Navarro A.M., Mera-Navarrete A. Effectiveness of an Adaptive Deep Learning-Based Intrusion Detection System // IEEE Access. 2024. vol. 12. pp. 184010–184027. DOI: 10.1109/ACCESS.2024.3512363.
8. Kotenko I., Saenko I., Lauta O., Kribel A. A Proactive Protection of Smart Power Grids against Cyberattacks on Service Data Transfer Protocols by Computational Intelligence Methods // Sensors. 2022. vol. 22. no. 19. 7506 p. DOI: 10.3390/s22197506.
9. Desnitsky V.A., Kotenko I.V., Nogin S.B. Detection of Anomalies in Data for Monitoring of Security Components in the Internet of Things // Proceedings of the XVIII IEEE International Conference on Soft Computing and Measurements (SCM). 2015. pp. 189–192. DOI: 10.1109/SCM.2015.7190452.
10. Aydin Z. Detecting Cybersecurity Threats in Digital Energy Systems Using Deep learning for Imbalanced Datasets // International Journal of Energy Economics and Policy. 2025. vol. 15. no. 3. pp. 614–628. DOI: 10.32479/ijeep.19649.
11. Cremer F., Sheehan B., Fortmann M., et al. Cyber risk and cybersecurity: a systematic review of data availability // The Geneva Papers on Risk and Insurance - Issues and Practice. 2022. vol. 47. pp. 698–736. DOI: 10.1057/s41288-022-00266-6.
12. Kotenko I., Chechulin A. Computer attack modeling and security evaluation based on attack graphs // Proceedings of the IEEE 7th International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS). 2013. pp. 614–619. DOI: 10.1109/IDAACS.2013.6662998.
13. Homaei M., Mogollon-Gutierrez O., Sancho J.C., et al. A review of digital twins and their application in cybersecurity based on artificial intelligence // Artificial Intelligence Review. 2024. vol. 57. 201 p. DOI: 10.1007/s10462-024-10805-3.
14. McLaughlin K.L. The power of digital twins in the cybersecurity mesh // EDPACS. 2023. vol. 68. no. 6. pp. 35–39. DOI: 10.1080/07366981.2023.2263214.
15. Krishnaveni S., Chen T.M., Sathiyanarayanan M., et al. CyberDefender: an integrated intelligent defense framework for digital-twin-based industrial cyber-physical systems // Cluster Computing. 2024. vol. 27. pp. 7273–7306. DOI: 10.1007/s10586-024-04320-x.
16. Salim M.M., Camacho D., Park J.H. Digital Twin and federated learning enabled cyberthreat detection system for IoT networks // Future Generation Computer Systems. 2024. vol. 161. pp. 701–713. DOI: 10.1016/j.future.2024.07.017.
17. Sousa B., Arieiro M., Pereira V., Correia J., Lourenco N., Cruz T. ELEGANT: Security of Critical Infrastructures With Digital Twins // IEEE Access. 2021. vol. 9. pp. 107574–107588. DOI: 10.1109/ACCESS.2021.3100708.
18. De Hoz D.J., Temperekidis A., Katsaros P., Konstantinou C. An IoT Digital Twin for Cyber-Security Defence Based on Runtime Verification // Lecture Notes in Computer Science. Springer. 2022. vol. 13701. pp. 556–574. DOI: 10.1007/978-3-031-19849-6_31.
19. Erkek I., Irmak E. Enhancing Cybersecurity of a Hydroelectric Power Plant Through Digital Twin Modeling and Explainable AI // IEEE Access. 2025. vol. 13. pp. 41887–41908. DOI: 10.1109/ACCESS.2025.3547672.
20. Li Y., Guan P., Li T., Larsen K.G., Aiello M., Pedersen T.B., Huang T., Zhang Y. Digital Twin for Secure Peer-to-Peer Trading in Cyber-Physical Energy Systems // IEEE Transactions on Network Science and Engineering. 2025. vol. 12. no. 2. pp. 669–683. DOI: 10.1109/TNSE.2024.3507956.
21. Patel T., Jadav N.K., Rathod T., Tanwar S., Garg D., Shahinzadeh H. AI-based Secure Intrusion Detection Framework for Digital Twin-enabled Critical Infrastructure //

Proceedings of the 14th International Conference on Information and Knowledge Technology (IKT). 2023. pp. 24–29. DOI: 10.1109/IKT62039.2023.10433057.
22. Ma J., Guo Y., Fang C., Zhang Q. Digital-Twin-Based CPS Anomaly Diagnosis and Security Defense Countermeasure Recommendation // IEEE Internet of Things Journal. 2024. vol. 11. no. 10. pp. 18726–18738. DOI: 10.1109/JIOT.2024.3366904.
23. Ghenai C., Husein L.A., Nahlawi M.A., Hamid A.K., Bettayeb M. Recent trends of digital twin technologies in the energy sector: A comprehensive review // Sustainable Energy Technologies and Assessments. 2022. vol. 54. 102837 p. DOI: 10.1016/j.seta.2022.102837.
24. Yu W., Patros P., Young B., Klinac E., Walmsley T.G. Energy digital twin technology for industrial energy management: Classification, challenges and future // Renewable and Sustainable Energy Reviews. 2022. vol. 161. 112407 p. DOI: 10.1016/j.rser.2022.112407.
25. Ismail F.B., Al-Faiz H., Hasini H., Al-Bazi A., Kazem H.A. A comprehensive review of the dynamic applications of the digital twin technology across diverse energy sectors // Energy Strategy Reviews. 2024. vol. 52. 101334 p. DOI: 10.1016/j.esr.2024.101334.
26. Hashmi R., Liu H., Yavari A. Digital Twins for Enhancing Efficiency and Assuring Safety in Renewable Energy Systems: A Systematic Literature Review // Energies. 2024. vol. 17. no. 11. 2456 p. DOI: 10.3390/en17112456.
27. Stadtmann F., Rasheed A., Kvamsdal T., et al. Digital Twins in Wind Energy: Emerging Technologies and Industry-Informed Future Directions // IEEE Access. 2023. vol. 11. pp. 110762–110795. DOI: 10.1109/ACCESS.2023.3321320.
28. Semeraro C., Aljaghoub H., Abdelkareem M.A., Alami A.H., Olabi A.G. Digital twin in battery energy storage systems: Trends and gaps detection through association rule mining // Energy. 2023. vol. 273. 127086 p. DOI: 10.1016/j.energy.2023.127086.
29. Wanasinghe T.R., Wroblewski L., Petersen B.K., et al. Digital Twin for the Oil and Gas Industry: Overview, Research Trends, Opportunities, and Challenges // IEEE Access. 2020. vol. 8. pp. 104175–104197. DOI: 10.1109/ACCESS.2020.2998723.
30. Amaral J.V.S.D., Santos C.H.D., Montevechi J.A.B., De Queiroz A.R. Energy Digital Twin applications: A review // Renewable and Sustainable Energy Reviews. 2023. vol. 188. 113891 p. DOI: 10.1016/j.rser.2023.113891.
31. Heluany J.B., Gkioulos V. A review on digital twins for power generation and distribution // International Journal of Information Security. 2023. vol. 23. pp. 1171–1195. DOI: 10.1007/s10207-023-00784-x.
32. Das O., Zafar M.H., Sanfilippo F., Rudra S., Kolhe M.L. Advancements in digital twin technology and machine learning for energy systems: A comprehensive review of applications in smart grids, renewable energy, and electric vehicle optimisation // Energy Conversion and Management: X. 2024. vol. 24. 100715 p. DOI: 10.1016/j.ecmx.2024.100715.
33. Ba L., Tangour F., Abbassi I.E., Absi R. Analysis of Digital Twin Applications in Energy Efficiency: A Systematic Review // Sustainability. 2025. vol. 17. no. 8. 3560 p. DOI: 10.3390/su17083560.
34. Koirala B., Cai H., Khayatian F., Munoz E., An J.G., Mutschler R., et al. Digitalization of Urban Multi-Energy Systems – Advances in Digital Twin Applications across Life-Cycle Phases // Advances in Applied Energy. 2024. vol. 16. 100196 p. DOI: 10.1016/j.adapen.2024.100196.
35. Gourisetti S.N.G., Bhadra S., Sebastian-Cardenas D.J., Touhiduzzaman Md., Ahmed O. A Theoretical Open Architecture Framework and Technology Stack for Digital Twins in Energy Sector Applications // Energies. 2023. vol. 16. no. 13. 4853 p. DOI: 10.3390/en16134853.

36. Boukaf M., Fadli F., Meskin N. A Comprehensive Review of Digital Twin Technology in Building Energy Consumption Forecasting // IEEE Access. 2024. vol. 12. pp. 187778–187799. DOI: 10.1109/ACCESS.2024.3498107.
37. Maksimovic M., Jokic S., Boskovic M.C. Innovative Horizons for Sustainable Smart Energy: Exploring the Synergy of 5G and Digital Twin Technologies // Process Integration and Optimization for Sustainability. 2025. vol. 9. pp. 431–470. DOI: 10.1007/s41660-024-00478-4.
38. Wang Y., Kang X., Chen Z. A survey of Digital Twin techniques in smart manufacturing and management of energy applications // Green Energy and Intelligent Transportation. 2022. vol. 1. no. 2. 100014 p. DOI: 10.1016/j.geits.2022.100014.
39. Masi M., Sellitto G.P., Aranha H., Pavleska T. Securing critical infrastructures with a cybersecurity digital twin // Software and Systems Modeling. 2023. vol. 22. pp. 689–707. DOI: 10.1007/s10270-022-01075-0.
40. Gulyamov S., Akhmedov A., Bazarov S., Ubaydullaeva A., Musaev S., Rodionov A., Odilkhujaev I. Using Digital Twins for Modeling and Testing Cybersecurity Scenarios in Smart Cities // Proceedings of the 6th International Conference on Control Systems, Mathematical Modeling, Automation and Energy Efficiency (SUMMA). 2024. pp. 655–658. DOI: 10.1109/SUMMA64428.2024.10803689.
41. Alhamam N., Rahman M.M.H., Aljughaiman A. A Comprehensive Review on Cybersecurity of Digital Twins Issues, Challenges, and Future Research Directions // IEEE Access. 2023. vol. 13. pp. 45106–45124. DOI: 10.1109/ACCESS.2025.3545004.
42. Coppolino L., Nardone R., Petruolo A., Romano L., Souvent A. Exploiting Digital Twin technology for Cybersecurity Monitoring in Smart Grids // Proceedings of the 18th International Conference on Availability, Reliability and Security. 2023. pp. 1–10. DOI: 10.1145/3600160.3605043.
43. Srivastava A., Liu C.-C., Ştefanov A., Basumallik S., et al. Digital Twins Serving Cybersecurity: More Than a Model: Cybersecurity as a Future Benefit of Digital Twins 2 // IEEE Power and Energy Magazine. 2024. vol. 22. no. 1. pp. 61–71. DOI: 10.1109/MPE.2023.3325196.
44. Olivares-Rojas J.C., Reyes-Archundia E., Gutierrez-Gnecchi J.A., et al. Towards Cybersecurity of the Smart Grid Using Digital Twins // IEEE Internet Computing. 2021. vol. 26. pp. 52–57. DOI: 10.1109/MIC.2021.3063674.
45. Sarker I.H., Janicke H., Mohsin A., Gill A.Q., Maglaras L. Explainable AI for cybersecurity automation, intelligence and trustworthiness in digital twin: Methods, taxonomy, challenges and prospects // ICT Express. 2024. vol. 10. pp. 935–958. DOI: 10.1016/j.icte.2024.05.007.
46. Suhail S., Iqbal M., Mclaughlin K. Digital-twin-driven deception platform: vision and way forward // IEEE Internet Computing. 2024. vol. 28. no. 4. pp. 40–47. DOI: 10.1109/MIC.2024.3406188.
47. Alcaraz C., Lopez J. Digital Twin: A Comprehensive Survey of Security Threats // IEEE Communications Surveys and Tutorials. 2022. vol. 24. no. 3. pp. 1475–1503. DOI: 10.1109/COMST.2022.3171465.
48. Holmes D., Papathanasaki M., Maglaras L., Ferrag M.A., Nepal S., Janicke H. Digital Twins and Cyber Security – solution or challenge? // Proceedings of the 2021 6th South-East Europe Design Automation, Computer Engineering, Computer Networks and Social Media Conference (SEEDA-CECNSM). 2021. pp. 1–8. DOI: 10.1109/SEEDA-CECNSM53056.2021.9566277.
49. Singh A. Enhancing Cybersecurity for Digital Twins: Challenges and Solutions // International Journal of Science and Technology. 2024. vol. 15. no. 4. pp. 1–9. DOI: 10.71097/ijsat.v15.i4.1651.
50. Kumar R., Aljuhani A., Javeed D., Kumar P., Islam S., Islam A.K.M.N. Digital Twins-enabled Zero Touch Network: A smart contract and explainable AI integrated

cybersecurity framework // Future Generation Computer Systems. 2024. vol. 156. pp. 191–205. DOI: 10.1016/j.future.2024.02.015.
51. Ravinder M., Kulkarni V. Smart Grid Anomaly Detection Using MFDA and Dilated GRU-based Neural Networks // Smart Grids and Sustainable Energy. 2025. vol. 10. 9 p. DOI: 10.1007/s40866-024-00216-2.
52. Alkuwari A.N., Al-Kuwari S., Albaseer A., Qaraqe M. Anomaly Detection on Smart Grids with Optimized Convolutional Long Short-Term Memory Model // IEEE Access. 2025. vol. 13. pp. 40399–40412. DOI: 10.1109/ACCESS.2025.3547037.
53. Yu L., Zhang X., Du L., Yue L. Anomaly Detection of Cyber Attacks in Smart Grid Communications Based on Residual Recurrent Neural Networks // Security and Privacy. 2025. vol. 8. no. 1. DOI: 10.1002/spy2.498.
54. Jithish J., Alangot B., Mahalingam N., Yeo K.S. Distributed Anomaly Detection in Smart Grids: A Federated Learning-Based Approach // IEEE Access. 2016. vol. 11. pp. 7157–7179. DOI: 10.1109/ACCESS.2023.3237554.
55. Abdel-Basset M., Moustafa N., Hawash H. Privacy-Preserved Generative Network for Trustworthy Anomaly Detection in Smart Grids: A Federated Semisupervised Approach // IEEE Transactions on Industrial Informatics. 2023. vol. 19. no. 1. pp. 995–1005. DOI: 10.1109/TII.2022.3165869.
56. Zhang Y., Fang X., Hordiichuk-Bublivska O., Beshley H., Beshley M. Modified Masking-Based Federated Singular Value Decomposition Method for Fast Anomaly Detection in Smart Grid Systems // Energies. 2023. vol. 16. no. 16. 5996 p. DOI: 10.3390/en16165996.
57. Fenza G., Gallo M., Loia V. Drift-Aware Methodology for Anomaly Detection in Smart Grid // IEEE Access. 2019. vol. 7. pp. 9645–9657. DOI: 10.1109/ACCESS.2019.2891315.
58. Li Y., Bai Y., Yang R., Feng Z., He W. Interpretable adaptive fault detection method for smart grid based on belief rule base // Scientific Reports. 2025. vol. 15. 7646 p. DOI: 10.1038/s41598-025-91897-x.
59. Hu P., Gao W., Li Y., Guo X., Hua F., Qiao L. Anomaly Detection and State Correction in Smart Grid Using EKF and Data Compensation Techniques // IEEE Sensors Journal. 2024. vol. 24. no. 8. pp. 12995–13009. DOI: 10.1109/JSEN.2024.3372973.
60. Gaggero G.B., Girdinio P., Marchese M. Artificial Intelligence and Physics-Based Anomaly Detection in the Smart Grid: A Survey // IEEE Access. 2025. vol. 13. pp. 23597–23606. DOI: 10.1109/ACCESS.2025.3537410.
61. Karimipour H., Geris S., Dehghantanha A., Leung H. Intelligent Anomaly Detection for Large-scale Smart Grids // Proceedings of the IEEE Canadian Conference of Electrical and Computer Engineering (CCECE). 2019. pp. 1–4. DOI: 10.1109/CCECE.2019.8861995.
62. Митяков Е.С. Разработка прототипа цифрового двойника автоматизированной системы управления интеллектуальной энергосетью для анализа угроз информационной безопасности // Computational Nanotechnology. 2025. Т. 12. №4. С. 116–123. DOI: 10.33693/2313-223X-2025-12-4-116-123.
63. Митяков Е.С. Метод обнаружения признаков угроз информационной безопасности объектов критической информационной инфраструктуры на основе цифровых двойников // Computational Nanotechnology. 2025. Т. 12. №3. С. 115–122. DOI: 10.33693/2313-223X-2025-12-3-115-122.
64. Kochergin S.V., Artemova S.V., Bakaev A.A., Mityakov E.S., Vegera Zh.G., Maksimova E.A. Cybersecurity of smart grids: Comparison of machine learning approaches training for anomaly detection // Russian Technological Journal. 2024. vol. 12. no. 6. pp. 7–19. DOI: 10.32362/2500-316X-2024-12-6-7-19.

65. Dong H., Kotenko I. Cybersecurity in the AI era: analyzing the impact of machine learning on intrusion detection // Knowledge and Information Systems. 2025. vol. 67. pp. 3915–3966. DOI: 10.1007/s10115-025-02366-w.
66. Abdi A., Bennouri H., Keane A. Cyber Resilience, Risk Management, and Security Challenges in Enterprise-Scale Cloud Systems: Comprehensive Review // Proceedings of the 13th Mediterranean Conference on Embedded Computing (MECO). 2024. pp. 1–8. DOI: 10.1109/MECO62516.2024.10577956.
67. Guide for Cybersecurity Event Recovery: NIST Special Publication 800-184. Gaithersburg, MD, USA: NIST, 2023. 53 p. URL: https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.800-184.pdf (дата обращения: 06.02.2026).
68. Vielberth M., Bohm F., Fichtinger I., Pernul G. Security Operations Center: A Systematic Study and Open Challenges // IEEE Access. 2020. vol. 8. pp. 227756–227779. DOI: 10.1109/ACCESS.2020.3045514.
69. Kotenko I., Saenko I., Lauta O., Ichetovkin E., Sadovnikov V., Li W. Analysis of Modern Research on Protection against Adversarial Attacks in Energy Systems // Информатика и автоматизация. 2025. Т. 24. №6. С. 1751–1809. DOI: 10.15622/ia.24.6.8.

Котенко Игорь Витальевич — д-р техн. наук, профессор, заслуженный деятель науки российской федерации, главный научный сотрудник, руководитель лаборатории, лаборатория проблем компьютерной безопасности, Санкт-Петербургский Федеральный исследовательский цент Российской академии наук (СПб ФИЦ РАН). Область научных интересов: безопасность компьютерных сетей, управление политиками безопасности, контроль доступа, аутентификация, анализ защищенности, обнаружение компьютерных атак, защита от вирусов и сетевых червей, анализ и верификация протоколов безопасности, анализ и верификация систем информационной безопасности, защита программного обеспечения от взлома, управление цифровыми правами, технологии моделирования и визуализации для противодействия кибертерроризму. 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, Федеральное государственное бюджетное учреждение науки «Санкт-Петербургский Федеральный исследовательский центр Российской академии наук» (СПБ ФИЦ РАН). Область научных интересов: автоматизированные информационные системы, информационная безопасность, искусственный интеллект, машинное обучение, 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.

Митяков Евгений Сергеевич — д-р экон. наук, профессор, заведующий кафедрой, кафедра предметно-ориентированные информационные системы, Федеральное государственное бюджетное образовательное учреждение высшего образования "МИРЭА - Российский технологический университет" (РТУ МИРЭА). Область научных интересов: информационная безопасность, безопасность критической информационной инфраструктуры, моделирование процессов управления инцидентами, экономическая безопасность, математическое моделирование. Число научных публикаций — 250. mityakov@mirea.ru; ул. Стромынка, 20, 107996, Москва, Россия; р.т.: +7(499)600-8080.

Поддержка исследований. Работа выполнена при частичной финансовой поддержке бюджетной темы FFZF-2025-0016.

DOI 10.15622/ia.25.3.8

I. KOTENKO, I. SAENKO, E. MITYAKOV

USING A DIGITAL TWIN FOR CYBER THREAT DETECTION IN INTELLIGENT POWER GRID CONTROL SYSTEMS

Kotenko I., Saenko I., Mityakov E. **Using a Digital Twin for Cyber Threat Detection in Intelligent Power Grid Control Systems.**

Abstract. The digital transformation of the energy sector, accompanied by the widespread deployment of smart grids, expands the cyber attack surface and increases the vulnerability of critical infrastructure. Traditional signature-based security systems demonstrate limited effectiveness against emerging threat types, while their verification on operational facilities entails unacceptable risks. This paper presents and experimentally validates an integrated adaptive framework based on a Digital Twin (DT), aimed at enhancing the effectiveness of cyber threat detection in intelligent power grid control systems. The framework includes a formal mathematical formulation of the detection problem considering three key metrics: Recall, False Positive Rate (FPR), and Total Time to Respond (TTR), together with a loss function for their joint optimization. The key element of the framework is a closed-loop adaptation governed by quality metrics, driven by a three-level mechanism (operational, tactical, and strategic modes): the performance results of the threat detection system on a real-world facility are used to automatically trigger incremental training, optimization, or generation of new scenarios within the secure virtual DT environment. The digital twin is formalized as a tuple of interconnected models of physical processes, communication infrastructure, control logic, and a threat knowledge base, enabling the framework to account for the multi-layer architecture of intelligent power grids and specific protocols (IEC 61850, Modbus). Experimental validation on a synthetic dataset simulating the operation of an industrial control system confirmed the applicability of the concept: the use of synthetic data generated in the DT environment reduced the false positive rate by 6.4% compared to training on static data, while activation of the adaptation mechanism contributed to a reduction in the aggregate model response time to incidents by 3.82%. The computation of the composite loss function showed a 16.8% reduction in the overall quality metric after adaptation. The proposed approach contributes to addressing the challenges of representative data scarcity and constraints on safe testing, and may be considered a step towards implementing the “security by design” paradigm in the energy sector. The obtained results constitute a proof of concept and substantiate the need for further research towards industrial deployment, including validation on real-world industrial scenarios.

Keywords: intelligent power grids, digital twin, adaptive threat detection, cyber-resilience, control systems.

References

1. Gehrmann C., Gunnarsson M. A Digital Twin Based Industrial Automation and Control System Security Architecture. IEEE Transactions on Industrial Informatics. 2020. vol. 16. no. 1. pp. 669–680. DOI: 10.1109/TII.2019.2938885.
2. Venkatachary S.K., Prasad J., Alagappan A., Andrews L.J.B., Raj R.A., Duraisamy S. Cybersecurity and cyber-terrorism challenges to energy-related infrastructures – Cybersecurity frameworks and economics – Comprehensive review. International Journal of Critical Infrastructure Protection. 2024. vol. 45. 100677 p. DOI: 10.1016/j.ijcip.2024.100677.
3. Diaba S.Y., Shafie-khah M., Elmusrati M. Cyber-physical attack and the future energy systems: A review. Energy Reports. 2024. vol. 12. pp. 2914–2932. DOI: 10.1016/j.egyr.2024.08.060.

4. Kotenko I., Saenko I., Lauta O., Kribel A. An Approach to Detecting Cyber Attacks against Smart Power Grids Based on the Analysis of Network Traffic Self-Similarity. Energies. 2020. vol. 13. no. 19. 5031 p. DOI: 10.3390/en13195031.
5. Izrailov K., Buinevich M., Kotenko I. Intelligent Detection of Cyber Attacks on Electrical Power Systems Based on Simulation and Graph-Based Modeling. Proceedings of the International Conference Intelligent Systems (INTELS'24). Communications in Computer and Information Science (CCIS). Springer. 2026. vol. 2603. pp. 231–245. DOI: 10.1007/978-3-032-04758-8_18.
6. Lin H. Computer network information security monitoring system. Proceedings of Second International Conference on Big Data, Computational Intelligence, and Applications (BDCIA 2024). 2025. vol. 13550. DOI: 10.1117/12.3059895.
7. Villegas-Ch W., Govea J., Gutierrez R., Navarro A.M., Mera-Navarrete A. Effectiveness of an Adaptive Deep Learning-Based Intrusion Detection System. IEEE Access. 2024. vol. 12. pp. 184010–184027. DOI: 10.1109/ACCESS.2024.3512363.
8. Kotenko I., Saenko I., Lauta O., Kribel A. A Proactive Protection of Smart Power Grids against Cyberattacks on Service Data Transfer Protocols by Computational Intelligence Methods. Sensors. 2022. vol. 22. no. 19. 7506 p. DOI: 10.3390/s22197506.
9. Desnitsky V.A., Kotenko I.V., Nogin S.B. Detection of Anomalies in Data for Monitoring of Security Components in the Internet of Things. Proceedings of the XVIII IEEE International Conference on Soft Computing and Measurements (SCM). 2015. pp. 189–192. DOI: 10.1109/SCM.2015.7190452.
10. Aydin Z. Detecting Cybersecurity Threats in Digital Energy Systems Using Deep learning for Imbalanced Datasets. International Journal of Energy Economics and Policy. 2025. vol. 15. no. 3. pp. 614–628. DOI: 10.32479/ijeep.19649.
11. Cremer F., Sheehan B., Fortmann M., et al. Cyber risk and cybersecurity: a systematic review of data availability. The Geneva Papers on Risk and Insurance - Issues and Practice. 2022. vol. 47. pp. 698–736. DOI: 10.1057/s41288-022-00266-6.
12. Kotenko I., Chechulin A. Computer attack modeling and security evaluation based on attack graphs. Proceedings of the IEEE 7th International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS). 2013. pp. 614–619. DOI: 10.1109/IDAACS.2013.6662998.
13. Homaei M., Mogollon-Gutierrez O., Sancho J.C., et al. A review of digital twins and their application in cybersecurity based on artificial intelligence. Artificial Intelligence Review. 2024. vol. 57. 201 p. DOI: 10.1007/s10462-024-10805-3.
14. McLaughlin K.L. The power of digital twins in the cybersecurity mesh. EDPACS. 2023. vol. 68. no. 6. pp. 35–39. DOI: 10.1080/07366981.2023.2263214.
15. Krishnaveni S., Chen T.M., Sathiyanarayanan M., et al. CyberDefender: an integrated intelligent defense framework for digital-twin-based industrial cyber-physical systems. Cluster Computing. 2024. vol. 27. pp. 7273–7306. DOI: 10.1007/s10586-024-04320-x.
16. Salim M.M., Camacho D., Park J.H. Digital Twin and federated learning enabled cyberthreat detection system for IoT networks. Future Generation Computer Systems. 2024. vol. 161. pp. 701–713. DOI: 10.1016/j.future.2024.07.017.
17. Sousa B., Arieiro M., Pereira V., Correia J., Lourenco N., Cruz T. ELEGANT: Security of Critical Infrastructures With Digital Twins. IEEE Access. 2021. vol. 9. pp. 107574–107588. DOI: 10.1109/ACCESS.2021.3100708.
18. De Hoz D.J., Temperekidis A., Katsaros P., Konstantinou C. An IoT Digital Twin for Cyber-Security Defence Based on Runtime Verification. Lecture Notes in Computer Science. Springer. 2022. vol. 13701. pp. 556–574. DOI: 10.1007/978-3-031-19849-6_31.
19. Erkek I., Irmak E. Enhancing Cybersecurity of a Hydroelectric Power Plant Through Digital Twin Modeling and Explainable AI. IEEE Access. 2025. vol. 13. pp. 41887–41908. DOI: 10.1109/ACCESS.2025.3547672.

20. Li Y., Guan P., Li T., Larsen K.G., Aiello M., Pedersen T.B., Huang T., Zhang Y. Digital Twin for Secure Peer-to-Peer Trading in Cyber-Physical Energy Systems. IEEE Transactions on Network Science and Engineering. 2025. vol. 12. no. 2. pp. 669–683. DOI: 10.1109/TNSE.2024.3507956.
21. Patel T., Jadav N.K., Rathod T., Tanwar S., Garg D., Shahinzadeh H. AI-based Secure Intrusion Detection Framework for Digital Twin-enabled Critical Infrastructure. Proceedings of the 14th International Conference on Information and Knowledge Technology (IKT). 2023. pp. 24–29. DOI: 10.1109/IKT62039.2023.10433057.
22. Ma J., Guo Y., Fang C., Zhang Q. Digital-Twin-Based CPS Anomaly Diagnosis and Security Defense Countermeasure Recommendation. IEEE Internet of Things Journal. 2024. vol. 11. no. 10. pp. 18726–18738. DOI: 10.1109/JIOT.2024.3366904.
23. Ghenai C., Husein L.A., Nahlawi M.A., Hamid A.K., Bettayeb M. Recent trends of digital twin technologies in the energy sector: A comprehensive review. Sustainable Energy Technologies and Assessments. 2022. vol. 54. 102837 p. DOI: 10.1016/j.seta.2022.102837.
24. Yu W., Patros P., Young B., Klinac E., Walmsley T.G. Energy digital twin technology for industrial energy management: Classification, challenges and future. Renewable and Sustainable Energy Reviews. 2022. vol. 161. 112407 p. DOI: 10.1016/j.rser.2022.112407.
25. Ismail F.B., Al-Faiz H., Hasini H., Al-Bazi A., Kazem H.A. A comprehensive review of the dynamic applications of the digital twin technology across diverse energy sectors. Energy Strategy Reviews. 2024. vol. 52. 101334 p. DOI: 10.1016/j.esr.2024.101334.
26. Hashmi R., Liu H., Yavari A. Digital Twins for Enhancing Efficiency and Assuring Safety in Renewable Energy Systems: A Systematic Literature Review. Energies. 2024. vol. 17. no. 11. 2456 p. DOI: 10.3390/en17112456.
27. Stadtmann F., Rasheed A., Kvamsdal T., et al. Digital Twins in Wind Energy: Emerging Technologies and Industry-Informed Future Directions. IEEE Access. 2023. vol. 11. pp. 110762–110795. DOI: 10.1109/ACCESS.2023.3321320.
28. Semeraro C., Aljaghoub H., Abdelkareem M.A., Alami A.H., Olabi A.G. Digital twin in battery energy storage systems: Trends and gaps detection through association rule mining. Energy. 2023. vol. 273. 127086 p. DOI: 10.1016/j.energy.2023.127086.
29. Wanasinghe T.R., Wroblewski L., Petersen B.K., et al. Digital Twin for the Oil and Gas Industry: Overview, Research Trends, Opportunities, and Challenges. IEEE Access. 2020. vol. 8. pp. 104175–104197. DOI: 10.1109/ACCESS.2020.2998723.
30. Amaral J.V.S.D., Santos C.H.D., Montevechi J.A.B., De Queiroz A.R. Energy Digital Twin applications: A review. Renewable and Sustainable Energy Reviews. 2023. vol. 188. 113891 p. DOI: 10.1016/j.rser.2023.113891.
31. Heluany J.B., Gkioulos V. A review on digital twins for power generation and distribution. International Journal of Information Security. 2023. vol. 23. pp. 1171–1195. DOI: 10.1007/s10207-023-00784-x.
32. Das O., Zafar M.H., Sanfilippo F., Rudra S., Kolhe M.L. Advancements in digital twin technology and machine learning for energy systems: A comprehensive review of applications in smart grids, renewable energy, and electric vehicle optimization. Energy Conversion and Management: X. 2024. vol. 24. 100715 p. DOI: 10.1016/j.ecmx.2024.100715.
33. Ba L., Tangour F., Abbassi I.E., Absi R. Analysis of Digital Twin Applications in Energy Efficiency: A Systematic Review. Sustainability. 2025. vol. 17. no. 8. 3560 p. DOI: 10.3390/su17083560.
34. Koirala B., Cai H., Khayatian F., Munoz E., An J.G., Mutschler R., et al. Digitalization of Urban Multi-Energy Systems – Advances in Digital Twin

Applications across Life-Cycle Phases. Advances in Applied Energy. 2024. vol. 16. 100196 p. DOI: 10.1016/j.adapen.2024.100196.
35. Gourisetti S.N.G., Bhadra S., Sebastian-Cardenas D.J., Touhiduzzaman Md., Ahmed O. A Theoretical Open Architecture Framework and Technology Stack for Digital Twins in Energy Sector Applications. Energies. 2023. vol. 16. no. 13. 4853 p. DOI: 10.3390/en16134853.
36. Boukaf M., Fadli F., Meskin N. A Comprehensive Review of Digital Twin Technology in Building Energy Consumption Forecasting. IEEE Access. 2024. vol. 12. pp. 187778–187799. DOI: 10.1109/ACCESS.2024.3498107.
37. Maksimovic M., Jokic S., Boskovic M.C. Innovative Horizons for Sustainable Smart Energy: Exploring the Synergy of 5G and Digital Twin Technologies. Process Integration and Optimization for Sustainability. 2025. vol. 9. pp. 431–470. DOI: 10.1007/s41660-024-00478-4.
38. Wang Y., Kang X., Chen Z. A survey of Digital Twin techniques in smart manufacturing and management of energy applications. Green Energy and Intelligent Transportation. 2022. vol. 1. no. 2. 100014 p. DOI: 10.1016/j.geits.2022.100014.
39. Masi M., Sellitto G.P., Aranha H., Pavleska T. Securing critical infrastructures with a cybersecurity digital twin. Software and Systems Modeling. 2023. vol. 22. pp. 689–707. DOI: 10.1007/s10270-022-01075-0.
40. Gulyamov S., Akhmedov A., Bazarov S., Ubaydullaeva A., Musaev S., Rodionov A., Odilkhujaev I. Using Digital Twins for Modeling and Testing Cybersecurity Scenarios in Smart Cities. Proceedings of the 6th International Conference on Control Systems, Mathematical Modeling, Automation and Energy Efficiency (SUMMA). 2024. pp. 655–658. DOI: 10.1109/SUMMA64428.2024.10803689.
41. Alhamam N., Rahman M.M.H., Aljughaiman A. A Comprehensive Review on Cybersecurity of Digital Twins Issues, Challenges, and Future Research Directions. IEEE Access. 2023. vol. 13. pp. 45106–45124. DOI: 10.1109/ACCESS.2025.3545004.
42. Coppolino L., Nardone R., Petruolo A., Romano L., Souvent A. Exploiting Digital Twin technology for Cybersecurity Monitoring in Smart Grids. Proceedings of the 18th International Conference on Availability, Reliability and Security. 2023. pp. 1–10. DOI: 10.1145/3600160.3605043.
43. Srivastava A., Liu C.-C., Ştefanov A., Basumallik S., et al. Digital Twins Serving Cybersecurity: More Than a Model: Cybersecurity as a Future Benefit of Digital Twins 2. IEEE Power and Energy Magazine. 2024. vol. 22. no. 1. pp. 61–71. DOI: 10.1109/MPE.2023.3325196.
44. Olivares-Rojas J.C., Reyes-Archundia E., Gutierrez-Gnecchi J.A., et al. Towards Cybersecurity of the Smart Grid Using Digital Twins. IEEE Internet Computing. 2021. vol. 26. pp. 52–57. DOI: 10.1109/MIC.2021.3063674.
45. Sarker I.H., Janicke H., Mohsin A., Gill A.Q., Maglaras L. Explainable AI for cybersecurity automation, intelligence and trustworthiness in digital twin: Methods, taxonomy, challenges and prospects. ICT Express. 2024. vol. 10. pp. 935–958. DOI: 10.1016/j.icte.2024.05.007.
46. Suhail S., Iqbal M., Mclaughlin K. Digital-twin-driven deception platform: vision and way forward. IEEE Internet Computing. 2024. vol. 28. no. 4. pp. 40–47. DOI: 10.1109/MIC.2024.3406188.
47. Alcaraz C., Lopez J. Digital Twin: A Comprehensive Survey of Security Threats. IEEE Communications Surveys and Tutorials. 2022. vol. 24. no. 3. pp. 1475–1503. DOI: 10.1109/COMST.2022.3171465.
48. Holmes D., Papathanasaki M., Maglaras L., Ferrag M.A., Nepal S., Janicke H. Digital Twins and Cyber Security – solution or challenge?. Proceedings of the 2021 6th South-East Europe Design Automation, Computer Engineering, Computer Networks

and Social Media Conference (SEEDA-CECNSM). 2021. pp. 1–8. DOI: 10.1109/SEEDA-CECNSM53056.2021.9566277.
49. Singh A. Enhancing Cybersecurity for Digital Twins: Challenges and Solutions. International Journal of Science and Technology. 2024. vol. 15. no. 4. pp. 1–9. DOI: 10.71097/ijsat.v15.i4.1651.
50. Kumar R., Aljuhani A., Javeed D., Kumar P., Islam S., Islam A.K.M.N. Digital Twins-enabled Zero Touch Network: A smart contract and explainable AI integrated cybersecurity framework. Future Generation Computer Systems. 2024. vol. 156. pp. 191–205. DOI: 10.1016/j.future.2024.02.015.
51. Ravinder M., Kulkarni V. Smart Grid Anomaly Detection Using MFDA and Dilated GRU-based Neural Networks. Smart Grids and Sustainable Energy. 2025. vol. 10. 9 p. DOI: 10.1007/s40866-024-00216-2.
52. Alkuwari A.N., Al-Kuwari S., Albaseer A., Qaraqe M. Anomaly Detection on Smart Grids with Optimized Convolutional Long Short-Term Memory Model. IEEE Access. 2025. vol. 13. pp. 40399–40412. DOI: 10.1109/ACCESS.2025.3547037.
53. Yu L., Zhang X., Du L., Yue L. Anomaly Detection of Cyber Attacks in Smart Grid Communications Based on Residual Recurrent Neural Networks. Security and Privacy. 2025. vol. 8. no. 1. DOI: 10.1002/spy2.498.
54. Jithish J., Alangot B., Mahalingam N., Yeo K.S. Distributed Anomaly Detection in Smart Grids: A Federated Learning-Based Approach. IEEE Access. 2016. vol. 11. pp. 7157–7179. DOI: 10.1109/ACCESS.2023.3237554.
55. Abdel-Basset M., Moustafa N., Hawash H. Privacy-Preserved Generative Network for Trustworthy Anomaly Detection in Smart Grids: A Federated Semisupervised Approach. IEEE Transactions on Industrial Informatics. 2023. vol. 19. no. 1. pp. 995–1005. DOI: 10.1109/TII.2022.3165869.
56. Zhang Y., Fang X., Hordiichuk-Bublivska O., Beshley H., Beshley M. Modified Masking-Based Federated Singular Value Decomposition Method for Fast Anomaly Detection in Smart Grid Systems. Energies. 2023. vol. 16. no. 16. 5996 p. DOI: 10.3390/en16165996.
57. Fenza G., Gallo M., Loia V. Drift-Aware Methodology for Anomaly Detection in Smart Grid. IEEE Access. 2019. vol. 7. pp. 9645–9657. DOI: 10.1109/ACCESS.2019.2891315.
58. Li Y., Bai Y., Yang R., Feng Z., He W. Interpretable adaptive fault detection method for smart grid based on belief rule base. Scientific Reports. 2025. vol. 15. 7646 p. DOI: 10.1038/s41598-025-91897-x.
59. Hu P., Gao W., Li Y., Guo X., Hua F., Qiao L. Anomaly Detection and State Correction in Smart Grid Using EKF and Data Compensation Techniques. IEEE Sensors Journal. 2024. vol. 24. no. 8. pp. 12995–13009. DOI: 10.1109/JSEN.2024.3372973.
60. Gaggero G.B., Girdinio P., Marchese M. Artificial Intelligence and Physics-Based Anomaly Detection in the Smart Grid: A Survey. IEEE Access. 2025. vol. 13. pp. 23597–23606. DOI: 10.1109/ACCESS.2025.3537410.
61. Karimipour H., Geris S., Dehghantanha A., Leung H. Intelligent Anomaly Detection for Large-scale Smart Grids. Proceedings of the IEEE Canadian Conference of Electrical and Computer Engineering (CCECE). 2019. pp. 1–4. DOI: 10.1109/CCECE.2019.8861995.
62. Mityakov E.S. [Development of a prototype of a digital twin of an automated smart grid control system for analyzing information security threats]. Komp'yuternye nanotekhnologii – Computational Nanotechnology. 2025. vol. 12. no. 4. pp. 116–123. DOI: 10.33693/2313-223X-2025-12-4-116-123. (In Russ.).
63. Mityakov E.S. [Method for detecting signs of information security threats to critical information infrastructure objects based on digital twins]. Komp'yuternye

nanotekhnologii – Computational Nanotechnology. 2025. vol. 12. no. 3. pp. 115–122. DOI: 10.33693/2313-223X-2025-12-3-115-122. (In Russ.).
64. Kochergin S.V., Artemova S.V., Bakaev A.A., Mityakov E.S., Vegera Zh.G., Maksimova E.A. Cybersecurity of smart grids: Comparison of machine learning approaches training for anomaly detection. Russian Technological Journal. 2024. vol. 12. no. 6. pp. 7–19. DOI: 10.32362/2500-316X-2024-12-6-7-19.
65. Dong H., Kotenko I. Cybersecurity in the AI era: analyzing the impact of machine learning on intrusion detection. Knowledge and Information Systems. 2025. vol. 67. pp. 3915–3966. DOI: 10.1007/s10115-025-02366-w.
66. Abdi A., Bennouri H., Keane A. Cyber Resilience, Risk Management, and Security Challenges in Enterprise-Scale Cloud Systems: Comprehensive Review. Proceedings of the 13th Mediterranean Conference on Embedded Computing (MECO). 2024. pp. 1–8. DOI: 10.1109/MECO62516.2024.10577956.
67. Guide for Cybersecurity Event Recovery: NIST Special Publication 800-184. Gaithersburg, MD, USA: NIST, 2023. 53 p. Available at: https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.800-184.pdf (accessed 06.02.2026).
68. Vielberth M., Bohm F., Fichtinger I., Pernul G. Security Operations Center: A Systematic Study and Open Challenges. IEEE Access. 2020. vol. 8. pp. 227756–227779. DOI: 10.1109/ACCESS.2020.3045514.
69. Kotenko I., Saenko I., Lauta O., Ichetovkin E., Sadovnikov V., Li W. [Analysis of Modern Research on Protection against Adversarial Attacks in Energy Systems]. Informatika i avtomatizatsiya – Informatics and Automation. 2025. vol. 24. no. 6. pp. 1751–1809. DOI: 10.15622/ia.24.6.8. (In Russ.).

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, security policy management, access control, authentication, security analysis, detection of computer attacks, protection against viruses and network worms, analysis and verification of security protocols, analysis and verification of information security systems, software protection against hacking, digital rights management, modeling and visualization technologies for countering cyberterrorism. The number of publications — 1000. ivkote@comsec.spb.ru; 39, 14th line V.O., 199178, Saint-Petersburg, Russia; office phone: +7(812)328-7181; fax: +7(812)328-4450.

Saenko Igor — Ph.D., Dr. Sci., Professor, Chief scientific officer, 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, artificial intelligence, machine learning, modeling theory and mathematical statistics, information theory. The number of publications — 500. ibsaen@comsec.spb.ru; 39, 14th line V.O., 199178, Saint-Petersburg, Russia; office phone: +7(812)328-7181; fax: +7(812)328-4450.

Mityakov Evgeniy — Ph.D., Dr. Sci., Professor, Head of the Department, Department of Subject-Oriented Information Systems, Federal State Budgetary Educational Institution of Higher Education "MIREA - Russian Technological University" (RTU MIREA). Research interests: information security, security of critical information infrastructure, modeling of incident management processes, economic security, mathematical modeling. The number of publications — 250. mityakov@mirea.ru; 20, Stromynka St., 107996, Moscow, Russia; office phone: +7(499)600-8080.

Acknowledgements. The reported study was partially funded by the budget project FFZF-2025-0016.

УДК 004.056 **DOI** 10.15622/ia.25.3.9

И.Б. Саенко , Е.С. Новикова , А.В. Мелешко , В.Е. Садовников

ОНТОЛОГИЧЕСКИЙ ПОДХОД К УПРАВЛЕНИЮ РИСКАМИ В СИСТЕМЕ УПРАВЛЕНИЯ ОЧИСТНЫМИ СООРУЖЕНИЯМИ

Саенко И.Б., Новикова Е.С., Мелешко А.В., Садовников В.Е. **Онтологический подход к управлению рисками в системе управления очистными сооружениями.**

Аннотация. Современные системы управления технологическими процессами на критически важных объектах инфраструктуры, включая очистные сооружения, всё чаще становятся мишенью сложных кибератак, способных нарушить экологическую безопасность и устойчивость работы. Особенность задачи управления рисками в промышленных киберфизических системах (КФС) заключается в необходимости учитывать высокий уровень разнородности устройств, наличие данных об уязвимостях в программно-аппаратном обеспечении промышленного оборудования. Возникает необходимость в определении подхода, который учитывает поведение сложных КФС, особенности атакующих воздействий на КФС и последствия нарушения технологических процессов КФС. В статье предлагается подход к моделированию киберугроз КФС на примере очистных сооружений, в основе которого лежит онтологическая модель, ориентированная на оценку киберугроз для КФС. Модель связывает данные об уязвимостях и атаках со спецификой технологических процессов, включая экологические последствия нарушения целостности процесса очистки воды. Ее применение позволяет анализировать киберриски путем определения возможных информационные угроз с учетом последствий их реализации. Применение разработанной онтологии демонстрируется путем исследования атак, определенных для макета системы управления очистки воды флотационным методом.

Ключевые слова: информационная безопасность, семантическое моделирование угроз, матрица угроз MITRE ATT&CK, системы управления технологическими процессами, водоочистные сооружения.

1. Введение. Современные системы управления технологическими процессами на критически важных объектах инфраструктуры, к которым относятся комплексы очистных сооружений, всё чаще подвергаются киберугрозам, способным нарушить их стабильную работу и привести к серьёзным последствиям экологического характера [1]. В условиях растущей цифровизации и интеграции промышленных систем с корпоративными сетями и облачными платформами традиционные методы защиты, основанные на периметровой безопасности и сигнатурном обнаружении, оказываются недостаточными для противодействия сложным, многоэтапным и целенаправленным атакам. Это требует перехода к более гибким и адаптивным подходам к обеспечению информационной безопасности и киберустойчивости. В системах очистных сооружений, где используются датчики, контроллеры, исполнительные механизмы и программные комплексы автоматизации, каждое звено может стать точкой проникновения для злоумышленника. При этом нарушение работы даже одного компонента, например,

системы дозирования реагентов или насосной станции, может привести к сбоям в процессе очистки, превышению нормативов сброса загрязняющих веществ или полной остановке объекта. Для построения эффективной системы управления информационными рисками необходима систематизация знаний о возможных угрозах безопасности КФС, последствиях от реализации данных угроз и о существующих мерах противодействия им. Одним из подходов к систематизации знаний предметной области является онтологическое моделирование. Онтологии позволяют интегрировать разнородную информацию о киберугрозах, уязвимостях, технических активах, мерах защиты и инцидентах, создавая единое информационное пространство для анализа киберрисков и принятия решений по противодействию им. Особенно актуальна такая интеграция в сложных промышленных системах, где необходимо учитывать как кибербезопасность, так и функциональную устойчивость, надёжность и безопасность технологических процессов с учетом их динамической природы [2, 3].

В данной статье предлагается подход к управлению рисками информационной безопасности в автоматизированных системах управления водоочистными сооружениями, отличающийся построением семантической модели, которая определяет ключевые сущности системы управления водоочистными сооружениями, связывает их с возможными атаками и позволяет формализовать возможные последствия нарушения ее кибербезопасности. Данная модель позволяет исследовать риски информационной безопасности, выявляя критические (уязвимые) элементов системы и обосновывая выбор контрмер с учетом возможных последствий, в т.ч. экологических. Кроме того, анализ релевантных исследований показал, что задача построения подобной семантической модели для систем управления водоочистными сооружениями другими исследователями ранее не решалась.

Работа структурирована следующим образом. В разделе 2 проанализированы релевантные работы в области применения онтологий для решения задач информационной безопасности, включая задачи управления киберрисками. В разделе 3 дана общая схема системы управления очистными сооружениями, приводится описание с точки зрения технической реализации подобных сооружений, включающее описание возможных датчиков и других важных подсистем, таких как подсистема сбора данных для мониторинга состояния объекта и подсистема обновлений программного обеспечения. В разделе 4 представлена разработанная онтология, в разделе 5 приведен сценарий практического применения онтологии для анализа рисков на примере

макета системы очистки воды методом флотации. В разделе 6 даны выводы и определено дальнейшее направление работ.

2. Анализ релевантных работ. В последние годы наблюдается рост исследований, связанных с применением онтологий для решения задач в области информационной безопасности и управления киберрисками. Широко применяются онтологии, обобщающие и систематизирующие знания в области кибербезопасности. Примером такой онтологии служит онтология UCO (Unified Cyber Ontology, UCO), цель которой – определить базу для стандартизированного представления информации об объектах кибербезопасности. В частности, она определяет такие понятия, как кибербразведка, защита компьютеров/сетей, анализ угроз, анализ вредоносных программ, исследование уязвимостей и т.д., устанавливая между ними связи. Определенные в ней абстрактные классы могут быть использованы для детализации каждой предметной области по отдельности. Например, онтология CASE (Cyber-investigation Analysis Standard Expression) [4], являющаяся расширением онтологии UCO, определяет область цифровой криминалистики и формирует основу для автоматизации процессов расследования киберпреступлений, формализуя такие понятия, как наблюдаемые события и артефакты, источники данных, позволяя ответить на вопросы расследования (кто, когда, как долго, где). В [5] предложена онтология, построенная на основе словаря языка STIX [6] и связывающая такие сущности, как атакующие техники и тактики, атака, кампания, инструменты, уязвимости, атакующие и т.д. Ее применение позволяет по используемым атакующим техникам и инструментам определить потенциальных злоумышленников и объединить их в кампанию.

Таксономия контрмер представлена в онтологии D3FEND [7], она связывает технические меры по обеспечению информационной безопасности с информационными ресурсами разного типа. В [8] представлена онтология, объединяющая знания, представленные в таких базах знаний, как MITRE ATT&CK [9], D3FEND [10], CWE [11] и CVE [12]. Таким образом, ее ключевыми элементами являются тактики и техники APT (MITRE ATT&CK), уязвимости (CVE), слабые места (CWE), контрмеры (D3FEND), инструменты анализа и сценарии угроз. Данная онтология предназначена для формирования целостного представления о киберугрозах и упрощения выбора соответствующих контрмер.

Дойникова Е., и др. [13] предложили онтологию метрик, используемых для оценки рисков нарушения информационной безопасности. Разработанная онтология определяет иерархию метрик

безопасности и задает связи между источниками данных для их расчета и применяемыми алгоритмами расчета. Калеги М., и др. [14] дополнили модель метрик контекстом их использования, который определяется через такие концепты, как решаемая задача, время, целевая аудитория, положение наблюдаемых событий и ресурсов в системе, а также вычислительные характеристики системы.

В [15] описана семантическая модель безопасности приложений, которая формализует такие сущности, как сетевые настройки, конфигурации приложений, методы и инструменты тестирования и связывает их с показателями безопасности, что позволяет оценить степень защищенности критически важного приложения в контексте его положения в инфраструктуре организации.

В последнее время фокус исследований сместился на разработку онтологий, которые связывают в единое целое информационные системы, наблюдаемые инциденты безопасности и процессы реагирования для автоматизации данных процессов. В [3, 16] представлена онтология, которая определяет формальную базу для расчета уровня риска информационной безопасности в режиме, близком к реальному времени. В ее основе лежит онтологическое представление правил STIX и рекомендаций DRM (Disaster Risk Management), которые определяют такие понятия, как риск, оценка риска, вероятность риска и его критичность, контрмеры.

В [17—20] представлены онтологии, которые предназначены для управления рисками с учетом особенностей различных предметных областей. Например, представленная в [17] онтология разработана с учетом рекомендаций стандарта по кибербезопасности разработки дорожных транспортных средств ISO/SAE 21434:2021. Причем авторы показали, что стандарт не в полной мере учитывает практические особенности реализации контрмер. Ключевой задачей разработанной онтологии является устранение данного недостатка путем определения связей между компонентами процессов оценки рисков информационной безопасности, возникающих в процессе разработки транспортных средств, при взаимодействии различных его участников. В ее основе лежит формализованное представление методологии моделирования угроз и выбора контрмер TARA (Threat Analysis and Risk Assessment), сценариев ущерба в части нарушения физической безопасности, конфиденциальности пользователей, функциональных компонент транспортных средств и событий безопасности. В [18] предпринята попытка совместить риски нарушения кибербезопасности и физической безопасности, возникающие в промышленных информационных системах

на уровне построения гибридной онтологической модели оценки рисков, включающей элементы, относящиеся к киберрискам, физическому ущербу и функциональным элементам системы. Применимость разработанной модели была продемонстрирована на примере системы управления охлаждением резервуара с ядерным топливом.

В [19] обсуждаются особенности семантического моделирования информационных угроз, характерных для медицинских устройств. Разработанная К.С. Буджио и др. онтология отвечает на вопросы о типе медицинского устройства, его разработчике, используемом программном обеспечении, об уязвимостях, связанных с ним, а также о потенциально уязвимых группах пользователей и возможных последствиях реализованных атак. В [20] рассмотрены особенности атак на системы, использующие модели искусственного интеллекта, и механизмы, предложенные для повышения их уровня защищенности. Представленная онтология основана на базах знаний MITRE ATT&CK и ATLAS, тактики и техники MITRE ATT&CK дополнены новыми угрозами, связанными с уязвимостями машинного обучения.

Для полноты обзора релевантных работ следует отметить исследование [21], в котором представлена онтология, предназначенная для структурирования информации о наборах данных, которые могут быть использованы при обучении моделей машинного обучения для решения задач информационной безопасности. Данная онтология не только позволяет связать различные наборы данных, но и систематизирует типы атак, которые используются для их разметки.

Таким образом, в научной литературе представлены разнообразные онтологии для управления информационной безопасностью, однако отсутствуют решения, которые связывают описание разнородных киберфизических промышленных систем, наблюдаемых аномалий, включая инциденты безопасности, расчет рисков и выбор контрмер в зависимости от последствий их реализации. Также следует отметить, что в научной литературе не представлены решения для семантического моделирования угроз в системе управления очистными сооружениями. Таким образом, данная работа направлена решение выявленной проблемы и представляет онтологию, отличающуюся возможностью анализировать киберриски в контексте возможных экологических последствий.

3. Структура системы управления процессом очистки воды. Рассматриваемая в рамках данной работы система управления процессами очистки воды представляет собой автоматизированную технологическую платформу, объединяющую физические компоненты обработки воды

с цифровыми системами управления, мониторинга и внешними сервисами. Ее структура представлена на рисунке 1.

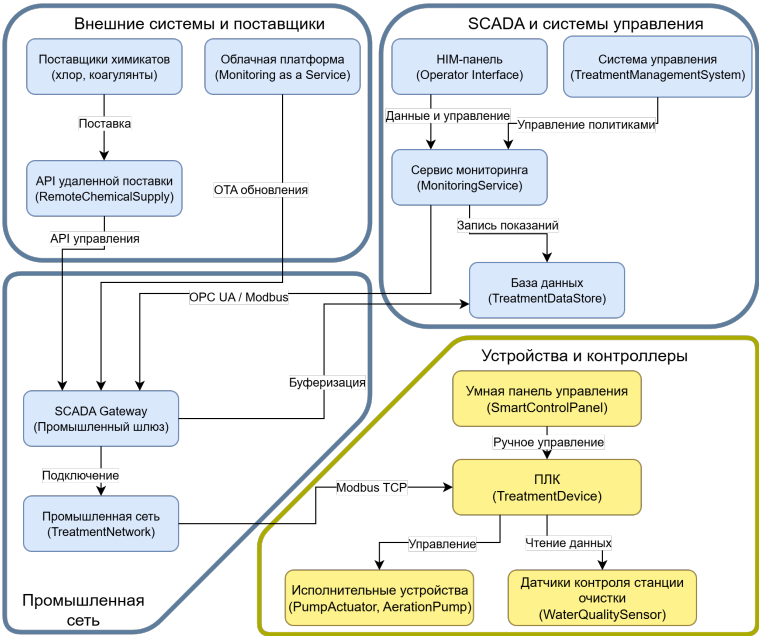


Рис. 1. Структурная схема системы управления водоочистными сооружениями

Центральным элементом системы является программируемый логический контроллер (ПЛК), который отвечает за выполнение алгоритмов управления процессами: запуск и остановку насосных станций, регулирование режимов аэрации, а также сбор данных с датчиков качества воды.

Для обеспечения точного контроля и автоматизации используется SCADA (Supervisory Control and Data Acquisition) система. Она включает в себя интерфейс оператора (HMI), через который осуществляется визуализация текущего состояния системы, а также программные управляющие сервисы, выполняющие сбор данных, их анализ и архивирование. Эта информация позволяет отслеживать ключевые показатели качества воды, выявлять отклонения и оптимизировать работу системы. База данных системы хранит всю историю работы: показания датчиков, параметры управления, события аварий и плановые операции, что позволяет проводить глубокий анализ производительности,

прогнозировать отказы оборудования и выстраивать стратегии технического обслуживания.

Связь между оборудованием и системой управления обеспечивается через промышленный шлюз (SCADA Gateway), который работает как мост между сетью оборудования и программными системами. Он поддерживает стандартные промышленные протоколы, такие как Modbus TCP и OPC UA, что гарантирует совместимость с различными типами оборудования. Шлюз также отвечает за безопасную передачу данных, буферизацию информации и защиту от перегрузок сети.

Современные системы управления водоочистными сооружениями также могут быть связаны с внешними поставщиками ресурсов, например, химикатов, необходимых для обеззараживания. Она может поддерживать OTA-обновления (over-the-air обновление) от облачной платформы, что обеспечивает постоянное совершенствование программного обеспечения без необходимости физического доступа к оборудованию.

4. Проектирование онтологии SecWat-O. Для управления рисками информационной безопасности разрабатываемая онтология должна уметь отвечать на следующие группы вопросов:

1. Какие атаки могут быть определены для системы управления очистными сооружениями?

2. Какие атакующие техники и тактики из базы знаний MITRE ATT&CK определены для заданной системы управления очистными сооружениями?

3. Для каких логических и физических компонентов заданной системы управления определены тактики и техники базы знаний MITRE ATT&CK?

4. Каковы последствия от реализации заданной угрозы?

5. Что является последствием компрометации заданного устройства?

6. Какая атака может привести к его компрометации?

Данные вопросы представляют собой вопросы компетентности онтологии SecWat-O. Они обуславливают необходимость формализовать следующие компоненты системы:

1. Физические и логические элементы системы управления водоочистных сооружений;

2. Информационно-коммуникационную среду, с помощью которой SCADA взаимодействует с системой очистных сооружений;

3. Атакующие техники и тактики, определенные для промышленных киберфизических систем;

4. Угрозы и последствия, возникающие в результате их реализации.

При проектировании онтологии был применен модульный подход, который позволяет импортировать и повторно использовать созданные ранее онтологии, и, как следствие, повышает совместимость разрабатываемой онтологии с существующими решениями в данной предметной области. Кроме того, такое решение дает возможность вести дальнейшую проработку компонентов независимо друг от друга и повторно использовать разработанную онтологию и/или ее отдельные компоненты в других проектах. На рисунке 2 представлена упрощенная схема основных компонент (модулей) разработанной онтологии SecWaT-O и связей между ними. Цветом выделены модули, отвечающие за моделирование физических и логических элементов системы управления водоочистных сооружений, атакующих воздействий и их последствий, и информационно-коммуникационную среду. Далее эти модули рассмотрены более подробно.

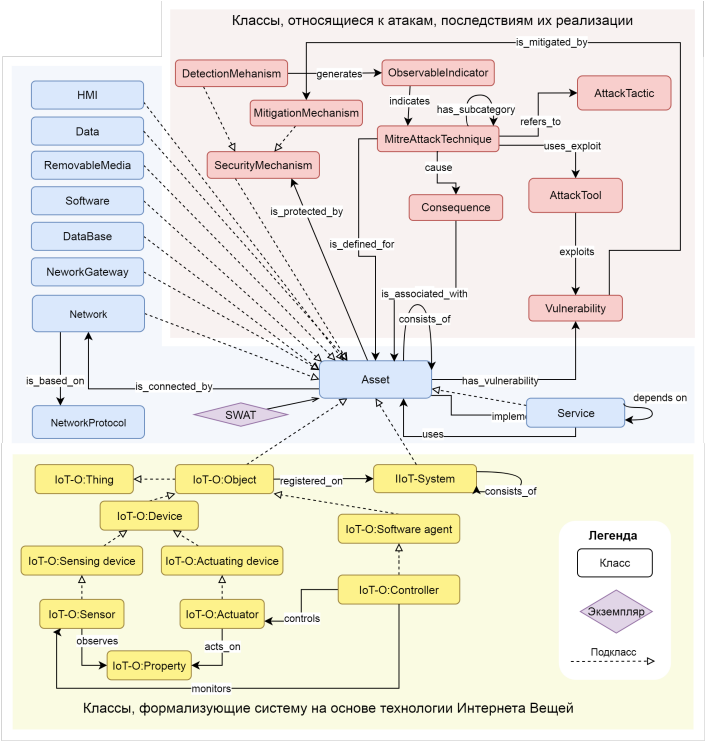


Рис. 2. Основные концепты онтологии SecWaT-O и связи между ними

4.1. Моделирование функционирования системы очистки воды. В качестве базы для моделирования устройств системы очистных сооружений была выбрана онтология IoT-O [22], которая определяет киберфизические системы через совокупность взаимодействующих устройств и выполняемые ими функции. Она определяет ключевые классы *Устройство* (Device) и *Программный агент* (Software Agent), которые формализуют понятия физических и логических элементов киберфизической системы. Класс "*Устройство*"(Device) имеет два подкласса, характеризующие два типа устройств:

– датчики (Sensing_Device) – устройства, которые собирают данные;

– актуаторы (Actuating_device) – устройства, взаимодействующие с системой.

Данные устройства управляются контроллером и в совокупности они формируют некоторую программно-аппаратную систему, выполняющую некоторую функцию или предоставляющую некоторый сервис.

В рамках общей онтологии SecWat-O (Рис. 2) реализуется импортирование онтологии IoT-O, классы которой формализуют систему на физическом уровне (на рис. 2 выделены желтым цветом).

В ходе построения онтологии для системы управления процессом флотационной очистки воды было специфицировано следующее аппаратное технологическое оборудование:

– датчики контроля состояния воды, к которым относятся датчики потока воды, датчики мутности воды, датчики уровня воды, датчики уровня pH, датчики взвешенных частиц, манометры;

– актуаторы, к которым относятся насосы подачи воды, краны, также регулирующие подачу воды, насосы подачи реагентов, мешалки, мотор скребка для шлама, компрессор для аэрации.

Координация данных устройств осуществляется с помощью общего контроллера управления всей станцией. На рисунке 3 представлена онтологическая модель компонентов технологического процесса флотационной очистки воды. Были определены следующие классы устройств:

– класс DensityMeter – датчик мутности воды, измеряющий плотность взвеси в воде;

– класс WaterLevelMeter – датчик уровня воды;

– класс WaterFlowMeter – датчик скорости потока;

– класс PHMeter – датчик уровня кислотности воды;

– класс PressureGauge – манометр;

– класс PumpActuator – исполнительный механизм насоса для подачи воды или подачи реагентов;

– класс VentActuator – управляемый кран насоса подачи воды;

– класс StirerActuator – исполняющий механизм мешалок, выполняющих перемешивание загрязненной воды;

– класс AirCompressor – компрессор для аэрации загрязненной воды;

– класс ScrapperMotor – мотор скребка для удаления шлама;

– класс WaterTreatmentNode – технологическое звено процесса, выполняющее определенную функцию, например, флотацию или флокулляцию.

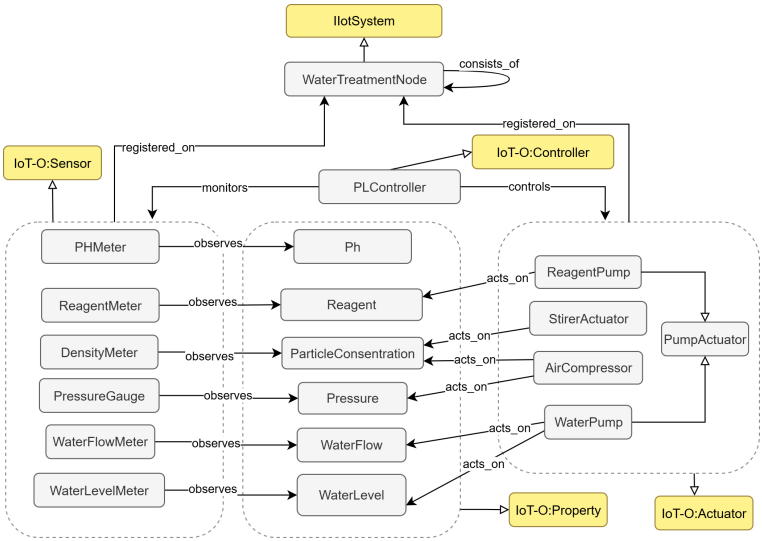


Рис. 3. Классы онтологии SecWaT-O, отражающие связи между технологическим оборудованием

Технологическая часть системы управления очисткой воды обычно интегрирована в информационную систему предприятия. Ресурсы информационной системы описываются через абстрактный класс *Ресурс* (Asset). Данный класс имеет такие подклассы, как *Сеть* (Network), *Хранилище данных* (DataBase), *Программное обеспечение* (Software), *Данные* (Data), *Человеко-машинные интерфейсы* (HMI), *Съемные носители данных* (Removable Media), *Механизм защиты* (Security Mechanism). Технологическая часть системы, состоящая

из промышленных датчиков и актуаторов, также является подклассом данного класса и представлена классом IoTSystem.

С ресурсами информационной системы связано понятие выполняемой функции, которое в разработанной онтологии определяется через класс *Функция* (Service). Примером такой функции может служить контроль мутности воды или очистка воды методом флотации, или удаленное обновление ПО. Для ее реализации может быть использован один и более ресурсов. Объекты класса *Ресурс* могут иметь некоторую уязвимость, которая может быть использована для реализации атакующего воздействия. Также ресурсы могут быть связаны с другими ресурсами с помощью коммуникационных сетей, которые, в свою очередь, также относятся к ресурсам. Таким образом, формально класс ресурс (Asset) определяется следующим образом:

$$\begin{aligned} Asset \sqsubseteq (&Thing \\ &\sqcap \forall implements.Service \\ &\sqcap \forall has\_vulnerability.Vulnerability \\ &\sqcap \forall is\_connected\_by.Network \\ &\sqcap \forall is\_protected\_by.SecurityMechanism \\ &\sqcap \forall consists\_of.Asset) \end{aligned} \tag{1}$$

На рисунке 2 показаны классы, относящиеся к информационно-коммуникационной части системы управления, отмечены синим цветом.

4.2. Моделирование кибер угроз и атак на систему очистных сооружений. В основе семантической модели угроз и атак лежит база знаний MITRE ATT&CK, которая систематизирует практические знания об атаках, техниках их выполнения, включая используемые уязвимости, а также возможные контрмеры. На основе данной базы знаний была разработана отдельная онтология MITRE-ATTACK, которая в дальнейшем импортируется в онтологию SecWat-O. Были выделены следующие классы:

– класс AttackTactic – цели атакующего;

– класс AttackTechnique – конкретные атакующие техники, которые использует злоумышленник для достижения своих целей, данный класс связан с абстрактным классом *Ресурс* (Asset), что позволяет учитывать техники, специфические для промышленных систем управления;

– класс Vulnerability – уязвимости, определенные для ресурса;

– класс AttackTool – инструменты, используемые для выполнения атакующего воздействия, включая эксплойты;

– класс ObservableIndicator – наблюдаемые индикаторы, которые сигнализируют о выполнении атаки;

– класс SecurityMechanism – механизм обеспечения кибербезопасности, он имеет два подкласса – механизм предотвращения атаки (MitigationMechanism) и механизм обнаружения атаки (Detection Mechanism), который генерирует наблюдаемые индикаторы атаки.

Для описания возможных последствий, возникающих вследствие выполнения атаки на заданный информационный ресурс, был определен класс *Последствия* (Consequence). Последствия задаются для скомпрометированных ресурсов. Поскольку ресурсы связаны между собой семантическими связями, имеется возможность построить цепочку связанных устройств, которые могут быть косвенно затронуты в результате выполняемых деструктивных действий. Классы и функциональные связи между ними, относящиеся к моделированию действий атакующего (на рис. 2 выделены красным цветом.

В результате разработанная онтология SecWat-O определяет элементы, связывающие две предметные области Интернета Вещей и атаки, показывая инфокоммуникационные связи между объектами, формализует понятие ресурс и бизнес-процессы. При ее построении осуществляется импортирование онтологий MITRE-ATTACK и IoT-O, а также их объединение в единую онтологию, что реализует модульный принцип построения.

В качестве инструмента разработки онтологии SecWat-O был выбран Protege. Проверка онтологии на непротиворечивость и наличие циклов была выполнена с помощью машины логического вывода HermiT 1.4.3. Проверка на полноту осуществлялась путем построения логических запросов на языке SPARQL, цель которых получить ответы на вопросы компетентности данной онтологии, определенных в начале раздела. Процесс построения запросов описан в следующем разделе.

5. Пример использования: моделирование атак на систему флотационной очистки воды. Для проверки полноты разработанной онтологии SecWaT-O была построена семантическая модель для макета системы управления очистки воды флотационным методом, представленного учебным стендом CE 587 (G.U.N.T., GmbH) [23].

Он оснащен следующими датчиками и исполнительными устройствами (Рис. 4):

– насос P1, установленный перед резервуаром для флокуляции, используемый для подачи неочищенной воды;

– датчик уровня воды LM1 в резервуаре для флотации;

– циркуляционный насос P2 перед воздушным барабаном, используемый для реализации системы аэрации резервуара для флотации;

– расходомер воды FM1, установленный после насоса P1;

– управляемый кран VA1, установленный после насоса P1;

– манометр PG1, установленный после воздушного компрессора AirC1;

– датчик скорости подачи воздуха FM2 перед воздушным барабаном, используемый для контроля подачи воздуха в барабан;

– датчик уровня воды LM2 в воздушном барабане для контроля работы система аэрации;

– манометр PG2 перед резервуаром флотации для контроля системы аэрации;

– датчик потока FM3 перед резервуаром флотации для контроля потока водно-воздушной смеси системы аэрации резервуара;

– датчик контроля уровня кислотности pH: датчик кислотности PH1;

– насосы подачи регентов P4, P4, P5, и механизмы для перемешивания воды в резервуаре флокуляции R1-R4.

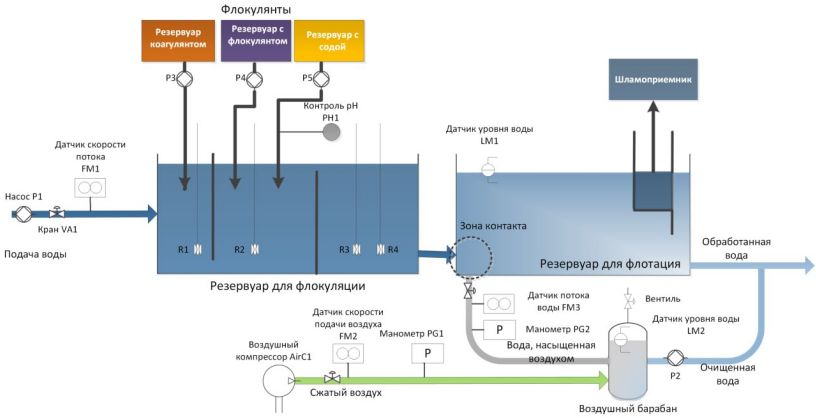


Рис. 4. Схема учебного стенда CE 587 с указанием установленных датчиков и актуаторов [23]

Рассматриваемый стенд не имеет подключения к сети Интернет, поэтому были рассмотрены сценарии атак с внутренним нарушителем:

– подключение съемного диска с вредоносным ПО к рабочей станции;

– сотрудник настроил систему таким образом, что процесс очистки нарушен;

– злоумышленник подключился к внутренней сети напрямую.

Нарушение целостности технологического процесса осуществляется путем подмены передаваемых значений сенсоров и/или команд актуаторов или путем их модификации. Согласно таксономии MITRE ATT&CK, атакующие техники, манипулирующие данными от таких устройств, относятся к тактикам *Ухудшение контроля технологического процесса* (Impair Process Control, ID: TA0106) и *Подавление ответной функции* (Inhibit Response Function, ID: TA0107).

Тактика *Ухудшение контроля технологического процесса* (ID: TA0106) состоит из методов, целью которых является нарушение логики управления и оказания определяющего воздействия на процессы, выполняемые системой. Ключевым объектом атак становится программно-логический контроллер, который непосредственно взаимодействуют с объектами физического процесса и определяет логику их функционирования. В группу техник, относящихся к тактике *Подавление ответной функции* (ID: TA0107), входят методы, целью которых является отключение функций обеспечения безопасности, защиты, контроля качества или предотвращение вмешательства оператора. Они направлены на активное выявление и предотвращение ожидаемых аварийных сигналов и ответных мер, возникающих в ответ на получаемые данные или команды от исполнительных устройств.

В таблице 1 предоставлены возможные последствия, определенные для учебного стенда CE 587 и возникающие в результате манипуляций с его датчиками.

Наиболее критичным узлом учебного стенда является контроллер. На текущий момент для атак, модифицирующих логику контроллера, специфических контрмер нет (объект онтологии – Mitigation Limited or Not Effective, ID: M 0816). Таким образом, защита таких ресурсов должна включать такие контрмеры, как контроль доступа к ресурсам информационно-коммуникационной системы в целом, применение механизмов аутентификации, формирование списков доступа, сегментацию компьютерной сети и т.д.

Таблица 1. Атаки на датчики и их возможные последствия на технологический процесс

Компонент стенда	**Датчик**	**Последствия**
Резервуар флокуляции	Кран подачи воды VA1	– повреждение насоса P1 (работа в сухом режиме); – переполнение/осушение резервуара флокуляции; – утечка загрязненной воды в результате переполнения резервуара флокуляции; – попадание неочищенной воды в водоемы.
	Насос P1	– переполнение/осушение резервуара флокуляции; – утечка загрязненной воды в результате переполнения резервуара флокуляции; – попадание неочищенной воды в водоемы.
	Расходомер воды FM1	– переполнение/осушение резервуара флокуляции; – утечка загрязненной воды в результате переполнения резервуара флокуляции; – попадание неочищенной воды в водоемы.
	Насосы подачи регентов P4, P4, P5	– избыточное/недостаточное количество флокулянтов и каугулянтов нарушает процесс очистки воды; – очищамая вода загрязняется флокулянтами/каугулянтами; – загрязненная вода попадает в водоемы.

Продолжение таблицы 1. Атаки на датчики и их возможные последствия на технологический процесс

Компонент стенда	**Датчик**	**Последствия**
	Механизмы перемешивания R1-R4	– нарушение процесса флокуляции, образуется меньшее число флокулянтов; – ухудшается качество очищаемой воды; – загрязненная вода попадает в водоемы.
	Датчик уровня кислотности PH1	– изменяются параметры процесса флокуляции, зависящие от уровня кислотности воды; – ухудшается качество очищаемой воды; – загрязненная вода попадает в водоемы.
Резервуар флотации	Датчик уровня воды LM1	– переполнение/осушение резервуара флотации; – утечка загрязненной воды в результате переполнения резервуара флокуляции; – попадание неочищенной воды в водоемы.
	Манометр PG1	– нарушение работы системы аэрации; – избыточная/недостаточная подача воздуха; – недостаточное удаление шлака из очищаемой воды, переполнения резервуара флокуляции; – попадание неочищенной воды в водоемы.

Продолжение таблицы 1. Атаки на датчики и их возможные последствия на технологический процесс

Компонент стенда	**Датчик**	**Последствия**
	Датчик уровня воды LM2	– нарушение работы системы аэрации; – избыточная/недостаточная подача воды; – недостаточное удаление шлака из очищаемой воды, переполнения резервуара флокуляции; – попадание неочищенной воды в водоемы; – выдавливание воды из воздушного барабана.
	Насос P2	– нарушение работы системы аэрации; – избыточная/недостаточная подача воды; – недостаточное удаление шлама из очищаемой воды; – переполнение резервуара флотации; – попадание неочищенной воды в водоемы; – выдавливание воды из воздушного барабана.
	Манометр PG2	– нарушение работы системы аэрации; – избыточная/недостаточная подача воды; – недостаточное удаление шлама из очищаемой воды; – переполнение резервуара флотации; – попадание неочищенной воды в водоемы.

Продолжение таблицы 1. Атаки на датчики и их возможные последствия на технологический процесс

Компонент стенда	**Датчик**	**Последствия**
	Датчик потока FM3	– нарушение работы системы аэрации; – избыточная/недостаточная подача воды; – недостаточное удаление шлама из очищаемой воды; – переполнение резервуара флотации; – попадание неочищенной воды в водоемы.

На рисунке 5 представлены экземпляры онтологии SecWat-O, которые описывают датчики и актуаторы учебного стенда CE 587, используемые обозначения сенсоров и актуаторов представлены выше. На рисунке 6 представлены экземпляры классов части онтологии, относящаяся к MITRE ATT&CK, в частности показаны индивиды, имеющие класс *Тактики*, *Техники*, *Противодействия*, а также указаны связи между ними. Идентификаторы индивидов совпадают с идентификаторами объектов базы знаний MITRE ATT&CK.

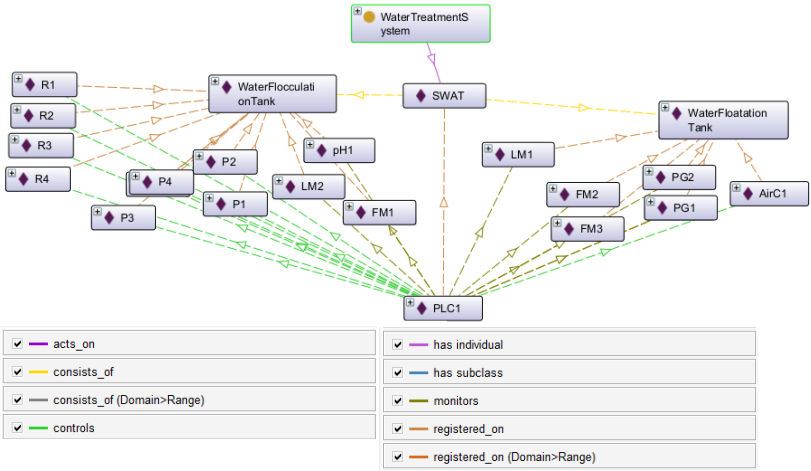


Рис. 5. Экземпляры онтологии, представляющие структурные компоненты учебного стенда CE 587

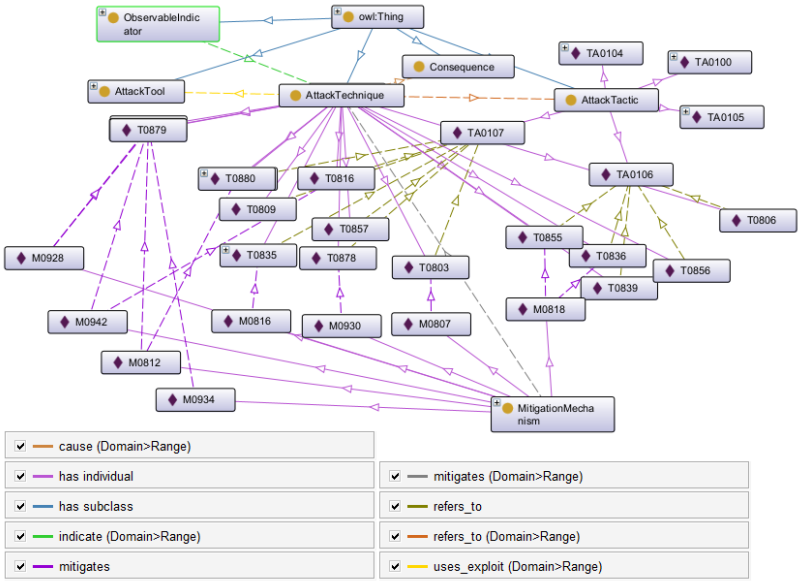


Рис. 6. Экземпляры онтологии, относящиеся к базе знаний MITRE ATT&CK

Для проверки полноты разработанной онтологии был разработан ряд запросов, построенных с учетом вопросов компетентности данной онтологии, сформулированных в разделе 4. Запросы позволяют ответить на вопросы о связях между устройствами, наличии атакующих тактик и техник, возможных для данного стенда, каковы возможные меры по их устранению и каковы возможные последствия успешной реализации атак. Например, запрос, показанный в листинге 1, возвращает список устройств, функционирование которых связано с показаниями датчика уровня воды LM1, а на рисунке 7 представлен результат запроса для учебного стенда. Нарушение целостности функционирования одного из этих устройств, например, путем подмены значений, потенциально может вызвать неисправности в функционировании связных устройств.

```
PREFIX IoT-O: <http://www.semanticweb.org/eve/ontologies/2025/8/IoT-O#>
SELECT ?actuator ?property  WHERE {
            ?actuator IoT-O:acts_on ?property.
            IoT-O:LM1 IoT-O:observes ?property}
```

Листинг 1. Поиск устройств, функционирование которых взаимосвязано через заданное свойство системы

actuator	property
P1	flotation_water_level
P2	flotation_water_level

Рис. 7. Результат SPARQL запроса, представленного в листинге 1

Запрос, представленный в листинге 2, возвращает перечень возможных атакующих техник, определенных для контроллера. На рисунке 8 представлен результат запроса для текущей версии онтологии.

```
PREFIX SecWat-O: <http://www.semanticweb.org/eve/ontologies/2025/8/SecWat-O#>
PREFIX MITRE-ATTACK: <http://www.semanticweb.org/eve/ontologies/2025/8/
     MItreAttack#>
PREFIX IoT-O: <http://www.semanticweb.org/eve/ontologies/2025/8/IoT-O#>
SELECT ?attack_tech ?description ?device ?device_type   WHERE {
        ?attack_tech rdfs:comment ?description.
    ?attack_tech SecWat-O:is_defined_for ?device.
        ?device rdf:type  ?device_type.
        ?device_type rdfs:subClassOf  IoT-O:Controller}
```

Листинг 2. Запрос, возвращающий список возможных атакующих техник для устройства типа Controller

attack_tech	description	device	device_type
T0803	"Block Command Message"	PLC1	PLController
T0806	"Brute Force I/O"	PLC1	PLController
T0856	"Spoof Reporting Message"	PLC1	PLController
T0836	"Modify Parameter"	PLC1	PLController
T0878	"Alarm Suppresion"	PLC1	PLController
T0804	"Block Reporting Message"	PLC1	PLController

Рис. 8. Результат SPARQL запроса, показанного в листинге 2

Запрос, показанный в листинге 3 возвращает список устройств, для которых определены уязвимости в CVE, инструменты, которые их эксплуатируют. Для учебного стенда примером такой уязвимости служит уязвимость CVE-2020-15782, определенная для контроллера приводов семейства SIMATIC компании Siemens AG. Результат запроса дан на рисунке 9, из него следует, что для эксплуатации уязвимости CVE-2020-15782 может быть использован червь PLC-Blaster. Запрос, представленный на листинге 4, возвращает перечень возможных последствий, определенных для заданного устройства, а именно насоса P1, который управляет подачей воды в резервуар флокуляции. Список последствий формируется на основе анализа последствий, заданных как для заданного устройства, так и для устройств, с которыми он связан

физически: резервуар флокуляции и сама система очистки воды в целом. Результат запроса представлен на рисунке 10.

```
PREFIX SecWat-O: <http://www.semanticweb.org/eve/ontologies/2025/8/SecWat-O#>
PREFIX MITRE-ATTACK: <http://www.semanticweb.org/eve/ontologies/2025/8/
     MItreAttack#>
PREFIX IoT-O: <http://www.semanticweb.org/eve/ontologies/2025/8/IoT-O#>

SELECT ?vulnerability ?tool ?device ?technique WHERE {
    ?technique SecWat-O:is_defined_for ?device.
    ?technique MITRE-ATTACK:uses_exploit ?tool.
    ?tool   MITRE-ATTACK:exploits ?vulnerability.
    ?vulnerability rdf:type MITRE-ATTACK:Vulnerability }
```

Листинг 3. Запрос, возвращающий список устройств, для которых определены уязвимости в MITRE ATT&CK

vulnerability	tool	device	technique
CVE-2020-15782	PLC-Blaster	PLC1	T0835

Рис. 9. Результат SPARQL запроса, показанного в листинге 3

```
PREFIX SecWat-O: <http://www.semanticweb.org/eve/ontologies/2025/8/SecWat-O#>
PREFIX MITRE-ATTACK: <http://www.semanticweb.org/eve/ontologies/2025/8/
     MItreAttack#>
PREFIX IoT-O: <http://www.semanticweb.org/eve/ontologies/2025/8/IoT-O#>
SELECT * WHERE {
    {   SELECT ?id ?desc WHERE {
           ?id rdfs:comment ?desc.
           ?id SecWat-O:is_associated_with ?system.
           ?system IoT-O:consists_of ?node.
           IoT-O:P1 IoT-O:registered_on ?node. }
    }   UNION
    {   SELECT ?id ?desc WHERE {
         ?id rdfs:comment ?desc.
         ?id SecWat-O:is_associated_with ?node.
         IoT-O:P1 IoT-O:registered_on ?node. }
    }
}
```

Листинг 4. Запрос, возвращает возможные последствия для заданного устройства и связанных с ним компонентов

id	
con001_SWAT	"Leakage of the waste water"
Con000_SWAT	"Malfunction of water treatment system"
Con00_Flocculation_tank	"OVerfiiling or drying of the tank reservour"
Con01_Floculation_tank	"Leackage of waste water"

Рис. 10. Результат SPARQL запроса, представленного в листинге 4

6. Заключение. В статье был предложен подход к управлению рисками информационной безопасности в системе управления водоочистными сооружениями, в основе которого лежит онтология SecWat-O. Онтология представляет собой семантическую модель, объединяющую данные о системе управления технологическим процессом и атакующими воздействиями на киберфизические системы. Многими исследователями [13] отмечается, что применение подходов, в основе которых лежат семантические модели, актуально ввиду растущей сложности таких систем и наличию множества зависимостей между ее элементами. Таким образом, онтология является эффективным инструментом анализа информационных рисков, позволяющим моделировать логические зависимости между компонентами системы, анализировать возможные угрозы и интерпретировать их последствия в контексте технологического процесса.

Предложенная в статье онтология SecWat-O объединяет в себе ключевые знания о предметной области очистки воды, представленные в виде описания технических и логических активов системы управления процессом очистки и определения связей между ними, типов киберугроз, уязвимостей и атакующих воздействий, представленных в виде тактик и техник базы знаний MITRE ATT&CK.

Одной из ключевых особенностей разработанной онтологии является учет специфики технологии очистки воды, что позволяет использовать ее для задач оценки рисков, мониторинга безопасности, в т.ч. экологической безопасности, и принятия решений в условиях кибератак.

Проверка полноты разработанной онтологии SecWat-O была выполнена путем моделирования функционирования упрощенной системы управления водоочистными сооружениями, представленного учебным стендом флотационной очистки воды CE 587. Были подробно описаны основные элементы данного стенда, датчики, исполнительные устройства. Определены атакующие сценарии, такие как подключение съемного диска с вредоносным ПО к рабочей станции, настройка системы

с нарушением процесса очистки и подключение злоумышленника к внутренней сети. Были разработаны SPARQL запросы, реализующие компетентностные вопросы онтологии. С их помощью было показано, что наиболее критичными атаками являются атаки, направленные на изменение логики функционирования контроллера, для которых в настоящее время нет эффективных контрмер.

Задачи дальнейших исследований связаны с автоматизацией синхронизации с базами MITRE ATT&CK, обновлением информации о самой системе водоочистных сооружений, а также уточнением классов онтологии, связанных с наблюдаемыми событиями и выбором контрмер. Последнее позволит строить адаптивную систему противодействия атакам в режиме, близком к реальному, с учетом последствий нарушения безопасности технологического процесса.

Литература

1. 11 recent cyber attacks on the water and wastewater sector // Wisdiam. 2024. URL: https://wisdiam.com/publications/recent-cyber-attacks-water-wastewater/ (дата обращения: 17.09.2025).
2. Almoabady T.A., et al. Protecting digital assets using an ontology based cyber situational awareness system // Frontiers in Artificial Intelligence. 2024. vol. 7. DOI: 10.3389/frai.2024.1394363.
3. Sanchez-Zas C., et al. A methodology for ontology-based interoperability of dynamic risk assessment frameworks in IoT environments // Internet of Things. 2024. vol. 27. 101267 p. DOI: 10.1016/j.iot.2024.101267.
4. Getting Started just the basics // CASE Ontology. 2025. URL: https://caseontology.org/ontology/start.html (дата обращения: 17.09.2025).
5. STIX 2.1 ontology model POC with Stardog using knowledge graphs and RDF // Smart City Cyber Security. 2024. URL: https://smartcitysecurity.net/stix-2-1-ontology-model-poc-with-stardog-using-knowledge-graphs-and-rdf/ (дата обращения: 17.09.2025).
6. Introduction to STIX // GitHub. 2024. URL: https://oasis-open.github.io/cti-documentation/stix/intro.html (дата обращения: 17.09.2025).
7. D3FEND Analytic Technique Taxonomy // MITRE. 2024. URL: https://d3fend.mitre.org/acf/ (дата обращения: 17.09.2025).
8. Akbar K.A., et al. The Design and Application of a Unified Ontology for Cyber Security / In: Muthukkumarasamy V., Sudarsan S.D., Shyamasundar R.K. (eds) // Information Systems Security. Lecture Notes in Computer Science. Springer. 2023. vol. 14424. pp. 23–41.
9. MITRE ATTCK. 2025. URL: https://attack.mitre.org/ (дата обращения: 17.09.2025).
10. D3FEND. A knowledge graph of cybersecurity countermeasures // MITRE. 2025. URL: https://d3fend.mitre.org/ (дата обращения: 17.09.2025).
11. CWE: Common Weakness Enumeration // MITRE. 2025. URL: https://cwe.mitre.org/ (дата обращения: 17.09.2025).
12. CVE: Common Vulnerabilities and Exposures // MITRE. 2025. URL: https://www.cve.org/ (дата обращения: 17.09.2025).
13. Doynikova E., et al. Ontology of Metrics for Cyber Security Assessment // Proceedings of the 14th International Conference on Availability, Reliability and Security (ARES '19). DOI: 10.1145/3339252.3341496.

14. Khaleghi M., et al. Context-Aware Ontology-based Security Measurement Model // Journal of Information Security and Applications. 2022. vol. 67. 103199 p. DOI: 10.1016/j.jisa.2022.103199.
15. Mozzaquatro B.A., et al. An Ontology-Based Cybersecurity Framework for the Internet of Things // Sensors. 2018. vol. 18. no. 9. 3053 p. DOI: 10.3390/s18093053.
16. Sanchez-Zas C., et al. Ontology-based approach to real-time risk management and cyber-situational awareness // Future Generation Computer Systems. 2023. vol. 141. pp. 462–472. DOI: 10.1016/j.future.2022.12.006.
17. Khalil K., et al. CyberROAD: A cybersecurity risk assessment ontology for automotive domain aligned with ISO/SAE21434:2021 // Journal of Information Security and Applications. 2025. vol. 90. 104015 p. DOI: 10.1016/j.jisa.2025.104015.
18. Alanen J., et al. Hybrid ontology for safety, security, and dependability risk assessments and Security Threat Analysis (STA) method for industrial control systems // Reliability Engineering System Safety. 2022. vol. 220. 108270 p. DOI: 10.1016/j.ress.2021.108270.
19. Bughio K.S., et al. Developing a Novel Ontology for Cybersecurity in Internet of Medical Things-Enabled Remote Patient Monitoring // Sensors. 2024. vol. 24. no. 9. 2804 p. DOI: 10.3390/s24092804.
20. Preuveneers D., et al. An Ontology-Based Cybersecurity Framework for AI-Enabled Systems and Applications // Future Internet. 2024. vol. 16. no. 3. 69 p. DOI: 10.3390/fi16030069.
21. Dash S., et al. From Data to Defense: How Ontology Fuels AI in Cyber Threat Detection // Proceedings of the 8th International Conference on Advances in Artificial Intelligence (ICAAI '24). 2025. pp. 121–133. (ICAAI '24). DOI: 10.1145/3704137.3704176.
22. Seydoux N., et al. IoT-O, a Core-Domain IoT Ontology to Represent Connected Devices Networks / In: Blomqvist E., Ciancarini P., Poggi F., Vitali F. (eds) // Knowledge Engineering and Knowledge Management. Lecture Notes in Computer Science. Springer. 2016. vol. 10024. pp. 561–576.
23. Novikova E., et al. Dataset Generation Methodology: Towards Application of Machine Learning in Industrial Water Treatment Security // SN Computer Science. 2024. vol. 5. 373 p. DOI: 10.1007/s42979-024-02704-9.

Саенко Игорь Борисович — 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». Область научных интересов: автоматизированные информационные системы, информационная безопасность, искусственный интеллект, машинное обучение, теория моделирования и математическая статистика, теория информации. Число научных публикаций — свыше 500. ibsaen@comsec.spb.ru, http://www.comsec.spb.ru; 14-я линия В.О., д. 39, Санкт-Петербург, 199178, Российская Федерация; тел.: +7(812)328-7181, факс: +7(812)328-4450.

Новикова Евгения Сергеевна — канд. техн. наук, старший научный сотрудник, лаборатория проблем компьютерной безопасности, Федеральное государственное бюджетное учреждение науки «Санкт-Петербургский Федеральный исследовательский центр Российской академии наук». Область научных интересов: безопасность информационных систем, обнаружение аномалий методами машинного обучения, обеспечение конфиденциальное машинное обучение. Число научных публикаций — более 150. novikova@comsec.spb.ru; 14-я линия В.О., 39, 199178, Санкт-Петербург, Российская Федерация; р.т.: +7(812)328-7181; факс: +7(812)328-4450

Мелешко Алексей Викторович — младший научный сотрудник, Международный центр цифровой криминалистики, Федеральное государственное бюджетное учреждение науки

«Санкт-Петербургский Федеральный исследовательский центр Российской академии наук». Область научных интересов: обнаружение атак методами машинного обучения, безопасность беспроводных сенсорных сетей, безопасность аддитивных производств. Число публикаций - 55. meleshko.a@iias.spb.su; 14-я линия В.О., 39, 199178, Санкт-Петербург, Российская Федерация; р.т.: +7(812)328-7181; факс: +7(812)328-4450

Садовников Владимир Евгеньевич — аспирант, лаборатория проблем компьютерной безопасности, Федеральное государственное бюджетное учреждение науки «Санкт-Петербургский Федеральный исследовательский центр Российской академии наук». Область научных интересов: обнаружение компьютерных атак, машинное обучение. Число научных публикаций — 25. bladimir1998@mail.ru; 14-я линия В.О., д. 39, Санкт-Петербург, 199178, Российская Федерация; тел.: +7(812)328-7181, факс: +7(812)328-4450.

Поддержка исследований. Работа выполнена при поддержке гранта Российского научного фонда № 23-11-20024 (https://rscf.ru/project/23-11-20024/) и Санкт-Петербургского научного фонда в СПб ФИЦ РАН.

DOI 10.15622/ia.25.3.9

I.B. Saenko , E.S. Novikova , A.V. Meleshko , V.E. Sadovnokov

AN ONTOLOGICAL APPROACH TO RISK MANAGEMENT IN WASTEWATER TREATMENT SYSTEM

Saenko I.B., Novikova E.S., Meleshko A.V., Sadovnikov V.E. **An Ontological Approach to Risk Management in Wastewater Treatment System.**

Abstract. Modern process control systems at critical infrastructure facilities, including wastewater treatment plants, are increasingly becoming the target of sophisticated cyberattacks aimed at compromising environmental safety and operational sustainability. Risk management in industrial cyber-physical systems (CPS) requires taking into account the high level of device heterogeneity and the presence of vulnerabilities in the software and hardware of industrial equipment. This necessitates developing an approach that takes into account the behavior of complex CPS, the specifics of attack patterns against CPS, and the consequences of disrupting CPS processes. This article proposes an approach to modeling cyberthreats in CPS using the example of a wastewater treatment system. This approach is based on an ontological model focused on assessing cyberthreats to CPS. The model links vulnerability and attack data with the specifics of the process, including the environmental consequences of disrupting the integrity of the water treatment process. Its application enables cyber risk analysis by identifying potential information threats and considering their consequences. The application of the developed ontology is demonstrated by investigating attacks defined for a prototype of a flotation water treatment system.

Keywords: information security, semantic threat modeling, MITRE ATT&CK threat matrix, industrial control systems, water treatment facilities.

References

1. 11 recent cyber attacks on the water and wastewater sector. Wisdiam. 2024. Available at: https://wisdiam.com/publications/recent-cyber-attacks-water-wastewater/ (accessed 17.09.2025).
2. Almoabady T.A., et al. Protecting digital assets using an ontology based cyber situational awareness system. Frontiers in Artificial Intelligence. 2024. vol. 7. DOI: 10.3389/frai.2024.1394363.
3. Sanchez-Zas C., et al. A methodology for ontology-based interoperability of dynamic risk assessment frameworks in IoT environments. Internet of Things. 2024. vol. 27. 101267 p. DOI: 10.1016/j.iot.2024.101267.
4. Getting Started just the basics. CASE Ontology. 2025. Available at: https://caseontology.org/ontology/start.html (accessed 17.09.2025).
5. STIX 2.1 ontology model POC with Stardog using knowledge graphs and RDF. Smart City Cyber Security. 2024. Available at: https://smartcitysecurity.net/stix-2-1-ontologymodel-poc-with-stardog-using-knowledge-graphs-and-rdf/ (accessed 17.09.2025).
6. Introduction to STIX. GitHub. 2024. Available at: https://oasis-open.github.io/cti-documentation/stix/intro (accessed 17.09.2025).
7. D3FEND Analytic Technique Taxonomy. MITRE. 2024. Available at: https://d3fend.mitre.org/acf/ (accessed 17.09.2025).
8. Akbar K.A., et al. The Design and Application of a Unified Ontology for Cyber Security. In: Muthukkumarasamy V., Sudarsan S.D., Shyamasundar R.K. (eds). Information Systems Security. Lecture Notes in Computer Science. Springer. 2023. vol. 14424. pp. 23–41.

9. MITRE ATTCK. 2025. Available at: https://attack.mitre.org/ (accessed 17.09.2025).
10. D3FEND. A knowledge graph of cybersecurity countermeasures. MITRE. 2025. Available at: https://d3fend.mitre.org/ (accessed 17.09.2025).
11. CWE: Common Weakness Enumeration. MITRE. 2025. Available at: https://cwe.mitre.org/ (accessed 17.09.2025).
12. CVE: Common Vulnerabilities and Exposures. MITRE. 2025. Available at: https://www.cve.org/ (accessed 17.09.2025).
13. Doynikova E., et al. Ontology of Metrics for Cyber Security Assessment. Proceedings of the 14th International Conference on Availability, Reliability and Security (ARES '19). DOI: 10.1145/3339252.3341496.
14. Khaleghi M., et al. Context-Aware Ontology-based Security Measurement Model. Journal of Information Security and Applications. 2022. vol. 67. 103199 p. DOI: 10.1016/j.jisa.2022.103199.
15. Mozzaquatro B.A., et al. An Ontology-Based Cybersecurity Framework for the Internet of Things. Sensors. 2018. vol. 18. no. 9. 3053 p. DOI: 10.3390/s18093053.
16. Sanchez-Zas C., et al. Ontology-based approach to real-time risk management and cyber-situational awareness. Future Generation Computer Systems. 2023. vol. 141. pp. 462–472. DOI: 10.1016/j.future.2022.12.006.
17. Khalil K., et al. CyberROAD: A cybersecurity risk assessment ontology for automotive domain aligned with ISO/SAE21434:2021. Journal of Information Security and Applications. 2025. vol. 90. 104015 p. DOI: 10.1016/j.jisa.2025.104015.
18. Alanen J., et al. Hybrid ontology for safety, security, and dependability risk assessments and Security Threat Analysis (STA) method for industrial control systems. Reliability Engineering System Safety. 2022. vol. 220. 108270 p. DOI: 10.1016/j.ress.2021.108270.
19. Bughio K.S., et al. Developing a Novel Ontology for Cybersecurity in Internet of Medical Things-Enabled Remote Patient Monitoring. Sensors. 2024. vol. 24. no. 9. 2804 p. DOI: 10.3390/s24092804.
20. Preuveneers D., et al. An Ontology-Based Cybersecurity Framework for AI-Enabled Systems and Applications. Future Internet. 2024. vol. 16. no. 3. 69 p. DOI: 10.3390/fi16030069.
21. Dash S., et al. From Data to Defense: How Ontology Fuels AI in Cyber Threat Detection. Proceedings of the 8th International Conference on Advances in Artificial Intelligence (ICAAI '24). 2025. pp. 121–133. (ICAAI '24). DOI: 10.1145/3704137.3704176.
22. Seydoux N., et al. IoT-O, a Core-Domain IoT Ontology to Represent Connected Devices Networks. In: Blomqvist E., Ciancarini P., Poggi F., Vitali F. (eds). Knowledge Engineering and Knowledge Management. Lecture Notes in Computer Science. Springer. 2016. vol. 10024. pp. 561–576.
23. Novikova E., et al. Dataset Generation Methodology: Towards Application of Machine Learning in Industrial Water Treatment Security. SN Computer Science. 2024. vol. 5. 373 p. DOI: 10.1007/s42979-024-02704-9.

Saenko Igor Borisovich — Ph.D., Dr. Sci., Professor, Chief scientific officer, 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, artificial intelligence, machine learning, modeling theory and mathematical statistics, information theory. The number of publications — 500. ibsaen@comsec.spb.ru; 39, 14th line V.O., 199178, Saint-Petersburg, Russia; office phone: +7(812)328-7181; fax: +7(812)328-4450.

Novikova Evgenia Sergeevna — Ph.D., Senior researcher, Laboratory of Computer Security Problems, St. Petersburg Federal Research Center of the Russian Academy of Sciences (SPC RAS). Research interests: information system security, anomaly detection using machine

learning methods, providing confidential machine learning. The number of publications — 150. novikova@comsec.spb.ru; 39, 14th line V.O., 199178, Saint-Petersburg, Russia; office phone: +7(812)328-7181; fax: +7(812)328-4450.

Meleshko Aleksei Viktorovich — Junior research assistant, International Center for Digital Forensics, St. Petersburg Federal Research Center of the Russian Academy of Sciences (SPC RAS). Research interests: attack detection using machine learning, wireless sensor network security, additive manufacturing security. The number of publications — 55. meleshko.a@iias.spb.su; 39, , 199178, Saint-Petersburg, Russia; office phone: +7(812)328-7181; fax: +7(812)328-4450.

Sadovnokov Vladimir Evgenevich — Graduate student, Laboratory of Computer Security Problems, St. Petersburg Federal Research Center of the Russian Academy of Sciences (SPC RAS). Research interests: detection of computer attacks, machine learning. The number of publications — 25. bladimir1998@mail.ru; 39, 14th line V.O., 199178, Saint-Petersburg, Russia; office phone: +7(812)328-7181; fax: +7(812)328-4450.

Acknowledgements. The research is supported by the grant of Russian Science Foundation #23-11-20024, https://rscf.ru/project/23-11-20024/, and St. Petersburg Science Foundation.

УДК 004.89 **DOI** 10.15622/ia.25.3.10

Н.Л.Харина, М.Л.Долженкова, З.Б.Иминова

ЗАЩИЩЕННАЯ СИСТЕМА БИОМЕТРИЧЕСКОЙ АУТЕНТИФИКАЦИИ ДЛЯ ВСТРАИВАЕМЫХ СИСТЕМ С ОГРАНИЧЕННЫМИ ВЫЧИСЛИТЕЛЬНЫМИ РЕСУРСАМИ

Харина Н.Л., Долженкова М.Л., Иминова З.Б. **Защищенная система биометрической аутентификации для встраиваемых систем с ограниченными вычислительными ресурсами.**

Аннотация. В настоящее время технологии биометрической аутентификации такие как распознавание лица, отпечатков пальцев, радужной оболочки глаза и вен ладони, активно применяются в банковских приложениях, медицинских системах, комплексах контроля доступа и государственных информационных ресурсах. Использование таких технологий должно сопровождаться защитой биометрической информации при обработке и хранении особенно в условиях растущего количества целевых атак на персональные данные. Одним из перспективных подходов к обеспечению безопасности биометрической информации является использование гомоморфного шифрования. В работе рассматривается система автономной защищенной биометрической аутентификации по рисунку вен ладони пользователя в ближнем инфракрасном спектре для встроенных систем, реализуемых с использованием технологии обеспечения конфиденциальности для защиты биометрических данных. Проведенные исследования показали практическую реализуемость системы биометрической аутентификации с применением легковесной сиамской нейронной сети для выделения вектора признаков и гомоморфного шифрования, обеспечивающего конфиденциальность во встраиваемых системах на процессорах с ограниченными вычислительными ресурсами на основе одноплатных компьютеров, таких как Raspberry Pi. Подтверждена устойчивость к искажениям, включая повороты, шум, смещения и варьирование параметров шифрования. Показатели эффективности разработанной системы оценивались на сборном датасете, состоящем из изображений датасетов CASIA-MS-PalmprintV1 и Tongji Contactless Palmvein Dataset и составляют: точность классификации – 98,2%, EER – 0,73 %, AUC – 0,96, время – 1,2 секунды, что соответствует уровню требований к подобным системам. Анализ результатов показывает, что разработанная система уступает по точности клиент-серверным решениям, построенным на тяжеловесных сверточных нейронных сетях из-за ограниченности в вычислительных ресурсах и как следствие необходимости применения легковесной нейронной сети, но при этом обеспечена защищенная обработка биометрических данных с сохранением допустимого времени отклика. Сокращение времени обработки при реализации на одноплатных компьютерах может быть реализовано как использованием специализированных вычислительных ускорителей, так и применением других библиотек, поддерживающих гомоморфные вычисления, менее требовательных к вычислительным ресурсам.

Ключевые слова: биометрическая аутентификация, обработка изображений, сверточная нейронная сеть, сиамская нейронная сеть, конфиденциальная обработка биометрических признаков, гомоморфное шифрование.

1. Введение. Современные технологии биометрической аутентификации становятся важнейшим компонентом информационной безопасности, обеспечивая надежный и удобный способ подтверждения личности пользователя при доступе к цифровым системам.

Биометрические методы, такие как распознавание лица, отпечатков пальцев, радужной оболочки глаза и вен ладони, активно применяются в банковских приложениях, медицинских системах, комплексах контроля доступа и государственных информационных ресурсах [1 – 3].

В последнее время широкое распространение получил метод аутентификации по рисунку вен ладони. Данный вид биометрии является неизменным в течение жизни пользователя, не требует дорогостоящего оборудования для реализации, демонстрирует высокую точность, устойчивость к подделке и низкий уровень ложных срабатываний. Обзор разработок с использованием данного вида биометрии и технологий искусственного интеллекта для выделения признаков и сравнения с базовыми шаблонами демонстрирует актуальность данной технологии, высокие показатели точности распознавания и идентификации пользователя. Среди таких разработок можно выделить некоторые работы. Авторы [4] предлагают модель аутентификации по венам ладони с использованием сверточных нейронных сетей (CNN) с применением байесовской оптимизации. В работе [5] авторы используют адаптивный фильтр Габора и триплет CNN. Авторы [6] предлагают подход MSMDGAN® CNN, который состоит из многомасштабной и многонаправленной генеративно-состязательной сети (MSMDGAN) для аугментации данных и CNN для идентификации вен на ладони. В работе [7] приведен пример применения CNN на основе ZFNet архитектуры для распознавания рисунка вен ладони. Авторы [8] разработали Focal Contrastive Palm Vein Network для построения биометрической системы аутентификации. В работе [9] предложен многомасштабный преобразователь вен (MSVT) для обучения устойчивым и многомасштабным признакам, который состоит из сверточного блока, захватывающего локальную информацию, и блока самовнимания, извлекающего масштабные зависимости между изображениями с разными масштабами, для учета взаимосвязи между метками использована графовая сверточная сеть (GCNLE). Авторы [10] предлагают подход VeinGuard, состоящий из локальной трансформерной генеративно-состязательной сети (LTGAN), которая обучается распределению изображений вен и генерирует высококачественные изображения вен ладони, и очистителя, состоящего из обучаемой остаточной сети и предварительно обученного генератора на основе LTGAN для удаления частей изображений, не относящихся к рисунку вен. Авторы [11] провели сравнительный анализ существующих моделей CNN, таких как LeNet, AlexNet, ResNet, ZFNet, выделяя по производительности ZFNet.

Большинство из рассмотренных вариантов используют тяжеловесные модели, что оправдано для повышения точности распознавания, такие алгоритмы используются для построения распределенных и клиент-серверных архитектур систем биометрической аутентификации (СБА). Однако наряду с распределенными и клиент-серверными архитектурами СБА находят свое применение автономные или встроенные системы [12], требующие применения легковесных нейронных сетей из-за ограниченных вычислительных ресурсов, но таких разработок не так много.

Наряду с очевидными преимуществами, СБА предъявляют особые требования к защите пользовательских данных. Биометрические признаки являются уникальными и неизменяемыми, поэтому утечка может привести к необратимым последствиям. Это делает конфиденциальную обработку и хранение биометрических данных критически важным аспектом архитектуры подобных систем, особенно в условиях растущего количества целевых атак на персональные данные [13 – 15]. Традиционные алгоритмы шифрования не применимы к биометрическим данным, так как компрометация секретного ключа приводит к утечке биометрических данных. Более того, небольшие изменения в исходных биометрических данных, такие как неизбежные расхождения между несколькими измерениями биометрических данных, относящихся к одной и той же характеристике одного и того же человека, приводят к кардинальным изменениям в зашифрованных данных. Поэтому биометрические данные невозможно обрабатывать и сравнивать в зашифрованном виде и необходимо их расшифровать перед сравнением, что создает риск их использования потенциальными злоумышленниками. Следовательно, для защиты биометрических данных требуются другие технологии повышения конфиденциальности биометрического распознавания. К ним относятся такие технологии как:

– отменяемые биометрические данные – преднамеренное повторяющееся искажение исходных биометрических данных, которые позволяют сравнивать биометрические данные в преобразованном виде;

– биометрические криптосистемы – предназначены для безопасной привязки цифрового ключа к биометрическим данным или генерации цифрового ключа из биометрических данных;

– гомоморфное шифрование – технология, позволяющая выполнять операции над зашифрованными данными без промежуточного расшифровывания [16];

– применение нечетких экстракторов для формирования ключа для стандартных криптографических методов [17, 18].

Применение нечетких экстракторов заключается в использовании кодов, способных обнаруживать и корректировать ошибки.

При выборе технологии должны соблюдаться основные требования к конфиденциальности биометрических данных:

– необратимость: восстановление исходного биометрического шаблона из сохраненных справочных данных, т.е. защищенного шаблона, должно быть трудным в вычислительном отношении, в то время как создание защищенного биометрического шаблона должно быть простым;

– несвязность: разные версии защищенных биометрических шаблонов могут быть созданы на основе одних и тех же биометрических данных (возможность обновления), в то время как защищенные шаблоны не должны допускать перекрестного сопоставления (разнообразия) [19].

Одним из наиболее перспективных подходов к обеспечению безопасности биометрической информации является использование гомоморфного шифрования – криптографической технологии, позволяющей выполнять операции над зашифрованными данными без предварительной расшифровки. Такой подход предоставляет возможность защищенной обработки данных на всех этапах жизненного цикла: от хранения и передачи до сравнения признаков. На фоне активного внедрения методов машинного обучения интеграция гомоморфного шифрования в биометрические системы аутентификации получила широкое распространение в научных и прикладных разработках, например, в [20] представлена СБА по радужке глаз с применением гомоморфного шифрования, в [21, 22] рассмотрены варианты защищенной биометрической аутентификации по лицу. В работе [23] авторы предложили использовать защищенную гомоморфным шифрованием аутентификацию по лицу для IoT систем, в [24] авторы оценивают устойчивость к атакам СБА на основе отпечатка пальца, рассматривают варианты защиты биометрических данных. В [25] авторы провели обзор моделей глубокого обучения, применяемых для аутентификации пользователей по рисунку вен пальца и ладони за последние 5 лет, демонстрирующий высокую точность аутентификации пользователей при надежной защите биометрических данных.

2. Постановка задачи. Требуется разработать автономную СБА для встроенных систем, обеспечивающую конфиденциальность биометрических данных. В качестве биометрического признака использовать рисунок вен ладони пользователя.

2.1. Математическая постановка задачи.

Входные данные. На вход системы поступает изображение ладони пользователя, полученное с камеры ближнего инфракрасного диапазона (NIR):

$$I_{raw}(x, y) \in R_{H \times W},$$

где: H, W – высота и ширина исходного изображения, $I_{raw}(x,y)$ – интенсивность пикселя в оттенках серого.

Выходные данные. На выходе система принимает бинарное решение:

$$d \in \{0,1\},$$

где: $d = 1$ – аутентификация успешна (пользователь подтверждён), $d = 0$ – аутентификация не успешна (доступ запрещён).

Этапы обработки данных. Выделение области интереса (ROI) – прямоугольная область ладони, содержащая венозный рисунок:

$$ROI = CropAndRotate(I_{raw}, \theta, P_1, P_2),$$

где: P_1, $P2$ – опорные точки (впадины между пальцами), Θ – угол поворота для нормализации.

Бинаризация ROI с помощью пороговой обработки:

$$ROI\_BINARY(x, y) = \begin{cases} 1, & \textit{если } ROI(x, y) \geq T, \\ 0, & \textit{если } ROI(x, y) < T. \end{cases}$$

Скелетизация бинарного изображения *ROI_BINARY*:

$$ROI\_Skel = Skeletonize(ROI\_BINARY),$$

где *ROI_Skel* – бинарное изображение венозного рисунка толщиной в один пиксель.

Извлечение признаков с помощью сиамской нейросети. Скелетизированное изображение *ROI*_Skel подаётся на вход свёрточной нейросети *f*:

$$v = f(ROI\_Skel) \in R_{128},$$

где v – вектор признаков длиной 128, компактно описывающий венозный рисунок.

Квантование и гомоморфное шифрование. Вектор признаков v нормализуется к диапазону $[-1,1]$ и квантуется в целые числа с фиксированной разрядностью n:

$$v_q = Quantize(v) \in Z^n.$$

Затем каждый элемент шифруется с использованием открытого ключа *public_key* в рамках схемы полного гомоморфного шифрования (FHE), поддерживающей сложение и умножение:

$$c = Enc_{public\_key}(v_q) \in C^n,$$

где C – пространство шифротекстов.

Сравнение с эталоном в зашифрованном виде. Для аутентификации выполняется конкатенация зашифрованного входного вектора c и эталонного зашифрованного вектора c_{ref}, соответствующего идентификатору пользователя id:

$$c_{concat} = c \| c_{ref},$$

где || – операция конкатенации, выполняемая над шифротекстами.

Полученный вектор подаётся на классификатор g, реализованный как часть свёрточной нейросети, адаптированной для работы с зашифрованными данными:

$$w = g(c_{concat}) \in [0,1],$$

где w – вероятность принадлежности одной и той же ладони.

Принятие решения. Итоговое решение принимается путём сравнения с порогом τ:

$$d = \begin{cases} 1, & \text{если } w \geq \tau, \\ 0, & \text{если } w < \tau. \end{cases}$$

Порог τ подбирается эмпирически на этапе валидации как значение, обеспечивающее баланс между вероятностями ложного допуска и ложного отказа.

2.2. Аппаратно-программная платформа:

Камера ближнего инфракрасного диапазона (NIR) для захвата изображений вен ладони.

СБА должна функционировать на устройстве с ограниченными вычислительными ресурсами – одноплатный компьютер (Raspberry Pi 4/5 или аналог).

Операционная система: Linux (Raspberry Pi OS).

2.3. Функциональные требования:

1. Захват и предобработка изображения ладони пользователя в NIR-спектре.
2. Извлечение биометрического шаблона.
3. Извлечение вектора признаков.
4. Выполнение криптографического преобразования над вектором признаков для обеспечения защищенной обработки и хранения эталонных шаблонов в базе данных.
5. Сравнение входного биометрического шаблона с эталонным шаблоном, хранящимся в системе.
6. Принятие бинарного решения на основе порога схожести (Аутентификация успешна/неуспешна).

2.4. Требования к производительности и точности:

1. Время выполнения полного цикла аутентификации ≤ 2 секунды для реализации на ПК и ≤ 4 секунды при реализации на Raspberry Pi 4/5 или аналог.
2. Точность классификации (аутентификации) – Accuracy $\geq 98\%$ (на тестовой выборке).
3. Вероятность ошибок первого рода (FRR) и второго рода (FAR) должны быть минимизированы при соблюдении условия точности.

2.5. Требования к безопасности и конфиденциальности:

1. Биометрические шаблоны (исходные и эталонные) должны храниться и обрабатываться только в защищенном (необратимом) виде.
2. Метод криптографического преобразования – полное гомоморфное шифрование.
3. Криптографические ключи не должны покидать устройство.

3. Структура системы биометрической аутентификации.

Общая структура системы биометрической аутентификации представлена на рисунке 1 и состоит из следующих модулей:

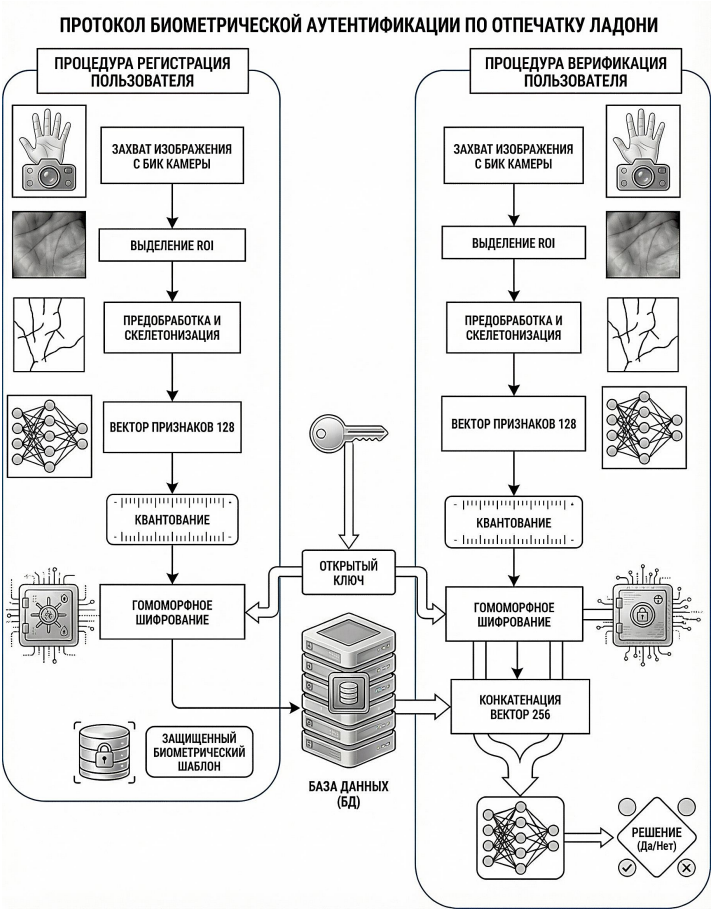


Рис. 1. Структура системы биометрической аутентификации

– модуль захвата и предобработки изображений ладони в инфракрасном спектре;

– модуль выделения области интереса (ROI) и предварительной обработки изображения (адаптивное выравнивание гистограммы с ограниченным контрастом, бинаризация, скелетизация);

– модуль извлечения признаков с использованием сверточной нейросети;

– модуль квантования и гомоморфного шифрования признаков с использованием открытого ключа;

– модуль сравнения с эталоном и принятия решения сверточной моделью.

СБА может работать в двух режимах: внесение биометрического шаблона пользователя в базу данных и режим верификации пользователя.

В режиме внесения данных изображение с БИК камеры поступает в блок выделения области интереса (ROI), проходит все этапы предобработки, результатом является скелет вен ладони. Далее с использованием сверточной нейросети осуществляется выделение вектора признаков с последующим квантованием и зашифрованием. Защищенный биометрический шаблон подлежит записи в базу данных.

В режиме аутентификации пользователь предъявляет свой идентификатор и прикладывает ладонь к сканеру. Изображение ладони проходит описанные выше процедуры и для вынесения решения об аутентификации происходит объединение защищенного биометрического шаблона пользователя с его эталонным из базы данных.

4. Модуль выделения области интереса и предварительной обработки изображения. Выделение области интереса является базовым этапом в структуре биометрической СБА. Цель данного этапа – локализация центральной части ладони, содержащей устойчивую и информативную васкулярную структуру, а также нормализация положения руки по масштабу и ориентации. В рамках проектируемой архитектуры применяется геометрический алгоритм, включающий бинаризацию, морфологическую обработку, определение опорных точек и коррекцию наклона изображения [26].

На вход подается полутоновое изображение ладони с NIR камеры (Рис. 2(a)). Для выделения объекта интереса применяется метод автоматической бинаризации по критерию Otsu. Алгоритм Otsu определяет пороговое значение яркости, при котором достигается максимальное разделение между классами «объект» (ладонь) и «фон» (Рис. 2(b)). Применение данной методики позволяет адаптировать процесс сегментации к различным условиям освещенности и яркости.

К полученному бинарному изображению применяется морфологическая эрозия с ядром [0,1,0;1,1,1;0,1,0], за которой следует операция поэлементного вычитания из исходной бинарной маски. Результатом является тонкий замкнутый контур ладони. Далее выполняется фильтрация контуров по длине: строятся контуры всех объектов и удаляются все объекты, кроме объекта с максимальной длиной. Это позволяет устранить шум и оставить только основной контур руки (Рис. 2(c)).

После получения контура ладони рассчитывается координата центра запястья, определяемая как середина между двумя крайними точками нижней границы контура. Далее вычисляются так называемые особые точки ладони – впадины между пальцами, в частности между указательным и средним, а также между безымянным и мизинцем. Для поиска проводится полный обход внешнего контура, в ходе которого для каждой точки измеряется расстояние до центра запястья. По сформированному профилю расстояний осуществляется сглаживание и выявление локальных минимумов, соответствующих впадинам между пальцами. Эти точки служат ориентирами для выравнивания изображения.

На основе координат найденных особых точек определяется угол наклона ладони относительно горизонтальной оси (Рис. 2(d)). Изображение поворачивается таким образом, чтобы линия, соединяющая особые точки, стала горизонтальной. Это позволяет устранить наклон руки при захвате и привести все изображения к единой ориентации (Рис. 2(f)).

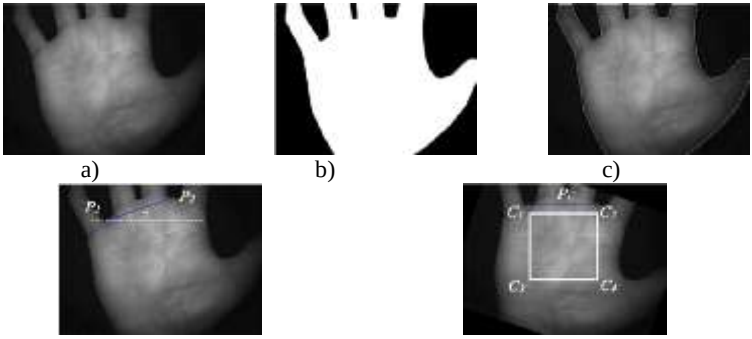


Рис. 2. Процесс выделения ROI: a – исходное изображение; b – бинаризация; c – определение контура; d – определение угла наклона; f – выделение прямоугольной области

После коррекции положения определяется прямоугольная область интереса. Ее верхняя граница проходит через линию, соединяющую найденные впадины между пальцами. Левые и правые границы определяются как вертикальные линии, проходящие через соответствующие особые точки. Нижняя граница устанавливается на фиксированном расстоянии вниз от линии впадин, определенном эмпирически на основе анализа анатомических особенностей.

В результате формируется прямоугольная ROI, содержащая центральную часть ладони и исключающая пальцы и запястье (Рис. 3(a)).

Для усиления видимости сосудистых структур внутри ROI осуществляется фильтрация последовательным применением следующих методов: контрастная ограниченная адаптивная коррекция гистограммы (CLAHE) с параметрами clipLimit=2.0, tileGridSize=(8,8), позволяющая увеличить контраст изображения без значительного усиления шума; медианное размытие (Median Blur) с ядром размытия (3,3), что позволяет уменьшить уровень шумов; размытие по Гауссу (Gaussian Blur), с ядром размытия (3,3), что позволяет уменьшить уровень шумов и снизить детализацию; повторное применение CLAHE с параметрами clipLimit=2.0, tileGridSize=(8,8), для того чтобы увеличить контраст еще сильнее.

Полученное изображение ROI (Рис. 3(a)) дополнительно преобразуется в бинарное представление с целью отделения венозного рисунка от фона. После бинаризации применяется операция скелетизации с помощью алгоритма Zhang-Suen [27], приводящая сосудистые линии к однопиксельной толщине (Рис. 3(b)). Это позволяет сохранить топологическую структуру рисунка и устранить избыточную информацию, не несущую полезных признаков.



a) b)

Рис. 3. Процесс скелетизации: a – ROI; b – результат скелетизации

Скелетизированное бинарное изображение используется в качестве входа для сверточной нейросети на следующем этапе обработки.

5. Модуль извлечения признаков с использованием сверточной нейросети. Извлечение признаков из изображения ладони осуществляется с использованием сверточной нейросети, построенной по архитектуре сиамской модели [28]. Сиамская нейросеть состоит из двух каналов с идентичными весами нейросетей в каждом и применяется обычно для задач измерения сходства между парами входных данных, кроме того, эта модель менее требовательна к вычислительным ресурсам по сравнению с моделями сверточных нейронных сетей, используемых в задачах классификации при построении СБА [4 – 11].

Основная задача нейросетевого блока – преобразование скелетизированного изображения области интереса (ROI) в компактное векторное представление, сохраняющее индивидуальные особенности венозного рисунка. Полученный вектор признаков используется в дальнейших модулях системы для сравнения с эталонными шаблонами в зашифрованной форме.

Модель реализована в виде базовой сверточной нейросети, которая применяется в обеих ветвях сиамской архитектуры с разделяемыми весами, структура представлена на рисунке 4. Структура каждой ветви включает:

1. Четыре сверточных блока:

– слой с 32 фильтрами 3×3, «ReLU», «BatchNormalization», «MaxPooling»;

– слой с 64 фильтрами 3×3, «ReLU», «BatchNormalization», «MaxPooling»;

– слой с 128 фильтрами 3×3, «ReLU», «BatchNormalization», «MaxPooling»;

– слой с 256 фильтрами 3×3, «ReLU», «BatchNormalization»;

2. Глобальный усредняющий пуллинг («Global Average Pooling»);
3. Полносвязный слой из 256 нейронов с «ReLU» и регуляризацией;
4. «Dropout» (p=0.5) для предотвращения переобучения;
5. Полносвязный слой на 128 нейронов с «ReLU».

Выходом каждой ветви является вектор признаков длины 128, компактно описывающий сосудистый рисунок ладони. Два вектора признаков объединяются путем конкатенации в единый вектор размерности 256. Далее этот вектор поступает в классификационную часть модели:

1. Полносвязный слой из 256 нейронов («ReLU»), «Dropout»;
2. Полносвязный слой из 128 нейронов («ReLU»);
3. Выходной слой из одного нейрона с сигмоидной активацией.

Классификатор предсказывает вероятность того, что пара входных изображений принадлежит одному пользователю.

Модель обучалась с использованием оптимизатора Adam. Данный выбор обусловлен его эффективностью в задачах метрического обучения и способностью адаптировать скорость обучения для разреженных признаков скелетизированных изображений. Начальная скорость обучения (learning rate) выбиралась эмпирическим путем и составляла порядка 10^{-3}. В процессе обучения применялось экспоненциальное снижение скорости для более точной

настройки весов на финальных этапах. Веса инициализировались случайным образом (обучение с нуля), без использования предобученных на других задачах моделей, что позволяет сети максимально адаптироваться к специфике биометрических шаблонов.

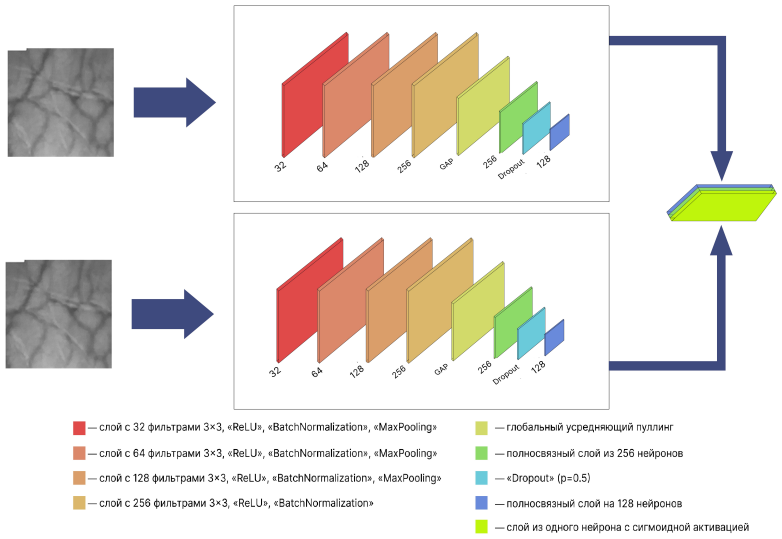


Рис.4 Архитектура нейросети

6. Модуль квантования и гомоморфного шифрования признаков с использованием открытого ключа. Основная задача данного этапа – обеспечение конфиденциальности биометрических данных посредством гомоморфного шифрования, реализуемого в асимметричной криптографической модели.

На первом этапе выполняется квантование элементов признакового вектора до целочисленного диапазона, совместимого с используемой гомоморфной схемой. Квантование осуществляется по линейному правилу с предварительной нормализацией значений к заданному диапазону. Это позволяет минимизировать искажения и сохранить структуру признаков в пределах допустимой точности. Глубина квантования может быть 8, 12 или 16 бит.

Для задач машинного обучения наиболее применимы схемы гомоморфного шифрования, поддерживающие операции над целыми числами. Это связано с тем, что признаки моделей могут быть приведены к целочисленному виду с помощью квантования, после

чего допускается выполнение линейной и сверточной арифметики в зашифрованном виде. В связи с этим, особый интерес для применения в защищенных биометрических системах представляет подход, реализованный в «Concrete ML», сочетающий возможность обработки тензоров, совместимость с популярными моделями и поддержание баланса между производительностью и защищенностью данных. Применение подобных библиотек позволяет выстраивать архитектуры, в которых конфиденциальность биометрических признаков обеспечивается на всем протяжении жизненного цикла данных – от извлечения до сравнения.

Далее применяется асимметричное шифрование квантованного вектора, основанное на схеме FHE (полное гомоморфное шифрование), которая является реализацией TFHE с поддержкой операций над тензорами, с использованием открытого ключа, сформированного в рамках асимметричной схемы гомоморфного шифрования.

В асимметричном шифровании (включая FHE):

Секретный ключ – используется для расшифровки. Хранится только у владельца, открытый ключ – используется для шифрования данных.

В FHE открытый ключ устроен так, что позволяет не только шифровать, но и производить над шифртекстом вычисления (сложение, умножение) без необходимости расшифровки. После шифрования каждый элемент признакового вектора преобразуется в соответствующий шифртекст.

7. Блок сравнения с эталоном и принятия решения сверточной моделью. После получения зашифрованного признакового вектора производится процесс сравнения с эталонным шаблоном, а также принятие решения о соответствии. В данной архитектуре используется заранее обученная сверточная нейросеть с сиамской структурой, адаптированная для обработки признаков в гомоморфно зашифрованной форме. Хранение эталонных шаблонов осуществляется в зашифрованном виде с использованием той же схемы гомоморфного шифрования. Все операции сравнения проводятся исключительно в зашифрованной области, без раскрытия признаков как на этапе загрузки, так и в процессе вычислений.

На первом этапе извлекается соответствующий эталон, привязанный к идентификатору пользователя. Вместо вычитания, два признаковых вектора – полученный в текущей сессии и эталонный – объединяются с помощью операции конкатенации. В результате

формируется вектор длины 256, содержащий полную информацию о паре сравниваемых признаков.

Конкатенированный вектор подается на классификационный блок сверточной нейросети, реализующий оценку вероятности соответствия. Сиамская архитектура модели позволяет обрабатывать пары признаков и выявлять скрытые взаимосвязи между ними. В рамках реализации модель была предварительно обучена на обширной выборке положительных и отрицательных пар изображений ладоней, благодаря чему достигнуты низкие значения ошибок распознавания.

Система использует TFHE-подход с Table Lookup для реализации полного цикла нейросетевых вычислений в зашифрованном виде, включая конкатенацию (как тривиальную операцию), линейные слои (через умножение/сложение) и нелинейные активации (через табличные функции или полиномиальные аппроксимации).

Все операции выполняются в зашифрованной форме, сверточная нейросеть функционирует над гомоморфно зашифрованными данными, не раскрывая признаки ни на одном этапе обработки. Результатом работы модели является значение вероятности совпадения. При превышении порога считается, что аутентификация успешно подтверждена.

8. Описание датасета. Для проведения эксперимента был собран датасет из изображений датасетов CASIA-MS-PalmprintV1 [29] и Tongji Contactless Palmvein Dataset (изображения ладоней 300 различных персон).

Датасет CASIA – это эталонная база данных для тестирования многоспектральных биометрических систем. Содержит изображения ладоней обеих рук 100 различных персон в 6 спектральных диапазонах (видимый спектр, 460, 530, 700, 850 и 940 нм). Сбор данных датасета проводился в две сессии, временной интервал между которыми составлял более одного месяца, что позволяет тестировать алгоритмы на устойчивость к временным изменениям. В рамках каждой сессии делалось по 3 образца, между которыми допускались небольшие изменения в положении руки для увеличения внутриклассового разнообразия. Таким образом, датасет содержит 7200 полутоновых изображений размером 576 ×768 пикселей в формате JPEG. При сборке образцов была использована специально разработанная бесконтактная многоспектральная камера (без фиксирующих штырей), которая позволяет варьировать положение руки для имитации

реальных условий использования. Кроме того, датасет имеет четкую структуру именования файлов.

Tongji Contactless Palmvein Dataset – это крупномасштабная публичная база данных изображений ладонных вен. Содержит изображения ладоней обеих рук 300 персон (192 мужчины и 108 женщин), полученные с помощью ближнего инфракрасного (NIR) излучения. Сбор данных датасета проводился в две сессии, временной интервал между которыми 61 день, что позволяет тестировать алгоритмы на устойчивость к временным изменениям (возрастным или физиологическим). В рамках каждой сессии делалось по 10 снимков каждой ладони, между которыми допускались небольшие изменения в положении руки для увеличения внутриклассового разнообразия. Таким образом, датасет содержит 12000 полутоновых изображений в формате JPEG размером 600 ×800 пикселей.

Целью эксперимента является оценка эффективности алгоритма аутентификации, поэтому датасет собран таким образом, чтобы исключить ошибки выделения ROI и явные дефекты изображений. Объединенный датасет изображений включает 326 пользователей, по 6 изображений каждой ладони на каждого пользователя и состоит из 1956×2 = 3912 изображений. Для формирования тестовой выборки случайным образом выделены 50 пользователей (600 изображений), оставшийся датасет (3312 изображений) был разделен для обучения и валидации в соотношении 80% и 20% соответственно.

9. Проведение эксперимента. На один из каналов подается ROI, а на второй – вектор признаков пользователя, зарегистрированного в базе данных. Такой способ сравнения позволяет, в процессе функционирования системы, сократить время отклика нейросети почти в два раза. Из каждого изображения формировались пары:

– положительные – изображения одной и той же ладони;

– отрицательные – изображения ладоней разных пользователей.

Таким образом, обучающая выборка включала равное количество положительных и отрицательных пар, что обеспечило сбалансированность классов.

Обучение проводилось с использованием функции потерь бинарной кроссэнтропии. В качестве оптимизатора использовался алгоритм Adam с параметрами по умолчанию: beta_1 = 0.9 (коэффициент для первого момента), beta_2 = 0.999 (коэффициент для второго момента), weight_decay (L2-регуляризация) – kernel_regularizer=l2(1e-4). Версия фреймворка TensorFlow 2.18.

Используется функция train_test_split из scikit-learn с параметром test_size=0.2. random_state=42. Обучение осуществлялось на GPU с автоматическим сохранением лучших весов по валидационной выборке. Размер батча составлял 32, обучение велось в течение 35 эпох.

Для тестирования алгоритма использован указанный выше датасет. Эксперименты проводились по следующему сценарию. На вход подавалось по очереди одно из 12 изображений каждого пользователя, формировались по 10 положительных и отрицательных пар, что позволило имитировать 500 «подходов» пользователей к БИК-камере без учета аугментации.

Для количественной оценки эффективности разработанной СБА выбраны ключевые показатели:

– общая точность (accuracy)

$$accuracy = \frac{TP+TN}{TP+TN+FP+FN};$$

где TN – количество правильных отказов пользователям, которых нет в базе, FP – количество ошибочных предоставлений доступа пользователям, которых нет в базе, TP – количество правильных предоставлений доступа пользователям, которые есть в базе, FN – количество ошибочных отказов в доступе пользователям, которые есть в базе:

– уровень ложных допусков (FAR)

– $FAR = \frac{FP}{FP+TN}$;

– уровень ложных отказов (FRR)

– $FRR = \frac{FN}{TP+FN}$;

– EER – показатель, определяющий баланс FAR и FRR;

– F1-мера, характеризующая баланс между ложноположительными и ложноотрицательными ошибками классификатора:

$$F_1 = 2 \times \frac{Precision \times Recall}{Precision+Recall},$$

где:

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

– площадь под ROC-кривой (AUC).

Модель оценивалась при различных порогах принятия решения, для чего построен график, отображающий FAR, FRR и EER от порогового значения (Рис. 5).

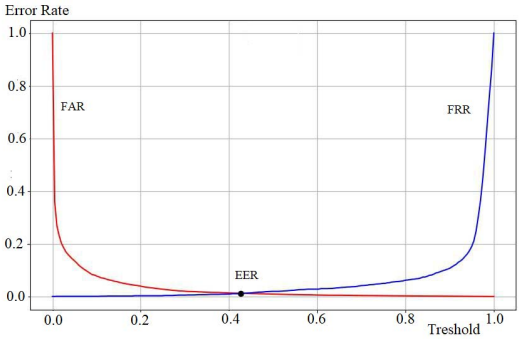


Рис. 5. Зависимость FAR и FRR от порогового значения (Treshold)

Значения показателей оценки эффективности биометрической системы аутентификации были получены при пороге 0,52, который был подобран эмпирически на валидационной выборке, и представлены в таблице 1.

Таблица 1. Показатели эффективности системы аутентификации

accuracy	FAR	FRR	F1	EER	AUC
98,2 %	0,67 %	1,38 %	98%	0,73 %	0,96

Для более подробной оценки качества классификации дополнительно была построена ROC-кривая, отражающая зависимость между долей истинно положительных срабатываний («True Positive Rate») и ложноположительных срабатываний («False Positive Rate»). Полученная кривая продемонстрировала стремительный рост TPR при низких значениях FPR, что свидетельствует о высокой дискриминационной способности модели. Значение площади под ROC-кривой (AUC) составило 0,96, что подтверждает высокую точность работы модели на всей области значений порога. ROC-кривая представлена на рисунке 6.

Время обработки биометрических данных с момента снятия изображения с камеры до шифрования данных составляет 0,06 с. Время, затраченное на шифрование, квантование и классификацию колеблется в зависимости от глубины квантования: для 16 бит – 2,0-2,2 с., при 8-

битном квантовании – 0,9-1,2 с. При этом снижение глубины квантования до 8 бит не привело к значимым потерям точности. Исследование проводилось на стандартной рабочей станции с процессором Intel Core i5 и 16 ГБ оперативной памяти без использования графических ускорителей.

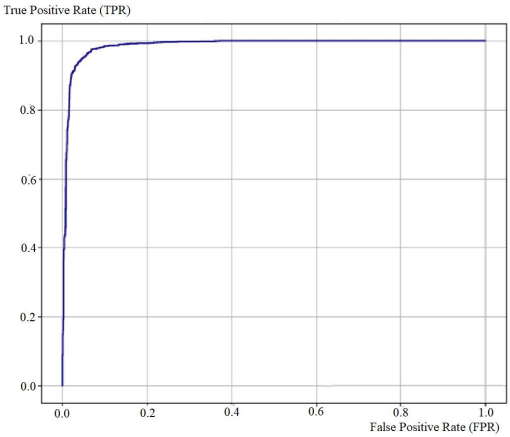


Рис. 6. ROC-кривая

Для оценки устойчивости СБА были проведены дополнительные эксперименты, направленные на изучение ее поведения при наличии искажений во входных данных, а также при варьировании параметров шифрования. Исследование проводилось на той же выборке, что и в основном эксперименте, с применением искусственно модифицированных изображений и измененных конфигураций квантования.

В качестве факторов искажения входных изображений рассматривались:

1. Смещение области ROI в пределах ±5 пикселей по осям x и y;
2. Поворот изображения на ±5 и ±10 градусов;
3. Добавление гауссовского шума с дисперсией 0.01;
4. Изменение контрастности (CLAHE с различными параметрами);
5. Небольшие масштабные искажения ROI (±10%).

Внесение в базу данных искаженных изображений позволило увеличить число «подходов» до 5000.

Показатели точности при введении искажений снижались незначительно. Наибольшее влияние оказывало смещение ROI:

при смещении на 5 пикселей accuracy снижалась на 1,3 %, а FRR возрастала до 2,4 %. При других видах искажений снижение точности не превышало 1 %. Это свидетельствует об устойчивости модели к небольшим ошибкам локализации и колебаниям условий съемки.

В таблице 2 приведены параметры точности (accuracy) и средняя ошибка аутентификации (EER) предложенной СБА с аналогичными системами, представленными в обзоре.

Таблица 2. Сравнение результатов с работы СБА по рисунку вен ладони

Источник	год	модель	dataset	Accuracy (%)	EER(%)
[4]	2021	CNN	Casia	99,4	0,0683
[5]	2021	CNN	Casia, PUT	99,83	0,0556
[6]	2021	GAN+ CNN	Casia, PUT	98 98,82	нет данных
[7]	2021	CNN	MS-PolyU	94,65	нет данных
[8]	2023	CNN	PolyU, Tongji	нет данных	0,28 1,07
[9]	2023	GCNLE+ Attention	PolyU, VERA	98,82 98,27	нет данных
[10]	2023	LST+GAN	PolyU, Tongji	93,6 88,7	нет данных
[11]	2023	ResNet-32, SingleNet	Cassia, PolyU	99,2 100	нет данных
[12]	2024	Deep Autoencoder + Siamese network	Tecnocampus Hand Image	98,79	нет данных
Предложенный	2025	Siamese network	Cassia + Tongji	98,2	0,73

Сравнение результатов показывает, что разработанная СБА уступает по точности системам, разработанным в [4 – 6, 11] из-за ограниченности в вычислительных ресурсах и необходимости применения легковесной нейронной сети, но при этом обеспечена защищенная обработка биометрических данных и возможность использования для встраиваемых систем, результаты [7 – 10] сопоставимы по точности классификации с предложенной СБА. В работе [12] представлена биометрическая система, являющаяся

наиболее близким прототипом, реализованная на одноплатном компьютере Raspberry Pi 3 Model B с CPU. Проведем сравнительный анализ полученных результатов разработанной СБА и прототипа [12] в таблице 3.

Таблица 3 Сравнение параметров с известным подобным решением

Параметр для сравнения	[12]	Предложенный
Обработка биометрических данных	Применение глубокого автоэнкодера для предварительной обработки	Предварительная обработка, выделение рисунка вен
Время обработки биометрических данных	Raspberry Pi 3 – 1,8 с. Desktop GPU -0,36 с.	Raspberry Pi 5 – 0,18 с. Desktop без GPU -0,06 с.
Классификация	Сиамская нейронная сеть	Сиамская нейронная сеть
Время классификации	Raspberry Pi 3– 3,8 с. Desktop GPU – 0,001 с.	Raspberry Pi 5– 0,21 с. Desktop без GPU -0,07 с.
Accuracy	98,79%	98,2%
EER	Нет данных	0,73%
Защита биометрических данных	отсутствует	Гомоморфное шифрование
Время шифрования		Raspberry Pi 5 – 3,62 с. Desktop без GPU - 1,2 с.

Сравнивая разработанную систему с прототипом, можно сказать, предложенная СБА не уступает [12], как по точности, так и по времени обработки, но при этом обеспечивает защищенную обработку данных.

10. Заключение. Разработанная система биометрической аутентификации демонстрирует практическую реализуемость защищённой биометрической аутентификации с применением гомоморфного шифрования. Точность классификации составляет 98,2%, время отклика составляет 1,2 с. при реализации на ПК и 4 с. при реализации на Raspberry Pi 5, что соответствует постановке задачи. Результаты экспериментов подтверждают, что даже при варьировании параметров и наличии искажений входных данных разработанная архитектура сохраняет стабильное поведение, обеспечивая приемлемый уровень ошибок и устойчивость в условиях неполной достоверности входной информации. Это позволяет адаптировать ее под различные

сценарии эксплуатации – от встроенных систем с жесткими ограничениями по вычислительным ресурсам до облачных платформ с повышенными требованиями к точности. Сокращение времени обработки может быть обеспечено как использованием специализированных вычислительных ускорителей (GPU, TPU, FPGA), так и применением других библиотек, поддерживающих гомоморфные вычисления, менее требовательных к вычислительным ресурсам.

Литература

1. Ryu R., Yeom S., Herbert D., Dermoudy J. The design and evaluation of adaptive biometric authentication systems: Current status, challenges and future direction // ICT Express. 2023. vol. 9. no. 6. pp. 1183–1197. DOI: 10.1016/j.icte.2023.04.003
2. Boshoff D., Hancke G.P. A Classifications Framework for Continuous Biometric Authentication (2018–2024) // Computers & Security. 2024. vol. 150. 104285 p.
3. Nayar G.R., Thomas T. Partial palm vein based biometric authentication // Journal of Information Security and Applications. 2023. vol. 72. 103390 p.
4. Obayya M.I., El-Ghandour M., Alrowais F. Contactless Palm Vein Authentication Using Deep Learning With Bayesian Optimization // IEEE Access. 2020. vol. 9. pp. 1940–1957. DOI: 10.1109/ACCESS.2020.3045424.
5. Chen Y.-Y., Hsia C.-H., Chen P.-H. Contactless Multispectral Palm-Vein Recognition With Lightweight Convolutional Neural Network // IEEE Access. 2021. vol. 9. pp. 149796–149806. DOI: 10.1109/ACCESS.2021.3124631.
6. Qin H., El-Yacoubi M.A., Li Y., Liu C. Multi-Scale and Multi-Direction GAN for CNN-Based Single Palm-Vein Identification // IEEE Transactions on Information Forensics and Security. 2021. vol. 16. pp. 2652–2666. DOI: 10.1109/TIFS.2021.3059340.
7. Kaddoun S.S., Aberni Y., Boubchir L., Raddadi M., Daachi B. Convolutional Neural Algorithm for Palm Vein Recognition using ZFNet Architecture // Proceedings of the 4th International Conference on Bio-Engineering for Smart Technologies (BioSMART). 2021. pp. 1–4. DOI: 10.1109/BioSMART54244.2021.9677799.
8. Ma Y., et al. Focal Contrastive Learning for Palm Vein Authentication // IEEE Transactions on Instrumentation and Measurement. 2023. vol. 72. pp. 1–15. DOI: 10.1109/TIM.2023.3304689.
9. Qin H., Gong C., Li Y., Gao X., El-Yacoubi M.A. Label Enhancement-Based Multiscale Transformer for Palm-Vein Recognition // IEEE Transactions on Instrumentation and Measurement. 2023. vol. 72. pp. 1–17. DOI: 10.1109/TIM.2023.3261909.
10. Li Y., Ruan S., Qin H., Deng S., El-Yacoubi M.A. Transformer Based Defense GAN Against Palm-Vein Adversarial Attacks // IEEE Transactions on Information Forensics and Security. 2023. vol. 18. pp. 1509–1523. DOI: 10.1109/TIFS.2023.3243782.
11. Hernandez-Garcia R., Salazar-Jurado E.H., Barrientos R.J., Castro F.M., Ramos-Cozar J., Guil N. From Synthetic Data to Real Palm Vein Identification: a Fine-Tuning Approach // Proceedings of the IEEE 13th International Conference on Pattern Recognition Systems (ICPRS). 2023. pp. 1–7. DOI: 10.1109/ICPRS58416.2023.10179042.
12. Nunes M., Viegas E.K., Santin A.O. Towards a Feasible Palm Vein Verification Scheme Using Deep Autoencoder and Siamese Networks // 2024 International Conference on Machine Learning and Applications (ICMLA). 2024. pp. 483–489. DOI: 10.1109/ICMLA61862.2024.00071.
13. Manikandan V.M. Chapter 11 - A secure biometric authentication system for smart environment using reversible data hiding through encryption scheme // Machine Learning for Biometrics Concepts, Algorithms and Applications Cognitive Data Science in Sustainable Computing. 2022. pp. 201–216.

14. Zeng L., Shen P., Zhu X., Tian X., Chen C. A review of privacy-preserving biometric identification and authentication protocols // Computers & Security. 2025. vol. 150. 104309 p. DOI: 10.1016/j.cose.2024.104309.
15. Adil M., Farouk A., Ali A., Song H., Jin Z. Securing Tomorrow of Next-Generation Technologies with Biometrics, State-of-The-Art Techniques, Open Challenges, and Future Research Directions // Computer Science Review. 2025. vol. 57. 100750 p.
16. Melzi P., Rathgeb C., Tolosana R., Vera-Rodriguez R., Busch C. An Overview of Privacy-Enhancing Technologies in Biometric Recognition // ACM Computing Surveys. 2024. 12 p. DOI: 10.1145/3664596.
17. Wang X., Xie Y., Shui D., Ge S. An improved biometric authentication and key agreement scheme based on fuzzy extractor for Wireless Body Area Networks // Journal of Information Security and Applications. 2025. vol. 91. 104047 p. DOI: 10.1016/j.jisa.2025.104047.
18. Srinivas P.V.V.S, Yadavalli N., DurgaV., Kumar K., Raju P. Enhanced biometric template protection schemes using distance based fuzzy extractor // Computers & Security. 2025. vol. 157. 104573 p.
19. ISO/IEC 24745 / Information security, cybersecurity and privacy protection – Biometric information protection // ISO/IEC. 2022. 63 p.
20. Morampudi M.K., Prasad M.V.N.K., Verma M., Raju U.S.N. Secure and verifiable iris authentication system using fully homomorphic encryption // Computers & Electrical Engineering. 2021. vol. 89. 106924 p. DOI: 10.1016/ j.compeleceng.2020.106924.
21. Wu B., Zheng S., Dai P., Chen J., Yao Y. Searchable face recognition authentication based on homomorphic encryption // Journal of Information Security and Applications. 2025. vol. 94. 104208 p. DOI: 10.1016/j.jisa.2025.104208.
22. Wu L., Wang X.A., Liu J., Su Y., Zheng T., Liu W., Lei H., Tang D., Cao Y., Zhang J. Homomorphic Encryption for Machine Learning Applications with CKKS Algorithms: A Survey of Developments and Applications // Computers, Materials @ Continua. 2025. vol. 85. no. 1. DOI: 10.32604/cmc.2025.064346.
23. Wang H., Liu W., Wang X.A., Jiang W., Liu J., Yang X., Zhao W., Zheng K. Privacy-enhanced facial recognition for IoT based on homomorphic encryption // Internet of Things. 2025. vol. 34. 101757 p. DOI: 10.1016/j.iot.2025.101757.
24. Sumalatha U., Prakasha K.K., Prabhu S., Nayak V.C. A Comprehensive Review of Unimodal and Multimodal Fingerprint Biometric Authentication Systems: Fusion, Attacks, and Template Protection // IEEE Access. 2024. vol. 12. pp. 64300–64334. DOI: 10.1109/ACCESS.2024.3395417.
25. Hemis M., Kheddar H., Bourouis S., Saleem N. Deep learning techniques for hand vein biometrics: A comprehensive review // Information Fusion. 2025. vol. 114. 102716 p. DOI: 10.1016/j.inffus.2024.102716.
26. Kurbatova E., Kharina N., Zemtsov A., Plyaskin S. Investigating Palm Vein Pattern Recognition Methods // Proceedings of the 24th International Conference on Digital Signal Processing and its Applications (DSPA). 2022. pp. 1–5. DOI: 10.1109/DSPA53304.2022.9790783.
27. Boudaoud B.L., Solaiman B., Tari A. A modified ZS thinning algorithm by a hybrid approach // The Visual Computer. 2018. vol. 34. pp. 689–706. DOI: 10.1007/s00371-17-1407-4.
28. Прозоров Д.Е., Земцов А.В. Применение легковесной сиамской нейросети для формирования вектора признаков в системе васкулярной аутентификации // Компьютерная оптика. 2023. Т. 47. №3. С. 433–441.
29. Note on CASIA Palmprint Database // CBSR. URL: http://www.cbsr.ia.ac.cn/english/ Palmprint%20Databases.asp.

Харина Наталья Леонидовна — канд. техн. наук, доцент, доцент кафедры радиоэлектронных средств, ВятГУ. Область научных интересов: компьютерное зрение, биометрическая аутентификация, искусственный интеллект в прикладных решениях, моделирование процессов различной природы с применением теории условных Марковских процессов. Число научных публикаций — 106. harina@vyatsu.ru; ул. Московская, д. 36, г. Киров, 610000, РФ; моб.т. +7(909)140-4746.

Долженкова Мария Львовна — канд. техн. наук, доцент, зав. кафедрой электронно-вычислительных машин, ВятГУ. Область научных интересов: системы поддержки принятия решений на знаниях, технологии глубокого обучения, искусственный интеллект в прикладных решениях. Число научных публикаций — 59. maryd@vyatsu.ru; ул. Московская, д. 36, г. Киров, 610000, РФ; моб.т. +7(912)825-2061.

Иминова Зарина Бахтияржановна — студент 4 курса, ВятГУ. Область научных интересов: компьютерное зрение, биометрическая аутентификация. stud143619@vyatsu.ru; ул. Московская, д. 36, г. Киров, 610000, РФ; моб.т. +7(922)662-7545.

DOI 10.15622/ia.25.3.10

N. HARINA, M. DOLZHENKOVA, Z. IMINOVA

SECURE BIOMETRIC AUTHENTICATION SYSTEM FOR EMBEDDED SYSTEMS WITH LIMITED COMPUTING RESOURCES

Harina N., Dolzhenkova M., Iminova Z. **Secure Biometric Authentication System for Embedded Systems with Limited Computing Resources.**

Abstract. Biometric authentication technologies such as face, fingerprint, iris, and palm vein recognition are currently widely used in banking applications, medical systems, access control systems, and government information resources. The use of such technologies must be accompanied by measures to protect biometric information during processing and storage, especially in the face of the growing number of targeted attacks on personal data. One promising approach to ensuring the security of biometric information is the use of homomorphic encryption. The paper presents an autonomous secure biometric authentication system based on the user's palm vein patterns in the near infrared spectrum for embedded solutions implemented using privacy-preserving technology to protect biometric data. Studies have shown the practical feasibility of a biometric authentication system using a lightweight Siamese neural network to extract a feature vector and homomorphic encryption to ensure confidentiality on processors with limited computing resources based on single-board computers such as Raspberry Pi. Resistance to distortion, including rotations, noise, displacements, and varying encryption parameters, has been confirmed. The performance of the developed system was evaluated on a combined dataset consisting of images from the CASIA-MS-PalmprintV1 and Tongji Contactless Palmvein datasets, yielding the following results: classification accuracy 98.2%, EER 0.73%, AUC 0.96, and processing time 1.2 seconds, which meet the requirements for such systems. Analysis of the results shows that the developed system is inferior in accuracy to client-server solutions based on heavyweight convolutional neural networks due to limited computing resources and the consequent need for a lightweight neural network, but at the same time it provides secure biometric data processing while maintaining an acceptable response time. Reducing the processing time when implemented on single-board computers can be achieved by using both specialized computing accelerators and by employing other libraries that support homomorphic computations and are less demanding on computing resources.

Keywords: biometric authentication, image processing, convolutional neural network, Siamese neural network, privacy-preserving biometric processing, homomorphic encryption.

References

1. Ryu R., Yeom S., Herbert D., Dermoudy J. The design and evaluation of adaptive biometric authentication systems: Current status, challenges and future direction. ICT Express. 2023. vol. 9. no. 6. pp. 1183–1197. DOI: 10.1016/j.icte.2023.04.003
2. Boshoff D., Hancke G.P. A Classifications Framework for Continuous Biometric Authentication (2018–2024). Computers & Security. 2024. vol. 150. 104285 p.
3. Nayar G.R., Thomas T. Partial palm vein based biometric authentication. Journal of Information Security and Applications. 2023. vol. 72. 103390 p.
4. Obayya M.I., El-Ghandour M., Alrowais F. Contactless Palm Vein Authentication Using Deep Learning With Bayesian Optimization. IEEE Access. 2020. vol. 9. pp. 1940–1957. DOI: 10.1109/ACCESS.2020.3045424.

5. Chen Y.-Y., Hsia C.-H., Chen P.-H. Contactless Multispectral Palm-Vein Recognition With Lightweight Convolutional Neural Network. IEEE Access. 2021. vol. 9. pp. 149796–149806. DOI: 10.1109/ACCESS.2021.3124631.
6. Qin H., El-Yacoubi M.A., Li Y., Liu C. Multi-Scale and Multi-Direction GAN for CNN-Based Single Palm-Vein Identification. IEEE Transactions on Information Forensics and Security. 2021. vol. 16. pp. 2652–2666. DOI: 10.1109/TIFS.2021.3059340.
7. Kaddoun S.S., Aberni Y., Boubchir L., Raddadi M., Daachi B. Convolutional Neural Algorithm for Palm Vein Recognition using ZFNet Architecture. Proceedings of the 4th International Conference on Bio-Engineering for Smart Technologies (BioSMART). 2021. pp. 1–4. DOI: 10.1109/BioSMART54244.2021.9677799.
8. Ma Y., et al. Focal Contrastive Learning for Palm Vein Authentication. IEEE Transactions on Instrumentation and Measurement. 2023. vol. 72. pp. 1–15. DOI: 10.1109/TIM.2023.3304689.
9. Qin H., Gong C., Li Y., Gao X., El-Yacoubi M.A. Label Enhancement-Based Multiscale Transformer for Palm-Vein Recognition. IEEE Transactions on Instrumentation and Measurement. 2023. vol. 72. pp. 1–17. DOI: 10.1109/TIM.2023.3261909.
10. Li Y., Ruan S., Qin H., Deng S., El-Yacoubi M.A. Transformer Based Defense GAN Against Palm-Vein Adversarial Attacks. IEEE Transactions on Information Forensics and Security. 2023. vol. 18. pp. 1509–1523. DOI: 10.1109/TIFS.2023.3243782.
11. Hernandez-Garcia R., Salazar-Jurado E.H., Barrientos R.J., Castro F.M., Ramos-Cozar J., Guil N. From Synthetic Data to Real Palm Vein Identification: a Fine-Tuning Approach. Proceedings of the IEEE 13th International Conference on Pattern Recognition Systems (ICPRS). 2023. pp. 1–7. DOI: 10.1109/ICPRS58416.2023.10179042.
12. Nunes M., Viegas E.K., Santin A.O. Towards a Feasible Palm Vein Verification Scheme Using Deep Autoencoder and Siamese Networks. 2024 International Conference on Machine Learning and Applications (ICMLA). 2024. pp. 483–489. DOI: 10.1109/ICMLA61862.2024.00071.
13. Manikandan V.M. Chapter 11 - A secure biometric authentication system for smart environment using reversible data hiding through encryption scheme. Machine Learning for Biometrics Concepts, Algorithms and Applications Cognitive Data Science in Sustainable Computing. 2022. pp. 201–216.
14. Zeng L., Shen P., Zhu X., Tian X., Chen C. A review of privacy-preserving biometric identification and authentication protocols. Computers & Security. 2025. vol. 150. 104309 p. DOI: 10.1016/j.cose.2024.104309.
15. Adil M., Farouk A., Ali A., Song H., Jin Z. Securing Tomorrow of Next-Generation Technologies with Biometrics, State-of-The-Art Techniques, Open Challenges, and Future Research Directions. Computer Science Review. 2025. vol. 57. 100750 p.
16. Melzi P., Rathgeb C., Tolosana R., Vera-Rodriguez R., Busch C. An Overview of Privacy-Enhancing Technologies in Biometric Recognition. ACM Computing Surveys. 2024. 12 p. DOI: 10.1145/3664596.
17. Wang X., Xie Y., Shui D., Ge S. An improved biometric authentication and key agreement scheme based on fuzzy extractor for Wireless Body Area Networks. Journal of Information Security and Applications. 2025. vol. 91. 104047 p. DOI: 10.1016/j.jisa.2025.104047.
18. Srinivas P.V.V.S, Yadavalli N., DurgaV., Kumar K., Raju P. Enhanced biometric template protection schemes using distance based fuzzy extractor. Computers & Security. 2025. vol. 157. 104573 p.
19. ISO.IEC 24745. Information security, cybersecurity and privacy protection – Biometric information protection. ISO.IEC. 2022. 63 p.

20. Morampudi M.K., Prasad M.V.N.K., Verma M., Raju U.S.N. Secure and verifiable iris authentication system using fully homomorphic encryption. Computers & Electrical Engineering. 2021. vol. 89. 106924 p. DOI: 10.1016/j.compeleceng.2020.106924.
21. Wu B., Zheng S., Dai P., Chen J., Yao Y. Searchable face recognition authentication based on homomorphic encryption. Journal of Information Security and Applications. 2025. vol. 94. 104208 p. DOI: 10.1016/j.jisa.2025.104208.
22. Wu L., Wang X.A., Liu J., Su Y., Zheng T., Liu W., Lei H., Tang D., Cao Y., Zhang J. Homomorphic Encryption for Machine Learning Applications with CKKS Algorithms: A Survey of Developments and Applications. Computers, Materials @ Continua. 2025. vol. 85. no. 1. DOI: 10.32604/cmc.2025.064346.
23. Wang H., Liu W., Wang X.A., Jiang W., Liu J., Yang X., Zhao W., Zheng K. Privacy-enhanced facial recognition for IoT based on homomorphic encryption. Internet of Things. 2025. vol. 34. 101757 p. DOI: 10.1016/j.iot.2025.101757.
24. Sumalatha U., Prakasha K.K., Prabhu S., Nayak V.C. A Comprehensive Review of Unimodal and Multimodal Fingerprint Biometric Authentication Systems: Fusion, Attacks, and Template Protection. IEEE Access. 2024. vol. 12. pp. 64300–64334. DOI: 10.1109/ACCESS.2024.3395417.
25. Hemis M., Kheddar H., Bourouis S., Saleem N. Deep learning techniques for hand vein biometrics: A comprehensive review. Information Fusion. 2025. vol. 114. 102716 p. DOI: 10.1016/j.inffus.2024.102716.
26. Kurbatova E., Kharina N., Zemtsov A., Plyaskin S. Investigating Palm Vein Pattern Recognition Methods. Proceedings of the 24th International Conference on Digital Signal Processing and its Applications (DSPA). 2022. pp. 1–5. DOI: 10.1109/DSPA53304.2022.9790783.
27. Boudaoud B.L., Solaiman B., Tari A. A modified ZS thinning algorithm by a hybrid approach. The Visual Computer. 2018. vol. 34. pp. 689–706. DOI: 10.1007/s00371-17-1407-4.
28. Prozorov D.E., Zemtsov A.V. [Application of a lightweight Siamese neural network to generate a feature vector in a vascular authentication system]. Kompyuternaya optika – Computer optics. 2023. vol. 47. no. 3. pp. 433–441. (In Russ.).
29. Note on CASIA Palmprint Database. CBSR. Available at: http://www.cbsr.ia.ac.cn/english/Palmprint%20Databases.asp

Harina Natalya — Ph.D., Associate professor, Docent, Department of Radio Electronic Devices, Federal State Budgetary Educational Institution of Higher Education "Vyatka State University" (VyatSU). Research interests: computer vision, biometric authentication, artificial intelligence in applied solutions, modeling of processes of various nature using the theory of conditional Markov processes. The number of publications — 106. harina@vyatsu.ru; 36, Moskovskaya St., 610000, Kirov, Russia; office phone: +7(909)140-4746.

Dolzhenkova Maria — Ph.D., Associate professor, Head of the Department, Department of Electronic Computing Machines, Federal State Budgetary Educational Institution of Higher Education "Vyatka State University" (VyatSU). Research interests: knowledge-based decision support systems, deep learning technologies, artificial intelligence in applied solutions. The number of publications — 59. maryd@vyatsu.ru; 36, Moskovskaya St., 610000, Kirov, Russia; office phone: +7(912)825-2061

Iminova Zarina — Student, Federal State Budgetary Educational Institution of Higher Education "Vyatka State University" (VyatSU). Research interests: computer vision, biometricauthentication. The number of publications — 0. stud143619@vyatsu.ru; 36, Moskovskaya St., 610000, Kirov, Russia; office phone: +7(922)662-7545.

УДК512.552.18+003.26 **DOI** 10.15622/ia.25.3.11

А.А. Молдовян, Н.А. Молдовян, Д.Н. Молдовян, А.А. Костина

ПОСТКВАНТОВЫЙ АЛГОРИТМ ЭЦП НА ОСНОВЕ ТРУДНОСТИ РЕШЕНИЯ СТЕПЕННЫХ УРАВНЕНИЙ

Молдовян А.А., Молдовян Н.А., Молдовян Д.Н., Костина А.А. **Постквантовый алгоритм ЭЦП на основе трудности решения степенных уравнений.**

Аннотация. Вычислительная сложность нахождения решений больших систем нелинейных уравнений (БСНУ) лежит в основе ряда постквантовых двухключевых криптосхем, включая алгебраические алгоритмы цифровой подписи с использованием матриц в качестве элементов секретного ключа. Такие алгоритмы относятся к вероятностным криптосхемам и характеризуются использованием некоторой матрицы **S** в качестве подгоночного элемента подписи. Последнее обусловливает актуальность рассмотрения стойкости к атакам, использующим известные подписи. При этом лобовой (прямой) атакой является атака на основе решения БСНУ. В известных алгебраических алгоритмах цифровой подписи в рамках лобовой атаки возникают недоопределенные БСНУ с большим числом решений, что создает предпосылки к атакам на основе эквивалентных ключей. В статье рассматривается построение постквантового алгоритма на конечных алгебрах квадратных матриц, при прямой атаке на который возникают близкие к сбалансированным (с равным числом уравнений и неизвестных) и переопределенные БСНУ (с числом уравнений, превышающим число неизвестных). Особенностью предложенного алгоритма является использование вспомогательных скрытых групп, вспомогательного параметра рандомизации ЭЦП, вычисляемого как значение сжимающей односторонней функции от значения **S**. Приводится оценка стойкости к лобовой атаке и нескольким вариантам атак на основе известных подписей.

Ключевые слова: постквантовая криптография, алгоритм ЭЦП, конечная алгебра матриц, система степенных уравнений, секретная группа.

1. Введение. Разработка практичных криптоалгоритмов с открытым ключом, обладающих достаточно высоким уровнем стойкости к атакам с использованием квантового вычислителя, является актуальной задачей постквантовой криптографии [1 – 3]. Существуют эффективные алгоритмы решения задачи дискретного логарифмирования и факторизации [4, 5], поэтому в качестве постквантовых примитивов используются другие вычислительно сложные задачи. Направления постквантовой криптографии задаются видом используемой вычислительно трудной задачи. Достаточно многочисленные исследования посвящены разработке двухключевых криптосхем на корректирующих кодах [6, 7], включая алгоритмы электронной цифровой подписи (ЭЦП) на основе трудности синдромного декодирования [8, 9], построенные путем преобразования протоколов с нулевым разглашением в схему ЭЦП. Однако такое построение приводит к большому размеру открытого ключа (более 600 Кбайт в [8] и более 150 Кбайт в [9] при уровне стойкости 2^{80}). С целью уменьшения длины открытого ключа предложены алгоритмы ЭЦП

на основе задачи синдромного декодирования ограниченных ошибок [10, 11], что позволило существенно уменьшить размер подписи (до 17 Кбайт в [12] и до 11 Кбайт в [13] при уровне стойкости 2^{128}).

В работе [14] рассматривается построение постквантовых алгоритмов ЭЦП на группах, в работе [15] – на булевых функциях. Перспективно использование алгебраических решеток [16, 17] и трудно обратимых функций [18] в качестве носителей постквантовых криптосхем.

Большое число работ посвящено разработке и анализу стойкости криптосхем с открытым ключом, включая алгоритмы ЭЦП, на трудно обратимых отображениях с потайной лазейкой [19 – 21]. Отображения данного типа служат открытым ключом, имеющим очень большой размер, и задаются в виде набора степенных многочленов над конечным полем сравнительно малого порядка. Для уменьшения размера открытого ключа многочлены описываются в виде упорядоченного набора коэффициентов и обычно используются многочлены второй степени. Для выполнения криптографических преобразований входные сообщения (в алгоритмах шифрования по открытому ключу) и значения хеш-функции от подписываемого сообщения (в алгоритмах ЭЦП) представляются в виде многомерных векторов. Прямой атакой на криптоалгоритмы данного типа является вычисление вектора-прообраза по известному выходному вектору, а именно, реализация обратного отображения без знания потайной лазейки, для чего требуется найти решение большой системы нелинейных уравнений (БСНУ), которая определяется набором многочленов, заданным как открытый ключ.

Для квантовых вычислителей неизвестны эффективные алгоритмы решения БСНУ, поэтому криптосхемы на отображениях, являющиеся стойкими в обычном смысле, являются постквантовыми. Однако криптосхемы данного вида обладают существенным с практической точки зрения недостатком – непрактично большим размером открытого ключа. Парадигма [22, 23] построения трудно обратимого отображения с потайной лазейкой как операции возведения в степень $n = 2, 3$ в векторном конечном поле с секретной модификацией и определению потайной лазейки, как операции извлечения корня степени n позволяет сократить длину открытого ключа в несколько десятков раз, но этот недостаток в полной мере не устраняется.

Парадигма разработки алгебраических схем подписи [24, 25], использующих вычислительную трудность решения БСНУ, обеспечивает малые длины открытого ключа и подписи с общей длиной менее 1 Кбайт при уровне стойкости 2^{256}. В качестве носителя

схем ЭЦП данного типа используются некоммутативные конечные алгебры с глобальной единицей, в которых определена ассоциативная операция умножения векторов.

В работе [25] подгоночный элемент ЭЦП в виде вектора **S** вычисляется по формуле:

$$\mathbf{S} = \mathbf{D}\mathbf{J}^{n}\mathbf{V}\mathbf{J}^{d}\mathbf{F}, \quad (1)$$

где **J** – генератор скрытой коммутативной группы <**J**>; $\mathbf{J}^{n}$ и $\mathbf{J}^{d}$ – векторы, выбираемые из группы <**J**> по случайным значениям многоразрядных неотрицательных целых чисел n и d; **D**, **V** и **F** – векторы, являющиеся элементами секретного ключа, которые некоммутативны с векторами из группы <**J**>.

В алгоритме из статьи [26] значение подгоночного вектора **S** задается выражением:

$$\mathbf{S} = \mathbf{D}\mathbf{G}^{n}\mathbf{L}^{u}\mathbf{J}^{d}\mathbf{F}, \quad (2)$$

где **G** и **J** – взаимно некоммутативные генераторы двух скрытых коммутативных групп <**G**> и <**J**>; $\mathbf{G}^{n}$ и $\mathbf{J}^{d}$ – векторы, выбираемые по случайным значениями случайных неотрицательных целых чисел n и d, имеющих большую разрядность (100 бит и более); **D** и **F** – элементы секретного ключа, используемые в качестве левого и правого маскирующих множителей, которые некоммутативны между собой и с каждым из векторов **G** и **J**; $\mathbf{L}^{u}$ – скалярный вектор, задаваемый как случайная степень u скалярного вектора **L**, являющегося элементом секретного ключа.

Другой важной особенностью алгоритмов ЭЦП с секретной группой является многократное вхождение вектора **S** в уравнение верификации. Этот прием позволяет обеспечить стойкость к подделке подписи путем решения проверочного уравнения относительно неизвестного значения **S**. Для обеспечения защищенности к такой атаке в некоторых алгоритмах [26] дополнительно используется операция возведения в степень $\rho = \Phi(\mathbf{S})$, где $\Phi(\mathbf{S})$ – некоторая сжимающая однонаправленная функция, например, коллизионно стойкая хэш-функция. В этом случае степень ρ играет роль дополнительного рандомизирующего параметра, который в явном виде не входит в набор значений, составляющих подпись. Поскольку значение ρ вычисляется после получения значения подгоночного элемента **S** и вносит дополнительную рандомизацию, то в проверочное

уравнение указанных алгоритмов ЭЦП требуется включить дополнительный подгоночный элемент подписи.

Формирование подгоночного элемента подписи **S** как по формуле (1), так и по формуле (2) позволяет задать достаточно высокий уровень стойкости к криптоанализу на основе известных подписей, однако прямая атака на алгоритмы из статей [24, 25] связана с решением недоопределенных БСНУ, в которых число степенных уравнений существенно меньше числа неизвестных. Последнее означает наличие большого числа решений, которые соответствуют различным секретным ключам, связанным с открытым ключом. Это означает наличие большого числа эквивалентных секретных ключей, а знание одного из них потенциально позволяет вычислить правильную подпись к произвольному электронному документу.

Несмотря на то, что вычисление даже одного эквивалентного ключа в случае алгоритмов [24, 25] является вычислительно невыполнимым, сам факт существования большого числа эквивалентных ключей создает предпосылки для потенциальных вычислительно эффективных атак с использованием эквивалентных ключей, например, подобных атаке [27], свободной от решения БСНУ. Кроме того, в случае недоопределенных БСНУ в рамках прямой атаки возникает возможность подбора и фиксирования значений части неизвестных, при которых потенциально снижается сложность решения БСНУ.

Потенциальные возможности использования указанных предпосылок в общем случае существенно нивелируются в алгоритмах, для которых БСНУ является сбалансированной (число уравнений равно числу неизвестных) или переопределенной (число неизвестных меньше числа уравнений). Говорить о полном устранении потенциальных атак на основе эквивалентных ключей будет некорректным, так как для этого надо было бы показать, что не существуют варианты представления связи заданного открытого ключа с некоторым секретным ключом с помощью альтернативных формул при обеспечении корректности работы модифицированной схемы подписи и сохранении заданного проверочного уравнения. Однако в случае алгебраических алгоритмов с секретной группой, в которых прямая атака связана с нахождением решения сбалансированных и переопределенных БСНУ, можно ожидать значительного повышения уровня стойкости к потенциальным атакам, использующих факт наличия эквивалентных ключей.

В настоящей работе рассматривается построение постквантового алгоритма ЭЦП на алгебре матриц с использованием вычислительной трудности решения сбалансированных или

переопределенных БСНУ (в зависимости от размерности используемых матриц).

2. Алгебраический носитель. Конечные алгебры матриц размера $\mu\times\mu$ могут быть рассмотрены как частные случаи конечных некоммутативных алгебр (размерности μ^2) с ассоциативной операцией умножения, задаваемой по специальным сильно прореженным таблицам умножения базисных векторов, которые легко записать по правилам матричного умножения. При задании матриц над конечным полем $GF(p)$ имеем конечную алгебру, представляющую собой некоммутативное конечное кольцо с единичным элементом. Совокупность всех невырожденных матриц образует мультипликативную группу этого кольца, порядок которой Ω, выражается следующей формулой:

$$\Omega = \prod_{i=0}^{\mu-1} p^i\,(p^{\mu-i}-1). \tag{3}$$

Если некоторое простое число q является делителем порядка Ω, то, согласно теореме Коши, множество невырожденных матриц содержит матрицу порядка q. В зависимости от значения простого числа p значение выражений в скобках формулы (3) содержит простые делители, разрядность которых может быть как меньше, так и значительно больше разрядности p. Максимальный размер делителя q, который задается специальным выбором простого числа p, имеет место для простых значений μ. При этом значение q выражается формулой:

$$q = p^{\mu-1} + p^{\mu-2} + p + 1. \tag{4}$$

При разрядности числа p, превышающей 15 бит, для случаев $\mu = 2, 3, 5, \ldots, 13$, используя формулу (2) путем многократного тестирования случайно выбираемых простых чисел r, имеющих разрядность $|r|$, легко найти простые числа $p = 2r + 1$ разрядности $|r| + 1$, для которых значение q, вычисленное по формуле (4), также является простым. Примеры нужных троек простых чисел (r, p, q) для значения $\mu = 3$ приведены в статье [25].

Существование матрицы **G** простого порядка q означает существование большого числа матриц **Y** такого порядка, которые могут быть найдены по формуле $\mathbf{Y} = \mathbf{X}^{-1}\mathbf{G}\mathbf{X}$ для каждой невырожденной матрицы **X**, некоммутативной с **G.** Матрица **Y** генерирует циклическую

группу <**Y**> порядка q. Легко показать, что матрица $\pi\mathbf{Y}$, где π является примитивным элементом в $GF(p)$, генерирует циклическую группу $<\pi\mathbf{Y}>$ порядка $p^{\mu-1}-1$. Таким образом, при упомянутом специальном выборе простого числа p в конечной алгебре матриц существуют множество попарно некоммутативных матриц достаточно больших простых порядков. Этот факт далее используется при построении постквантового алгоритма ЭЦП.

При рассмотрении вопроса стойкости разработанного алгоритма возникает задача сведения системы матричных уравнений к системе скалярных уравнений с минимизацией числа скалярных неизвестных. С неизвестной матрицей связаны μ^2 неизвестных элементов поля $GF(p)$, однако в случае выбора η случайных матриц из секретной коммутативной группы неопределенность в значительной степени снимается, благодаря заданию выбора из достаточно ограниченного подмножества матриц по сравнению с множеством всех невырожденных матриц, мощность которого равна Ω (см. формулу (3)). Указанные η матриц вносят всего $\mu^2 + (\eta - 1)\mu$ скалярных неизвестных. Действительно, генератор секретной группы является неизвестным, поэтому он задает μ^2 скалярных неизвестных. При этом, элементы каждой из остальных $\eta - 1$ неизвестных матриц могут быть описаны через значения элементов матрицы-генератора и μ скалярных переменных.

Последнее связано с тем, что множество матриц, входящих в коммутативную группу (в том числе группа, генерируемая матрицей, порядок которой равен многоразрядному натуральному числу), может быть описано по элементам некоторой фиксированной матрицы-представителя и μ скалярным переменным. В качестве такого представителя может быть взят, например, генератор **G** циклической группы <**G**>. На самом деле μ скалярные переменные, каждая из которых принимает значения $0, 1, \ldots, p-1$, описывают множество матриц, образующее коммутативное подкольцо порядка p^{μ}. При этом описывается и любое подмножество матриц, содержащихся в этом подкольце, включая и коммутативную группу <**G**>.

Действительно, рассмотрим матрицу $\mathbf{A} = \|a_{i,j}\|$, в которой задана нумерация строк и столбцов значениями $i,j = 0, 1, \ldots, \mu - 1$, следующего вида:

$$A = \|a_{i,j}\| = \begin{pmatrix} a & b & \ldots & z \\ z & a & b & \ldots \\ \ldots & z & a & b \\ b & \ldots & z & a \end{pmatrix}. \qquad (5)$$

Матрица такого вида задается следующим равенством, имеющим место для каждой фиксированной пары индексов i,j при всех значениях $k = 0, 1, \ldots, \mu - 1$:

$$a_{i,j} = a_{i+k,j+k}. \tag{6}$$

Утверждение. Для любого значения $\mu \geq 2$ и произвольных матриц $\mathbf{A} = \|a_{i,j}\|$ и $\mathbf{A}' = \|a'_{i,j}\|$ вида (5) имеет место равенство $\mathbf{AA}' = \mathbf{A}'\mathbf{A}$ и матрица $\mathbf{B} = \mathbf{AA}' = \|b_{i,j}\|$ также имеет вид (5).

Доказательство. Обозначим нулевую строку матрицы $\mathbf{A}$ ($\mathbf{A}'$) как последовательность $\alpha_0, \alpha_1, \ldots, \alpha_{\mu-1}$ ($\alpha'_0, \alpha'_1, \ldots, \alpha'_{\mu-1}$). Тогда можно записать $a_{i,j} = \alpha_{j-i}$ и $a'_{i,j} = \alpha'_{j-i}$ (где вычитание выполняется по модулю μ). В соответствии с правилами матричного умножения имеем для элементов матриц $\mathbf{B} = \mathbf{AA}' = \|b_{i,j}\|$ и $\mathbf{B}' = \mathbf{A}'\mathbf{A} = \|b'_{i,j}\|$ следующее:

$$b_{i,j} = \sum_{d=0}^{\mu-1} \alpha_{i,d}\alpha'_{d,j} = \sum_{d=0}^{\mu-1} \alpha_{d-i}\alpha'_{j-d}, \tag{7}$$

$$b'_{i,j} = \sum_{d=0}^{\mu-1} \alpha'_{i,d}\alpha_{d,j} = \sum_{d=0}^{\mu-1} \alpha'_{d-i}\alpha_{j-d}. \tag{8}$$

Подставляя в (8) вместо d индекс $d' = (i + j - d) \bmod \mu$, получаем:

$$b'_{i,j} = \sum_{d=0}^{\mu-1} \alpha'_{d-i}\alpha_{j-d} = \sum_{d'=i+j}^{i+j-\mu+1} \alpha'_{j-d'}\alpha_{d'-i}. \tag{9}$$

Легко видеть, что правые части в равенствах (7) и (9) отличаются только порядком слагаемых, поэтому $b'_{i,j} = b_{i,j}$. Поскольку последнее равенство имеет место для всевозможных пар индексов (i,j), то $\mathbf{AA}' = \mathbf{A}'\mathbf{A}$.

В соответствии с формулой (8) для элемента $b_{i+k,j+k}$ (где сложение выполняется по модулю μ) матрицы $\mathbf{B} = \mathbf{AA}'$ можно записать:

$$b_{i+k,j+k} = \sum_{d=0}^{\mu-1} \alpha_{d-i-k}\alpha'_{j+k-d} = \sum_{d=0}^{\mu-1} \alpha_{(d-k)-i}\alpha'_{j-(d-k)}. \tag{10}$$

Сделав замену переменных по формуле $d' = d - k \bmod \mu$, легко видеть, что правые части в равенствах (7) и (10) отличаются только порядком слагаемых, поэтому $b_{i+k,j+k} = b_{i,j}$, т. е. матрица $\mathbf{B}$ относится к матрицам типа (5).

Всевозможные матрицы вида (5) образуют коммутативное подкольцо порядка p^{μ}. Каждая матрица $\mathbf{W} = \|w_{i,j}\|$ этого подкольца может служить его представителем, через который могут быть выражены все матрицы $\mathbf{W}' = \|w'_{i,j}\|$, содержащиеся в подкольце, с помощью μ скалярных переменных $\pi_0, \pi_1, \ldots, \pi_{\mu-1}$, каждая из которых принимает значения $0, 1, \ldots, p-1$. Такое описание задается, например, следующей формулой:

$$w'_{i,j} = w_{i,j} + \pi_{j-i}. \tag{11}$$

Если рассмотреть матрицу $\mathbf{\Pi} = \|\pi_{i,j}\|$ типа (5) с нулевой строкой $\pi_0, \pi_1, \ldots, \pi_{\mu-1}$, то формулу (11) можно записать в виде $\mathbf{W}' = \mathbf{W} + \mathbf{\Pi}$. Из последней формулы для фиксированной невырожденной матрицы $\mathbf{X}$, некоммутативной с матрицей $\mathbf{W}$, имеем:

$$\mathbf{Y}' = \mathbf{X}^{-1}\mathbf{W}'\mathbf{X} = \mathbf{X}^{-1}\mathbf{W}\mathbf{X} + \mathbf{X}^{-1}\mathbf{\Pi}\mathbf{X}, \tag{12}$$

где матрицы $\mathbf{Y}'$ пробегают все значения в некотором коммутативном подкольце порядка p^{μ}, матрицы в котором имеют общий вид, а матрица $\mathbf{Y} = \mathbf{X}^{-1}\mathbf{W}\mathbf{X}$ является представителем этого подкольца, по которому можно описать все значения последнего с помощью скалярных переменных $\pi_0, \pi_1, \ldots, \pi_{\mu-1}$. Действительно, каждый элемент матрицы $\mathbf{X}^{-1}\mathbf{\Pi}\mathbf{X}$ в (12) выражается через элементы матрицы $\mathbf{X}$ и некоторую линейную комбинацию переменных $\pi_0, \pi_1, \ldots, \pi_{\mu-1}$. Для предполагаемого криптоаналитика вывод таких формул не составит значительной трудности в сравнении с вычислительной сложностью решения БСНУ.

3. Разработанный алгоритм ЭЦП. Основными вариантами используемого алгебраического носителя являются конечные алгебры матриц размера $\mu \times \mu$ при простых значениях $\mu = 2, 3, 5, 7$ и 11. При этом алгебра матриц задается над полем $GF(p)$, размер порядка которого составляет $|p| = 64$ бит (при $\mu = 3$), $|p| = 32$ бит ($\mu = 5$) и $|p| = 24$ бит ($\mu = 7$). Конкретное значение p берется таким, что $p = 2r + 1$ при некотором простом значении r, а значение q, вычисленное по формуле (4) также является простым. Выбор таких значений простого числа p позволяет существенно снизить

вычислительную сложность процедуры генерации секретного и открытого ключей. Также рассматриваются параметры реализации разработанной схемы ЭЦП на алгебрах матриц 4×4 и 6×6, однако эти две версии предложенного алгоритма требуют внесения некоторых изменений в процедуру формирования секретного и открытого ключа, которые связаны с тем, что для этих двух случаев формула (4) для получения максимального простого делителя мультипликативной группы алгебры матриц не имеет силы.

Практический интерес представляет также задание матриц над полями характеристики два $GF(2^g)$ при специально выбранных значениях степени расширения g двоичного поля. В настоящей работе рассматривается случай задания алгебры матриц над полем $GF(p)$. Секретный ключ формируется следующим образом:

1. Генерируются случайные невырожденные матрицы **G** и **J** порядка q, которые некоммутативны между собой.

2. Генерируются случайные попарно некоммутативные невырожденные матрицы **A**, **B**, **D** и **F**, которые также некоммутативны с матрицами **G** и **J**.

3. Выбираются два случайных неотрицательных целых числа $x < q$ и $z < q$, таких, что матрица:

$$\mathbf{Q} = \mathbf{G}^x\mathbf{J}^z, \tag{13}$$

имеет порядок $\omega = rq$ и некоммутативна с **A**, **B**, **D**, **F**, **G** и **J**.

4. Генерируются случайные целые числа $u < q$ *и* $v < q$, такие, что матрица:

$$\mathbf{U} = \mathbf{G}^u\mathbf{J}^v, \tag{14}$$

некоммутативна с матрицами **A**, **B**, **D**, **F**, **G**, **J**, **Q** и имеет порядок $\omega = rq$.

5. Выбираются случайные натуральные числа λ, β, α, γ, τ и ε, удовлетворяющие условиям $\lambda < \omega$, $\beta < \omega$, $\alpha < q$, $\gamma < q$, $\tau < q$ и $\varepsilon < q$.

Множество матриц (**A**, **B**, **D**, **F**, **G**, **J**, **Q**, **U**) и набор чисел (x, z, u, v, λ, β, α, γ, τ, ε) составляют секретный ключ, который используется для вычисления набора из девяти матриц (**Y**, **Z**, **K**, **N**, **V**, $\mathbf{T}_1$, $\mathbf{T}_2$, $\mathbf{T}_3$) по следующим формулам формирования открытого ключа:

$$\mathbf{Y} = \mathbf{A}\mathbf{G}^{\alpha}\mathbf{A}^{-1}; \qquad \mathbf{Z} = \mathbf{B}\mathbf{J}^{\alpha}\mathbf{B}^{-1}; \qquad \mathbf{K} = \mathbf{A}\mathbf{Q}^{\lambda}\mathbf{A}^{-1}; \tag{15}$$

$$\mathbf{N} = \mathbf{B}\mathbf{U}^{\lambda}\mathbf{B}^{-1}; \qquad \mathbf{V} = \mathbf{D}\mathbf{G}\mathbf{D}^{-1}; \qquad \mathbf{W} = \mathbf{F}^{-1}\mathbf{J}\mathbf{F}; \tag{16}$$

$$\mathbf{T}_1 = \mathbf{A}\mathbf{Q}^{\beta}\mathbf{G}^{\gamma}\mathbf{B}^{-1}; \quad \mathbf{T}_2 = \mathbf{F}^{-1}\mathbf{J}^{\varepsilon}\mathbf{B}^{-1}; \quad \mathbf{T}_3 = \mathbf{B}\mathbf{U}^{\beta}\mathbf{G}^{\tau}\mathbf{D}^{-1}. \tag{17}$$

Основной вклад в вычислительную трудоемкость процедуры формирования секретного ключа вносят шаги 3 и 4, на каждом из которых осуществляются многочисленные проверки различных пар целочисленных степеней (x, z) и (u, v), соответственно. Размеры открытого и секретного ключей представлены в таблице 1 (следует отметить, что для случаев $\mu = 4$ и $\mu = 6$ простое значение q вычисляется по формуле, отличной от (4)).

Таблица 1. Разрядность параметров p, q, ω и длина открытого и секретного ключей

μ	$\lvert p\rvert$, бит	$\lvert q\rvert$, бит	$\lvert\omega\rvert$, бит	Длина ключа, байт	
				Секретного	Открытого
2	128	128	255	672	576
3	64	128	191	754	648
4	64	128	254	1216	1152
5	32	128	159	968	900
6	32	128	190	1232	1296
7	24	144	167	1362	1323
11	16	160	175	2140	2178

Генерация цифровой подписи к электронному документу M выполняется по секретному ключу и следующему алгоритму (Рис. 1):

1. Выбрать случайные неотрицательные целочисленные значения k, t, w (меньшие, чем q) и y (меньшие, чем ω) и вычислить матрицы:

$$\mathbf{R}_1 = \mathbf{A}\mathbf{G}^{k}\mathbf{Q}^{y}\mathbf{J}^{t}\mathbf{B}^{-1}. \tag{18}$$

$$\mathbf{R}_2 = \mathbf{B}\mathbf{J}^{k}\mathbf{U}^{y}\mathbf{J}^{w}\mathbf{B}^{-1}. \tag{19}$$

2. Вычислить рандомизирующий элемент ЭЦП в виде 512-битного хеш-значения $e = e_1||e_2||e_3||e_4 = \Phi(M||\mathbf{R_1}||\mathbf{R_2})$, где Φ – некоторая специфицированная коллизионно стойкая хеш-функция, а значение e представлено в виде конкатенации четырех 128-битных чисел e_1, e_2, e_3 и e_4.

3. Вычислить подгоночные элементы подписи ξ и ψ:

$$\xi = k(e_1\alpha)^{-1} \bmod q. \tag{20}$$

$$\psi = (y - \beta - 1)(e_2\lambda)^{-1} \bmod \omega. \quad (21)$$

4. Найти значения третьего подгоночного элемента ЭЦП s и подгоночной степени n путем решения следующей системы из двух уравнений первой степени:

$$\begin{cases} \gamma + e_3 s + n = x \bmod q, \\ \tau + e_4 s + n = u \bmod q. \end{cases} \quad (22)$$

Требуемые значения устанавливаются по следующим двум формулам:

$$s = (u - \tau - x + \gamma)(e_4 - e_3)^{-1} \bmod q, \quad (23)$$

$$n = x - \gamma - e_3 s \bmod q. \quad (24)$$

5. Найти подгоночную степень d по формуле:

$$d = z + t - \varepsilon - \alpha e_2 \bmod q. \quad (25)$$

6. Вычислить значение подгоночной матрицы **S** по формуле:

$$\mathbf{S} = \mathbf{D}\mathbf{G}^n\mathbf{J}^d\mathbf{F}. \quad (26)$$

7. Вычислить значение $\rho = \Phi(S) \bmod q$, вносящее дополнительную рандомизацию в процесс формирования ЭЦП.

8. Найти значение вспомогательного подгоночного элемента ЭЦП в виде неотрицательного целого числа σ:

$$\sigma = w - d + v - \varepsilon - \alpha\rho \bmod q. \quad (27)$$

9. Вывести набор значений $(e, \xi, \psi, s, \sigma, \mathbf{S})$ в качестве сформированной цифровой подписи.

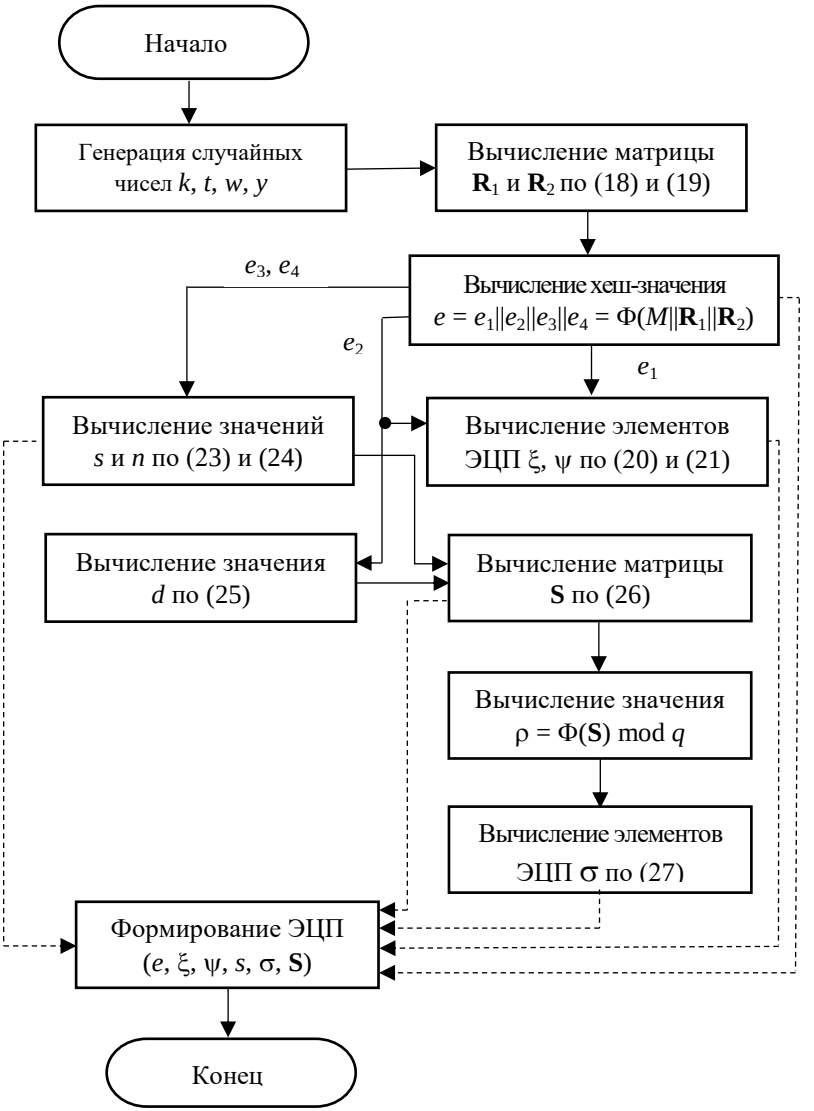


Рис. 1. Генерация ЭЦП

Размеры ЭЦП и отдельных ее элементов показаны в таблице 2 для различных версий разработанного алгоритма, отличающихся различными значениями параметров μ и p. Вычислительная сложность алгоритма генерации подписи в основном определяется восемью операциями возведения в целочисленную степень большого размера, выполняемыми в конечной алгебре матриц. Приближенные оценки вычислительной сложности представлены в таблице 3 с учетом зависимости сложности операций экспоненцирования от разрядности степени и параметров μ и $|p|$.

Таблица 2. Длина ЭЦП и ее отдельных элементов в байтах

μ	e	ξ	ψ	s	σ	**S**	ЭЦП
2	64	16	32	16	16	64	204
3	64	16	24	16	16	72	208
4	64	16	32	16	16	128	268
5	64	16	20	16	16	100	232
6	64	16	24	16	16	144	280
7	64	18	21	18	18	147	286
11	64	20	22	20	20	242	388

Процедура проверки подлинности ЭЦП (e, ξ, ψ, s, σ, **S**) к документу M выполняется с использованием открытого ключа (**Y**, **Z**, **K**, **N**, **V**, $\mathbf{T}_1$, $\mathbf{T}_2$, $\mathbf{T}_3$) по следующему алгоритму (Рис. 2):

1. Вычислить первую проверочную матрицу $\mathbf{R}_1'$ по формуле:

$$R_1^{'} = Y^{e_1\xi} K^{e_2\psi} T_1 V^{e_3 s} \mathrm{ST}_2 Z^{e_2}. \quad (28)$$

2. Вычислить значение $\rho = \Phi(\mathbf{S}) \bmod q$ и вторую проверочную матрицу $\mathbf{R}_2'$:

$$R_2^{'} = Z^{e_1\xi} N^{e_2\psi} T_3 V^{e_4 s} \mathrm{SW}^{\sigma} T_2 Z^{\rho}. \quad (29)$$

3. Вычислить $e' = e'_1||e'_2||e'_3||e'_4 = \Phi(M||\mathbf{R}_1'||\mathbf{R}_2')$.

4. Если $e' = e$, то подпись подлинная, иначе – ложная.

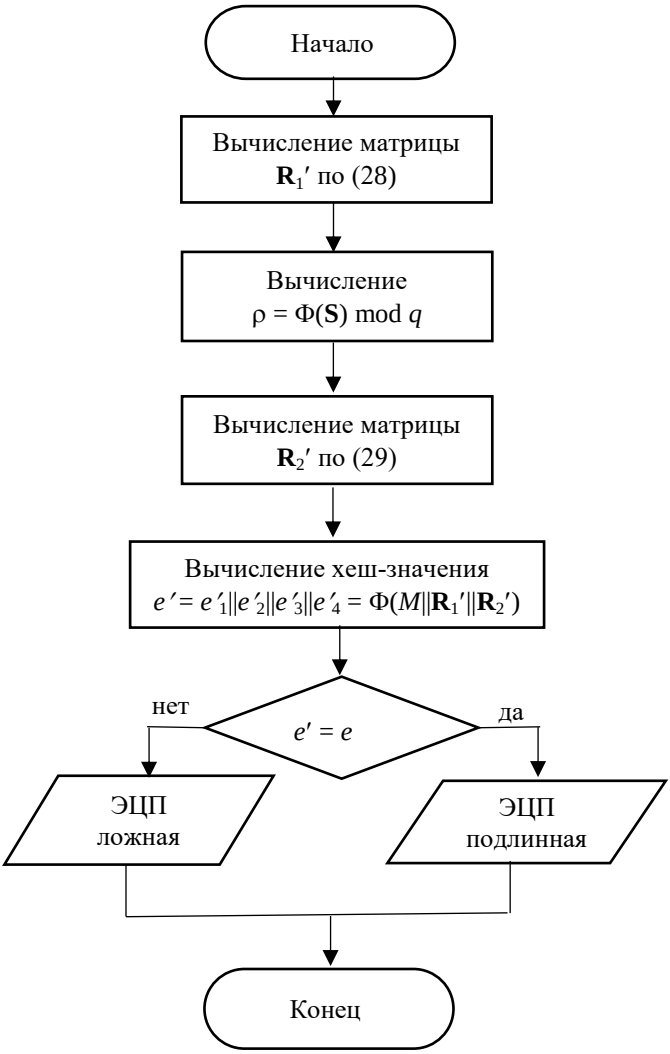


Рис. 2. Верификация ЭЦП

Вычислительная сложность процедуры верификации подписи задается в основном девятью операциями экспоненциирования, осуществляемыми в конечной алгебре квадратных матриц, играющей роль алгебраического носителя алгоритма ЭЦП. Грубые оценки вычислительной сложности процедуры верификации показаны в таблице 3 в приведенных единицах, а именно, в умножения по 16-битному модулю. Последнее позволяет учесть квадратичный рост сложности одной операции возведения в степень с ростом разрядности модуля.

Данные, приведенные в таблицах 2 и 3, показывают, что различные версии разработанного алгоритма вполне применимы на практике. В сопоставлении с известными в литературе постквантовыми алгоритмами-аналогами они представляют больший интерес для приложений, требующих фиксацию открытого ключа и/или цифровой подписи на бумажном носителе. По сравнению с известными постквантовыми алгоритмами ЭЦП разработанный алгоритм имеет минимальный суммарный размер подписи и открытого ключа, тогда как среди первых имеются алгоритмы с меньшим размером подписи при многократно большем размере открытого ключа [20, 21] и алгоритмы с меньшим размером открытого ключа при многократно большем размере подписи [8, 9].

Таблица 3. Вычислительная трудоемкость алгоритмов генерации и верификации ЭЦП для различных версий разработанного алгоритма

μ	Вычислительная сложность, умножений по 16-битному модулю	
	Генерация подписи	Верификация подписи
2	491500	540700
3	276500	304000
5	163200	182400
7	198500	222300
11	237300	266400

Для доказательства корректности работы предложенного постквантового алгоритма ЭЦП покажем, что проверочные матрицы $\mathbf{R}_1'$ и $\mathbf{R}_2'$, вычисляемые на первом и втором шагах процедуры верификации, равны рандомизирующим матрицам $\mathbf{R}_1$ и $\mathbf{R}_2$. Действительно, для подписи (e, ξ, ψ, s, σ, $\mathbf{S}$), вычисленной корректно и в полном соответствии с процедурой генерации ЭЦП, с учетом формул (13)-(18) имеем следующее значение матрицы $\mathbf{R}_1'$:

$$\mathbf{R}_1' = \mathbf{Y}^{e_1\xi}\mathbf{K}^{e_2\psi}\mathbf{T}_1\mathbf{V}^{e_3 s}\mathbf{S}\mathbf{T}_2\mathbf{Z}^{e_2} = (\mathbf{A}\mathbf{G}^{\alpha}A^{-1})^{e_1\xi}\left(\mathbf{A}\mathbf{Q}^{\lambda}\mathbf{A}^{-1}\right)^{e_2\psi}\mathbf{A}\mathbf{Q}^{\beta}\mathbf{G}^{\gamma}\mathbf{D}^{-1} \times$$
$$\times (\mathbf{D}\mathbf{G}\mathbf{D}^{-1})^{e_3 s}(\mathbf{D}\mathbf{G}^{n}\mathbf{J}^{d}\mathbf{F})\mathbf{F}^{-1}\mathbf{J}^{\varepsilon}\mathbf{B}^{-1}(\mathbf{B}\mathbf{J}^{\alpha}\mathbf{B}^{-1})^{e_2} =$$
$$= \mathbf{A}\mathbf{G}^{\alpha e_1\xi}\boldsymbol{Q}^{\lambda e_2\psi+\beta}\mathbf{G}^{\gamma+e_3 s+n}\mathbf{J}^{d+\varepsilon+\alpha e_2}\mathbf{B}^{-1} =$$

(далее с учетом формул (20)-(25) и (13) имеем следующее):

$$= \mathbf{A}\mathbf{G}^{\alpha e_1\frac{k}{\alpha e_1}}\mathbf{Q}^{\lambda e_2\frac{y-\beta-1}{\lambda e_2}+\beta}\mathbf{G}^{x}\mathbf{J}^{z+t}\mathbf{B}^{-1} = \mathbf{A}\mathbf{G}^{k}\boldsymbol{Q}^{y-1}(\mathbf{G}^{x}\mathbf{J}^{z})\mathbf{J}^{t}\mathbf{B}^{-1} =$$
$$= \mathbf{A}\mathbf{G}^{k}Q^{y-1}\mathbf{Q}\mathbf{J}^{t}\mathbf{B}^{-1} = \mathbf{A}\mathbf{G}^{k}\mathbf{Q}^{y}\mathbf{J}^{t}\mathbf{B}^{-1} = \mathbf{R}_1.$$

С учетом выражения (18) для матрицы $\mathbf{R}_2'$ имеем следующее:

$$\mathbf{R}_2' = \mathbf{Z}^{e_1\xi}\mathbf{N}^{e_2\psi}T_3V^{e_4 s}\mathrm{SW}^{\sigma}T_2Z^{\rho} = (\mathbf{B}\mathbf{J}^{\alpha}\mathbf{B}^{-1})^{e_1\xi}\left(\mathbf{B}\mathbf{U}^{\lambda}\mathbf{B}^{-1}\right)^{e_2\psi}\mathbf{B}\mathbf{U}^{\beta}\mathbf{G}^{\tau}\mathbf{D}^{-1} \times$$
$$\times (\mathbf{D}\mathbf{G}\mathbf{D}^{-1})^{e_4 s}(\mathbf{D}\mathbf{G}^{n}\mathbf{J}^{d}\mathbf{F})(\mathbf{F}^{-1}\mathbf{J}\mathbf{F})^{\sigma}\mathbf{F}^{-1}\mathbf{J}^{\varepsilon}\mathbf{B}^{-1}(\mathbf{B}\mathbf{J}^{\alpha}\mathbf{B}^{-1})^{\rho} =$$
$$= \mathbf{B}\mathbf{J}^{\alpha e_1\xi}\mathbf{U}^{\lambda e_2\psi+\beta}\mathbf{G}^{\tau+e_4 s+n}\mathbf{J}^{d+\sigma+\varepsilon+\alpha\rho}\mathbf{B}^{-1} =$$

(далее с учетом формул (20)-(25), (14) и (27) имеем следующее):

$$= \mathbf{B}\mathbf{J}^{k}\mathbf{U}^{y-1}\mathbf{G}^{u}\mathbf{J}^{v+w}\mathbf{B}^{-1} = \mathbf{B}\mathbf{J}^{k}\mathbf{U}^{y-1}(\mathbf{G}^{u}\mathbf{J}^{v})\mathbf{J}^{w}\mathbf{B}^{-1} =$$
$$= \mathbf{B}\mathbf{J}^{k}\mathbf{U}^{y-1}\mathbf{U}\mathbf{J}^{w}\mathbf{B}^{-1} = \mathbf{A}\mathbf{J}^{k}\mathbf{U}^{y}\mathbf{J}^{w}\mathbf{B}^{-1} = \mathbf{R}_2.$$

С учетом полученных равенств $\mathbf{R}_1 = \mathbf{R}_1'$ и $\mathbf{R}_2 = \mathbf{R}_2'$ имеем $e' = \Phi(M\|\mathbf{R}_1'\|\mathbf{R}_2') = \Phi(M\|\mathbf{R}_1\|\mathbf{R}_2) = e$, что в соответствии с шагом 4 процедуры верификации ЭЦП означает то, что корректно вычисленная подпись принимается за подлинную подпись. Последний факт доказывает корректность предложенного в данной статье постквантового алгоритма ЭЦП.

4. Оценка стойкости. Предложенный в данной статье алгоритм относится к недетерминированным криптосхемам. При этом используемый механизм рандомизации подписи создает предпосылки для вычисления некоторых элементов секретного ключа по известным подписям. Атака данного типа состоит в вычислении элементов секретного ключа, которые могут быть вычислены непосредственно по подгоночному элементу подписи $\mathbf{S}$ с использованием выражения (26) и/или по матрицам $\mathbf{R}_1'$ и $\mathbf{R}_2'$, вычисляемым по формулам (28) и (29) с использованием всех элементов ЭЦП и следующих равенств $\mathbf{R}_1' = \mathbf{R}_1 = \mathbf{A}\mathbf{G}^{k}\mathbf{Q}^{y}\mathbf{J}^{t}\mathbf{B}^{-1}$ и $\mathbf{R}_2' = \mathbf{R}_2 = \mathbf{B}\mathbf{J}^{k}\mathbf{U}^{y}\mathbf{J}^{w}\mathbf{B}^{-1}$.

Можно выделить несколько основных вариантов атаки на основе известных подписей в зависимости от того, какие из параметров $\mathbf{S}$, $\mathbf{R}_1$ и $\mathbf{R}_2$ и формулы для их вычисления используются для выполнения такой атаки. Указанные параметры могут использоваться отдельно или

в различных сочетаниях, обусловливая существование семи различных вариантов атаки на основе известных подписей.

Вариант 1. Пусть известна совокупность из η подписей, в которых содержатся подгоночные матрицы $\mathbf{S}_1$, $\mathbf{S}_2$, …, $\mathbf{S}_\eta$. Из формулы (26) видно, что секретные матрицы **D** и **F** используются при вычислении каждой подгоночной матрицы $\mathbf{S}_i$ ($i = 1, 2, \ldots, \eta$), т.е. являются фиксированными неизвестными, а матрицы $(\mathbf{G}^{n_i})$ и $(\mathbf{J}^{d_i})$ – при вычислении только одной матрицы $\mathbf{S}_i$, т.е. являются уникальными неизвестными. Матрицы $(\mathbf{G}^{n_i})$ принадлежат коммутативной группе $\langle\mathbf{G}\rangle$, в качестве представителя которой можно взять матрицу $(\mathbf{G}^{n_1})$. Тогда матрицу $(\mathbf{G}^{n_1})$ можно считать фиксированной неизвестной, вносящей в скалярную БСНУ μ^2 скалярных неизвестных, и выражать каждую из остальных матриц $(\mathbf{G}^{n_i})$ для $i = 2, 3, \ldots, \eta$ через μ уникальных скалярных неизвестных. Аналогичным образом примем матрицу $(\mathbf{J}^{n_1})$ за представителя уникальных матриц $(\mathbf{J}^{d_i})$. Тогда в наборе известных подписей будем иметь четыре фиксированные матричные неизвестные **D**, **F**, $(\mathbf{G}^{n_1})$ и $(\mathbf{J}^{d_i})$, которые совместно задают $4\mu^2$ скалярных неизвестных, и $2(\eta - 1)$ уникальных матричных неизвестных, принадлежащих группам $\langle\mathbf{G}\rangle$ и $\langle\mathbf{J}\rangle$. Каждая из уникальных матричных неизвестных вносит μ скалярных неизвестных, а совокупность уникальных матриц – $2\mu(\eta - 1)$ уникальных скалярных неизвестных. Число скалярных неизвестных обоих типов равно $4\mu^2 + 2\mu(\eta - 1)$. Число уравнений в поле $GF(p)$, задаваемых известными подписями, равно $\eta\mu^2$. Эти уравнения образуют скалярную БСНУ, из которой потенциально можно вычислить секретные матрицы **D**, **F**. Из условия равенства числа уравнений и неизвестных имеем следующее уравнение для вычисления требуемого числа подписей η_0:

$$\eta_0\mu^2 = 4\mu^2 + 2\mu(\eta_0 - 1). \qquad (30)$$

Из уравнения (30) имеем $\eta_0 = (4\mu - 2)/(\mu - 2)$. При $\mu = 2$ имеем бесконечное значение, которое показывает, что для этого случая при любом количестве известных ЭЦП число неизвестных в скалярной БСНУ, возникающей в рамках рассматриваемой атаки, существенно превышает число уравнений. Это означает, что в этом случае решение БСНУ не может существенно снять неопределенность значений элементов секретного ключа **D**, **F**. Оценки значения η_0, числа

уравнений и неизвестных в БСНУ и вычислительной сложности атаки для других значений μ приведены в таблице 4.

Вариант 2. Для каждой i-й ($i = 1, 2, \ldots, \eta$) известной подписи по формуле (29) можно вычислить значение рандомизирующей матрицы $\mathbf{R}_{2(i)}$, которое по формуле (19) задает μ^2 уравнений в $GF(p)$. Рассмотрим матрицы $(\mathbf{J}^{k_1})$ и $(\mathbf{U}^{y_1})$ в качестве представителей групп $<\mathbf{J}>$ и $<\mathbf{U}>$. Это устанавливает три фиксированные матричные неизвестные $\mathbf{B}$, $(\mathbf{J}^{k_1})$, $(\mathbf{U}^{y_1})$, которые совместно задают $3\mu^2$ скалярных неизвестных. Матрица $\mathbf{B}^{-1}$ не вносит новых скалярных неизвестных, ввиду того, что она выражается через элементы матрицы $\mathbf{B}$. Каждая из матриц $(\mathbf{J}^{w_1})$, $(\mathbf{J}^{k_i})$, $(\mathbf{J}^{w_i})$, $(\mathbf{U}^{y_i})$, где $i = 2, 3, \ldots, \eta$, задает μ уникальных скалярных неизвестных, т.е. в целом имеется $\mu + 3\mu(\eta - 1)$ неизвестных такого типа. Для вычисления нужного числа известных подписей имеем такое уравнение:

$$\eta_0\mu^2 = 3\mu^2 + \mu + 3\mu(\eta_0 - 1). \tag{31}$$

Из уравнения (31) имеем $\eta_0 = (3\mu - 2)/(\mu - 3)$. При $\mu = 3$ ($\mu = 2$) имеем бесконечное (отрицательное) значение, которое показывает, что для этих случаев рассматриваемый вариант атаки не может дать ограниченного числа решений ни при каком числе известных подписей. Оценки значения η_0 для второго варианта атаки на основе известных подписей приведены в таблице 4.

Вариант 3. Для каждой i-й ($i = 1, 2, \ldots, \eta$) известной ЭЦП, используя выражение (28), устанавливается значение матрицы $\mathbf{R}_{1(i)}$, которое задает μ^2 уникальных скалярных уравнений, записываемых по формуле (18). Каждая из неизвестных матриц $\mathbf{A}$, $\mathbf{B}^{-1}$, $(\mathbf{G}^{k_1})$, $(\mathbf{Q}^{y_1})$, $(\mathbf{J}^{t_1})$, где $(\mathbf{G}^{k_1})$, $(\mathbf{Q}^{y_1})$ и $(\mathbf{J}^{t_1})$ принимаются за представителей секретных коммутативных групп $<\mathbf{G}>$, $<\mathbf{Q}>$ и $<\mathbf{J}>$ соответственно, задает μ^2 фиксированных скалярных неизвестных, т. е. имеем $5\mu^2$ неизвестных такого типа. Каждая из матриц $(\mathbf{G}^{k_i})$, $(\mathbf{Q}^{y_i})$, $(\mathbf{J}^{t_i})$, где $i = 2, 3, \ldots, \eta$, является уникальной матричной неизвестной, поскольку связана только с одной подгоночной матрицей, а именно, с матрицей $\mathbf{S}_i$. Все уникальные матричные неизвестные вносят $3\mu(\eta - 1)$ уникальных скалярных неизвестных. Для вычисления значения η_0 имеем уравнение:

$$\eta_0\mu^2 = 5\mu^2 + 3\mu(\eta_0 - 1). \tag{32}$$

Из (32) имеем формулу $\eta_0 = (5\mu - 3)/(\mu - 3)$. Число уравнений (T) в БСНУ для различных значений μ представлено в таблице 4,

где Λ – значение стойкости к атаке по вариантам 1–3, оцененное для случая $\eta = \eta_0$ с использованием оценок [28] сложности решения БСНУ по числу неизвестных.

Таблица 4. Значение параметров η_0, T и уровня стойкости Λ к атакам на основе известных подписей

μ	Вариант 1			Вариант 2			Вариант 3		
	η_0	T	Λ	η_0	T	Λ	η_0	T	Λ
2	∞	∞	$>2^{256}$	∞	∞	$>2^{256}$	∞	∞	$>2^{256}$
3	10	90	$>2^{192}$	∞	∞	$>2^{256}$	∞	∞	$>2^{256}$
4	7	112	$>2^{256}$	10	160	$>2^{256}$	17	272	$>2^{256}$
5	6	150	$>2^{256}$	6,5	162	$>2^{256}$	11	275	$>2^{256}$
6	5,5	$\approx$200	$>2^{256}$	5,3	191	$>2^{256}$	9	324	$>2^{256}$
7	5,2	$\approx$255	$>2^{256}$	4,8	235	$>2^{256}$	8	392	$>2^{256}$

Дробные значения для параметра η_0 соответствуют случаям, которые следует трактовать как использование только части скалярных уравнений, задаваемых одной из известных подписей при числе последних, равном наименьшему натуральному значению $\eta \geq \eta_0$. В целом все версии разработанного алгоритма обладают высокой стойкостью к рассмотренным вариантам атаки на основе известных подписей. Заметим, что знак ∞ для числа уравнений и неизвестных в БСНУ соответствует случаям, когда ни при каком значении числа известных подписей нельзя добиться ситуации, в которой число уравнений превышает или равно числу неизвестных.

Можно предложить и другие варианты атаки на основе известных ЭЦП, использующие формирование БСНУ:

- по параметрам **S** и $\mathbf{R}_1$;
- по параметрам **S** и $\mathbf{R}_2$;
- по параметрам $\mathbf{R}_1$ и $\mathbf{R}_2$;
- по параметрам **S**, $\mathbf{R}_1$ и $\mathbf{R}_2$.

Детальное рассмотрение этих вариантов показывает, что для каждого из них версии разработанного алгоритма, соответствующие значениям $\mu = 2, 3, \ldots, 7$, обладают стойкостью $>2^{256}$.

В рамках лобовой атаки на разработанный алгоритм требуется решить БСНУ, задаваемую выражениями (13)-(17). Решение системы степенных уравнений в алгебре матриц связано с переходом к скалярной БСНУ, в которой степенные уравнения заданы в конечном поле $GF(p)$. При таком переходе число уравнений возрастает в μ^2 раз.

В системе матричных уравнений неизвестными являются элементы секретного ключа, т.е. девять неизвестных матриц

(**A**, **B**, **D**, **F**, **G**, **J**, **Q**, **U**) и десять неизвестных натуральных значений (x, z, u, v, λ, β, α, γ, τ, ε), входящих в матричные уравнения в качестве степеней, причем разрядность каждого из них равна или превышает 128 бит. Большая разрядность указанных степеней обусловливает практическую невозможность прямолинейного сведения системы матричных уравнений к скалярной БСНУ. Это сведение можно осуществить, вводя следующие дополнительные матричные неизвестные:

− $(\mathbf{G}^x)$, $(\mathbf{G}^u)$, $(\mathbf{G}^\alpha)$, $(\mathbf{G}^\gamma)$ и $(\mathbf{G}^\tau)$, каждая из которых при сведении к скалярной БСНУ задаст μ скалярных неизвестных при их представлении через элементы представителя **G** коммутативной секретной группы <**G**>, т. е. в целом перечисленные матрицы вносят 5μ неизвестных;

− $(\mathbf{J}^z)$, $(\mathbf{J}^v)$, $(\mathbf{J}^\alpha)$ и $(\mathbf{J}^\varepsilon)$, которые зададут 4μ неизвестных в скалярной БСНУ при их представлении через элементы представителя **J** коммутативной секретной группы <**J**>;

− $(\mathbf{Q}^\lambda)$ и $(\mathbf{Q}^\beta)$, которые зададут 2μ неизвестных в скалярной БСНУ при их представлении через элементы представителя **Q** коммутативной секретной группы <**Q**>;

− $(\mathbf{U}^\lambda)$ и $(\mathbf{U}^\beta)$, которые зададут еще 2μ неизвестных в скалярной БСНУ при их представлении через элементы представителя **U** коммутативной секретной группы <**U**>.

Каждая из матриц (**A**, **B**, **D**, **F**, **G**, **J**, **Q**, **U**) вносит μ^2 уникальных скалярных неизвестных, а все вместе задают $8\mu^2$ неизвестных в скалярной БСНУ. С учетом того, что каждая из 13 дополнительных матричных неизвестных определяет μ скалярных неизвестных, имеем следующую формулу для полного числа η_1 скалярных неизвестных:

$$\eta_1 = 8\mu^2 + 13\mu. \tag{33}$$

Выражения (13)-(17) задают 11 матричных уравнений, которые сводятся к $\eta_2 = 11\mu^2$ уравнениям в поле $GF(p)$. В таблице 5 представлены значения числа неизвестных η_1 и уравнений η_2 в скалярной БСНУ, трудность решения которой определяет уровень стойкости Λ' различных версий разработанного алгоритма к прямой атаке.

При $\mu = 2$ и $\mu = 3$ мы имеем недоопределенные БСНУ, при $\mu = 4$ и $\mu = 5$ – примерно сбалансированные, а при $\mu \geq 6$ – переопределенные. Критерием отнесения БСНУ к типу примерно сбалансированных может служить условие $2|\eta_1 - \eta_2| < \mu^2$, использованное при заполнении второй колонки таблице 5. В целом

версии разработанного алгоритма, соответствующие значениям $\mu \geq 4$ обладают более высоким уровнем защищенности к атакам на основе использования эквивалентных секретных ключей по сравнении с алгоритмами [24, 25], для которых имеет место неравенство $\eta_1 - \eta_2 > \mu^2$ и большое число различных решений, соответствующих большому числу эквивалентных секретных ключей. Нахождение конкретных значений последних не проще, чем решение БСНУ с числом уравнений η_2. Однако в общем случае следует сделать предположение о существовании атак, использующих сам факт существования эквивалентных ключей без конкретного их вычисления, которые аналогичны атаке, описанной в работе [27].

Таблица 5. Число неизвестных η_1 и уравнений η_2 в БСНУ, сложность решения которой определяет уровень стойкости Λ' к прямой атаке

μ	Тип БСНУ	η_1	η_2	Λ'
2	недоопределенная	58	44	$>2^{128}$
3	недоопределенная	111	99	$>2^{256}$
4	≈ сбалансированная	180	176	$>2^{256}$
5	≈ сбалансированная	265	275	$>2^{256}$
6	переопределенная	366	396	$>2^{256}$
7	переопределенная	483	539	$>2^{256}$

Для алгоритмов, стойкость которых определяется трудностью решения сбалансированных или переопределенных БСНУ, указанное предположение теряет актуальность, что в данной статье трактуется как повышение защищенности к потенциальным атакам на основе эквивалентных ключей. Вопрос полного устранения атак последнего типа связан с выполнением формального доказательства вычислительной невозможности модифицировать формулы (13)-(17) таким образом, что будет получена модифицированная схема ЭЦП, сохраняющая проверочные уравнения (28) и (29). Выполнение такого доказательства является самостоятельной задачей и связано с принципиальными трудностями, аналогичными тем, которые возникают при попытках доказательства того факта, что некоторый найденный алгоритм взлома криптоалгоритма является наилучшим (обладающий минимальной вычислительной сложностью) из всех потенциально возможных.

Актуальным для разработанного алгоритма является вопрос об оценке сложности подделки подписи путем использования элемента подписи **S** в качестве подгоночного параметра атаки. Действительно, в отличие от известных алгоритмов ЭЦП такого типа, например, [24, 25]

в предложенном алгоритме используются два проверочных уравнения, в которые матрица **S** входит как множитель первой степени, поэтому совместное решение указанных матричных уравнений при прочих фиксированных параметрах сводится к решению системы из $2\mu^2$ скалярных уравнений первой степени, что имеет полиномиальную по времени вычислительную сложность. Способы подделки подписи на основе решения системы линейных уравнений практически устраняются тем, что имеют место следующие два момента.

Во-первых, в разработанном алгоритме одним из параметров второго проверочного уравнения является 128-битная степень ρ, которая вычисляется по значению матрицы **S**, т. е. нет реализуемой возможности зафиксировать значение ρ, что позволило бы прийти к линейной системе скалярных уравнений. Рассмотрение степени ρ в качестве неизвестной обусловливает практическую невыполнимость сведения матричных уравнений к системе скалярных уравнений (линейных или нелинейных).

Во-вторых, фиксирование элементов подписи ξ, ψ, s, σ и значений матриц $\mathbf{R}_1$ и $\mathbf{R}_2$, с последующим вычислением хеш-значения $e = e_1||e_2||e_3||e_4 = \Phi(M||\mathbf{R}_1||\mathbf{R}_2)$ и фиксированием значения ρ, позволяет свести задачу решения системы из двух матричных уравнений (28) и (29), относительно неизвестной матрицы **S,** к решению системы скалярных линейных уравнений, однако вероятность того, что последняя система совместной, т.е. будет иметь решения, является пренебрежимо малой.

Представленные варианты атак на разработанный алгоритм имеют высокую вычислительную сложность, однако в дальнейшем следует уделить внимание разработке других атак и оценкам стойкости к ним. На настоящий момент представляется, что предложенный алгоритм имеет перспективы как кандидат на практичную постквантовую схему подписи, а полученные результаты дают основание для привлечения более широкого круга исследователей к анализу стойкости предложенного алгоритма ЭЦП.

Разработанный в данной работе постквантовый ЭЦП алгоритм представляет существенный практический интерес, однако на данный момент этот алгоритм и общий подход к построению алгебраических алгоритмов на основе трудности решения БСНУ не прошли достаточно длительной апробации с участием широкого круга криптографов. Авторы предполагают, что данная статья послужит побудительным мотивом для вовлечения читателя в процесс указанной апробации.

5. Заключение. Предложен новый постквантовый алгоритм ЭЦП на конечных алгебрах матриц размера $\mu\times\mu$. Все версии разработанного

алгоритма, соответствующие значениям параметра $\mu = 2, 3, 4, 5, 6, 7$ и 11, обладают высоким уровнем стойкости к лобовой атаке и к атаке на основе известных подписей. За счет обеспечения сбалансированности или переопределенности БСНУ для версий алгоритма, соответствующих значениям $\mu = 4, 5, 6, 7, 11$, повышается уровень защищенности к потенциальным атакам на основе эквивалентных ключей. Благодаря сравнительно малой длине подписи и открытого ключа предложенный алгоритм представляет интерес как прототип постквантового стандарта ЭЦП, ориентированного на широкое практическое применение.

Важной задачей дальнейшего исследования предложенного постквантового алгоритма ЭЦП является разработка специальных атак, учитывающих особенности его построения, и оценивание его стойкости к таким атакам.

С целью увеличения производительности и уменьшения схемотехнической сложности аппаратной реализации представляет практический интерес рассмотрение модификации разработанной схемы ЭЦП, использующей в качестве своего носителя конечные алгебры квадратных матриц, заданных над конечными полями характеристики два.

Литература

1. Post-Quantum Cryptography. 16-th International Conference, PQCrypto 2025, Taipei, Taiwan, April 8–10, 2025. Proceedings, Part I // Lecture Notes in Computer Science. Springer. 2023. vol. 15577. DOI: 10.1007/978-3-031-86599-2.
2. Post-Quantum Cryptography. 16-th International Conference, PQCrypto 2025, Taipei, Taiwan, April 8–10, 2025. Proceedings, Part II // Lecture Notes in Computer Science. Springer. 2023. vol. 15578. DOI: 10.1007/978-3-031-86602-9.
3. Post-Quantum Cryptography. 15-th International Conference, PQCrypto 2024, Oxford, UK, June 12–14, 2024, Proceedings // Lecture Notes in Computer Science. Springer. 2024. vol. 14771–14772. DOI: 10.1007/978-3-031-62743-9.
4. Yan S.Y. Quantum Computational Number Theory // Cham: Springer, 2015. 252 p. DOI: 10.1007/978-3-319-25823-2.
5. Yan S.Y. Quantum Attacks on Public-Key Cryptosystems // New York: Springer, 2013. 207 p. DOI: 10.1007/978-1-4419-7722-9.
6. Codes, Cryptology and Information Security. 4th International Conference, C2SI 2023, Rabat, Morocco, May 29–31, 2023 // Lecture Notes in Computer Science. Springer. 2023. vol. 13874. 265 p. DOI: 10.1007/978-3-031-33017-9.
7. Baldi M., Battaglioni M., Chiaraluce F., et al. A new path to code-based signatures via identification schemes with restricted errors // Advances in Mathematics of Communications. 2025. vol. 19. no. 5. pp. 1360□1381. DOI: 10.3934/amc.2024058.
8. Vysotskaya V.V., Chizhov I.V. The security of the code-based signature scheme based on the Stern identification protocol // Прикладная дискретная математика. 2022. №57. С. 67–90. DOI: 10.17223/20710410/57/5.
9. Ниткин И.С. Применение метода компактного описания подстановки для модификации схемы цифровой подписи на основе протокола аутентификации

Штерна // Информационно-управляющие системы. 2025. №6. С. 51–63. DOI: 10.31799/1684-8853-2025-6-51-63.
10. Bidoux L., Gaborit P., Kulkarni M., et al. Code-based signatures from new proofs of knowledge for the syndrome decoding problem // Designs, Codes and Cryptography. 2023. vol. 91. pp. 497–544. DOI: 10.1007/s10623-022-01114-3.
11. Baldi M., Bitzer S., Pavoni A., Santini P., Wachter-Zeh A., Weger V. Zero Knowledge Protocols and Signatures from the Restricted Syndrome Decoding Problem / In: Tang, Q., Teague, V. (eds) / Public-Key Cryptography – PKC 2024. PKC 2024 // Lecture Notes in Computer Science. Springer. 2024. vol. 14602. pp. 243–274. DOI: 0.1007/978-3-031-57722-2_8.
12. Feneuil T., Joux A., Rivain M. Shared permutation for syndrome decoding: new zero-knowledge protocol and code-based signature // Designs, Codes and Cryptography. 2023. vol. 91. pp. 563–608. DOI: 10.1007/s10623-022-01116-1.
13. Feneuil T., Joux A., Rivain M. Syndrome Decoding in the Head: Shorter Signatures from Zero-Knowledge Proofs // Lecture Notes in Computer Science. Springer. 2022. vol. 13508. pp. 541–572. DOI: 10.1007/978-3-031-15979-4_19.
14. Battarbee C., Kahrobaei D., Perret L., Shahandashti S.F. SPDH-Sign: Towards Efficient, Post-quantum Group-Based Signatures / In: Johansson T., Smith-Tone D. (eds) / Post-Quantum Cryptography. PQCrypto 2023 // Lecture Notes in Computer Science. Springer. 2023. vol. 14154. pp. 113–138. DOI: 10.1007/978-3-031-40003-2_5.
15. Agibalov G.P. ElGamal cryptosystems on Boolean functions // Прикладная дискретная математика. 2018. №42. С. 57–65. DOI: 10.17223/20710410/42/4.
16. Gartner J. NTWE: A Natural Combination of NTRU and LWE / In: Johansson T., Smith-Tone D. (eds) / Post-Quantum Cryptography. PQCrypto 2023 // Lecture Notes in Computer Science. Springer. 2023. vol. 14154. pp. 321–353. DOI: 10.1007/978-3-031-40003-2_12.
17. Lysakov I.V. Solving some cryptanalytic problems for lattice-based cryptosystems with quantum annealing method // Математические вопросы криптографии. 2023. Т. 14. №2. С. 111–122. DOI: 10.4213/mvk441.
18. Hamlin B., Song F. Quantum Security of Hash Functions and Property-Preservation of Iterated Hashing / In: Ding, J., Steinwandt, R. (eds) / Post-Quantum Cryptography. PQCrypto 2019 // Lecture Notes in Computer Science. Springer. 2019. vol. 11505. pp. 329–349. DOI: 10.1007/978-3-030-25510-7_18.
19. Ikematsu Y., Nakamura S., Takagi T. Recent progress in the security evaluation of multivariate public-key cryptography // IET Information Security. 2022. vol. 17. no. 2. pp. 210–226. DOI: 10.1049/ise2.12092.
20. Ding J., Petzoldt A., Schmidt D.S. Multivariate Cryptography / In: Multivariate Public Key Cryptosystems // Advances in Information Security. Springer. 2020. vol. 80. pp. 7–23. DOI: 10.1007/978-1-0716-0987-3_2.
21. Ding J., Petzoldt A., Schmidt D.S. Oil and Vinegar / In: Multivariate Public Key Cryptosystems // Advances in Information Security. Springer. 2020. vol. 80. pp. 89–150. DOI: 10.1007/978-1-0716-0987-3_5.
22. 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.
23. 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.
24. 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.

25. Молдовян А.А., Молдовян Д.Н., Костина А.А. Рандомизация в постквантовых алгоритмах ЭЦП с секретной группой // Информатика и автоматизация. 2025. Т. 24. №6. С. 1810–1835. DOI: 10.15622/ia.24.6.9.
26. Dinh K.L., Do T.B., Moldovyan N.A., Petrenko A.S. Typical Algebraic Signature Schemes with Two Hidden Groups / In: Dang, T.K., Kung, J., Chung, T.M. (eds) / Future Data and Security Engineering. FDSE 2025 // Communications in Computer and Information Science. Springer. 2026. vol 2709. pp. 83–99. DOI: 10.1007/978-981-95-4724-1_7.
27. Ma Y. Cryptanalysis of the cryptosystems based on the generalized hidden discrete logarithm problem // Computer Science Journal of Moldova. 2024. vol. 32. no. 2(95). pp. 289–3070. DOI: 10.56415/csjm.v.32.15.
28. Ding J., Petzoldt A. Current State of Multivariate Cryptography // IEEE Security and Privacy Magazine. 2017. vol. 15. no. 4. pp. 28–36. DOI: 10.1109/MSP.2017.3151328.

Молдовян Александр Андреевич — д-р тех. наук, профессор, главный научный сотрудник, лаборатория проблем компьютерной безопасности, СПб ФИЦ РАН. Область научных интересов: криптография, постквантовые криптоалгоритмы с открытым ключом, электронная цифровая подпись, криптографические протоколы, компьютерная безопасность. Число научных публикаций — 230. maa1305@yandex.ru; 14-я линия В.О., д. 39, г. Санкт-Петербург, 199178, РФ; р.т. +7(812)328-7181, fax +7(812)328-4450.

Молдовян Николай Андреевич — д-р тех. наук, профессор, главный научный сотрудник, лаборатория проблем компьютерной безопасности, СПб ФИЦ РАН. Область научных интересов: криптография, постквантовые алгоритмы цифровой подписи, псевдовероятностные шифры, компьютерная безопасность. Число научных публикаций — 250. nmold@mail.ru; 14-я линия В.О., д. 39, г. Санкт-Петербург, 199178, РФ; р.т. +7(812)328-7181, fax +7(812)328-4450.

Молдовян Дмитрий Николаевич — канд. тех. наук, доцент кафедры информационных систем Санкт-Петербургского государственного электротехнического университета «ЛЭТИ». Область научных интересов: криптография, постквантовые двухключевые криптоалгоритмы на алгебрах и нелинейных отображениях, цифровая подпись, информационная безопасность. Число научных публикаций — 120. mdn.spectr@mail.ru; ул. Профессора Попова, д. 5, г. Санкт-Петербург, 197022, РФ; р.т. +7(812)234-2772.

Костина Анна Александровна — научный сотрудник, лаборатория проблем компьютерной безопасности, 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., д. 39, г. Санкт-Петербург, 199178, РФ; р.т. +7(812)328-7181, fax +7(812)328-4450.

Поддержка исследований. Работа выполнена при финансовой поддержке РНФ: проект № 24-41-04006.

DOI 10.15622/ia.25.3.11

A. MOLDOVYAN, N. MOLDOVYAN, D. MOLDOVYAN, A. KOSTINA

POST-QUANTUM DIGITAL SIGNATURE ALGORITHM BASED ON THE DIFFICULTY OF SOLVING POWER EQUATIONS

Moldovyan A., Moldovyan N., Moldovyan D., Kostina A. **Post-Quantum Digital Signature Algorithm Based on the Difficulty of Solving Power Equations.**

Abstract. The computational complexity of finding solutions to large systems of nonlinear equations (LSNE) underlies a number of post-quantum public-key cryptosystems, including algebraic digital signature algorithms using matrices as secret key elements. Such algorithms belong to probabilistic cryptosystems and are characterized by the use of a certain matrix S as a fitting element of the signature. The latter makes it relevant to consider resistance to attacks using known signatures. In this case, a direct attack is based on the LSNE solution. In well-known algebraic digital signature algorithms, underdetermined LSNE with a large number of solutions arise in the framework of a direct attack, which creates the preconditions for attacks based on equivalent keys. The article considers the construction of a post-quantum algorithm on finite algebras of square matrices, which, when directly attacked, leads to close-to-balanced (with an equal number of equations and unknowns) and overdetermined LSNE (with a number of equations exceeding the number of unknowns). A special feature of the proposed algorithm is the use of auxiliary hidden groups, an auxiliary parameter for the randomization of the digital signature, which is calculated as the value of a compressive one-way function from the value of S. The article provides an assessment of the resistance to direct attacks and several variants of attacks based on known signatures.

Keywords: post-quantum cryptography, digital signature algorithm, finite matrix algebra, system of power equations, secret group.

References

1. Post-Quantum Cryptography. 16-th International Conference, PQCrypto 2025, Taipei, Taiwan, April 8–10, 2025. Proceedings, Part I. Lecture Notes in Computer Science. Springer. 2023. vol. 15577. DOI: 10.1007/978-3-031-86599-2.
2. Post-Quantum Cryptography. 16-th International Conference, PQCrypto 2025, Taipei, Taiwan, April 8–10, 2025. Proceedings, Part II. Lecture Notes in Computer Science. Springer. 2023. vol. 15578. DOI: 10.1007/978-3-031-86602-9.
3. Post-Quantum Cryptography. 15-th International Conference, PQCrypto 2024, Oxford, UK, June 12–14, 2024, Proceedings. Lecture Notes in Computer Science. Springer. 2024. vol. 14771–14772. DOI: 10.1007/978-3-031-62743-9.
4. Yan S.Y. Quantum Computational Number Theory. Cham: Springer, 2015. 252 p. DOI: 10.1007/978-3-319-25823-2.
5. Yan S.Y. Quantum Attacks on Public-Key Cryptosystems. New York: Springer, 2013. 207 p. DOI: 10.1007/978-1-4419-7722-9.
6. Codes, Cryptology and Information Security. 4th International Conference, C2SI 2023, Rabat, Morocco, May 29–31, 2023. Lecture Notes in Computer Science. Springer. 2023. vol. 13874. 265 p. DOI: 10.1007/978-3-031-33017-9.
7. Baldi M., Battaglioni M., Chiaraluce F., et al. A new path to code-based signatures via identification schemes with restricted errors. Advances in Mathematics of Communications. 2025. vol. 19. no. 5. pp. 1360□1381. DOI: 10.3934/amc.2024058.
8. Vysotskaya V.V., Chizhov I.V. [The security of the code-based signature scheme based on the Stern identification protocol]. Prikladnaya diskretnaya matematika –

Applied discrete Mathematics. 2022. no. 57. pp. 67–90. DOI: 10.17223/20710410/57/5. (In Russ.).
9. Nitkin I.S. [Application of the method of compact substitution description for modifying the digital signature scheme based on the Stern authentication protocol]. Informatsionno-upravlyayushchiye sistemy – Information and Control Systems. 2025. no. 6. pp. 51–63. DOI: 10.31799/1684-8853-2025-6-51-63. (In Russ.).
10. Bidoux L., Gaborit P., Kulkarni M., et al. Code-based signatures from new proofs of knowledge for the syndrome decoding problem. Designs, Codes and Cryptography. 2023. vol. 91. pp. 497–544. DOI: 10.1007/s10623-022-01114-3.
11. Baldi M., Bitzer S., Pavoni A., Santini P., Wachter-Zeh A., Weger V. Zero Knowledge Protocols and Signatures from the Restricted Syndrome Decoding Problem. In: Tang, Q., Teague, V. (eds). Public-Key Cryptography – PKC 2024. PKC 2024. Lecture Notes in Computer Science. Springer. 2024. vol. 14602. pp. 243–274. DOI: 0.1007/978-3-031-57722-2_8.
12. Feneuil T., Joux A., Rivain M. Shared permutation for syndrome decoding: new zero-knowledge protocol and code-based signature. Designs, Codes and Cryptography. 2023. vol. 91. pp. 563–608. DOI: 10.1007/s10623-022-01116-1.
13. Feneuil T., Joux A., Rivain M. Syndrome Decoding in the Head: Shorter Signatures from Zero-Knowledge Proofs. Lecture Notes in Computer Science. Springer. 2022. vol. 13508. pp. 541–572. DOI: 10.1007/978-3-031-15979-4_19.
14. Battarbee C., Kahrobaei D., Perret L., Shahandashti S.F. SPDH-Sign: Towards Efficient, Post-quantum Group-Based Signatures. In: Johansson T., Smith-Tone D. (eds). Post-Quantum Cryptography. PQCrypto 2023. Lecture Notes in Computer Science. Springer. 2023. vol. 14154. pp. 113–138. DOI: 10.1007/978-3-031-40003-2_5.
15. Agibalov G.P. [ElGamal cryptosystems on Boolean functions]. Prikladnaya diskretnaya matematika – Applied discrete Mathematics. 2018. no. 42. pp. 57–65. DOI: 10.17223/20710410/42/4. (In Russ.).
16. Gartner J. NTWE: A Natural Combination of NTRU and LWE. In: Johansson T., Smith-Tone D. (eds). Post-Quantum Cryptography. PQCrypto 2023. Lecture Notes in Computer Science. Springer. 2023. vol. 14154. pp. 321–353. DOI: 10.1007/978-3-031-40003-2_12.
17. Lysakov I.V. [Solving some cryptanalytic problems for lattice-based cryptosystems with quantum annealing method]. Matematicheskiye voprosy kriptografii – Mathematical issues of Cryptography. 2023. vol. 14. no. 2. pp. 111–122. DOI: 10.4213/mvk441. (In Russ.).
18. Hamlin B., Song F. Quantum Security of Hash Functions and Property-Preservation of Iterated Hashing. In: Ding, J., Steinwandt, R. (eds). Post-Quantum Cryptography. PQCrypto 2019. Lecture Notes in Computer Science. Springer. 2019. vol. 11505. pp. 329–349. DOI: 10.1007/978-3-030-25510-7_18.
19. Ikematsu Y., Nakamura S., Takagi T. Recent progress in the security evaluation of multivariate public-key cryptography. IET Information Security. 2022. vol. 17. no. 2. pp. 210–226. DOI: 10.1049/ise2.12092.
20. Ding J., Petzoldt A., Schmidt D.S. Multivariate Cryptography. In: Multivariate Public Key Cryptosystems. Advances in Information Security. Springer. 2020. vol. 80. pp. 7–23. DOI: 10.1007/978-1-0716-0987-3_2.
21. Ding J., Petzoldt A., Schmidt D.S. Oil and Vinegar. In: Multivariate Public Key Cryptosystems. Advances in Information Security. Springer. 2020. vol. 80. pp. 89–150. DOI: 10.1007/978-1-0716-0987-3_5.
22. 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.

23. 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.
24. Moldovyany A.A., Moldovyany D.N., Moldovyany N.A. [Post-Quantum Public-Key Cryptoschemes on Finite Algebras]. Informatika i avtomatizatsiya – Informatics and Automation. 2024. vol. 23. no.4. pp. 1246–1276. DOI: 10.15622/ia.23.4.12. (In Russ.).
25. Moldovyany A.A., Moldovyany D.N., Kostina A.A. [Randomization in Post-Quantum Digital Signature Algorithms with a Secret Group]. Informatika i avtomatizatsiya – Informatics and Automation. 2025. vol. 24. no. 6. pp. 1810–1835. DOI: 10.15622/ia.24.6.9. (In Russ.).
26. Dinh K.L., Do T.B., Moldovyan N.A., Petrenko A.S. Typical Algebraic Signature Schemes with Two Hidden Groups. In: Dang, T.K., Kung, J., Chung, T.M. (eds). Future Data and Security Engineering. FDSE 2025. Communications in Computer and Information Science. Springer. 2026. vol 2709. pp. 83–99. DOI: 10.1007/978-981-95-4724-1_7.
27. Ma Y. Cryptanalysis of the cryptosystems based on the generalized hidden discrete logarithm problem. Computer Science Journal of Moldova. 2024. vol. 32. no. 2(95). pp. 289–3070. DOI: 10.56415/csjm.v.32.15.
28. Ding J., Petzoldt A. Current State of Multivariate Cryptography. IEEE Security and Privacy Magazine. 2017. vol. 15. no. 4. pp. 28–36. DOI: 10.1109/MSP.2017.3151328.

Moldovyan Alexandr — 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 cryptography, electronic digital signature, cryptographic protocols, computer security. The number of publications — 230. maa1305@yandex.ru; 39, 14th Line V. O., 199178, Saint-Petersburg, Russia; office phone: +7(921)953-0373.

Moldovyan Nikolay — 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 digital signature algorithms, pseudo-probabilistic ciphers, computer security. The number of publications — 250. nmold@mail.ru; 39, 14th Line V. O., 199178, Saint-Petersburg, Russia; office phone: +7(812)328-7181.

Moldovyan Dmitry — Ph.D., Associate professor, Department of Information Systems, Federal State Autonomous Educational Institution of Higher Education St. Petersburg State Electrotechnical University "LETI" named after V.I. Ulyanov (Lenin) (SPbGETU "LETI"). Research interests: cryptography, post-quantum two-key cryptographic algorithms on algebras and nonlinear mappings, digital signature, information security. The number of publications — 120. mdn.spectr@mail.ru; 5, Professora Popova St., 197022, Saint-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 cryptosystems, electronic digital signatures, cryptographic protocols, and information security. The number of publications — 40. to.ann@inbox.ru; 39, 14th Line V. O., 199178, Saint-Petersburg, Russia; office phone: +7(812)328-7181.

Acknowledgements. This research is supported by RSF (project #24-41-04006).

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 материала выделение подразделов. В основную часть допускается помещать рисунки, таблицы, листинги и формулы. Правила их оформления подробно изложены на нашем сайте в разделе «Руководство для авторов».

Author guidelines

Interaction between authors and the Editorial board is carried out via the personal account on the website of the journal "Informatics and Automation" http://ia.spcras.ru/. Upon registration, authors are requested to fill in all the provided data fields in the registration form. Manuscripts should be prepared using MS Word 2007 or later, or LaTex. The length of the main text (up to the References section) must be between 20 and 30 pages, inclusive. Hyphenation is not allowed. Page numbers should not be inserted. The main body of the manuscript should be divided into sections. The following sections are mandatory: an Introduction, at least one "substantive" section, and a Conclusion. The use of subsections is permitted if justified by the content and structure of the material. Figures, tables, listings, and formulas may be included in the main body. Detailed rules for their formatting are available on our website in the "Author Guidelines" section.

Подписано к печати 05.06.2026.
Формат 60×90 1/16. Усл. печ. л. 22,3. Заказ № 176. Тираж 300 экз.
Дата выхода в свет: 05.06.2026.

Отпечатано в Редакционно-издательском центре ГУАП.
Адрес типографии: Россия, 190000, г. Санкт-Петербург, ул. Большая Морская, д. 67, лит. А
Подписной индекс П5513 по каталогу «Почта России»

Цена свободная

Signed to print 05.06.2026
Format 60×90 1/16. Cond. printed sheets 22.3.
Order No. 176. Circulation 300 copies.
Passed for print 05.06.2026.

Printed in Publishing center GUAP.
Address: 67 Bolshaya Morskaya Street., St.Petersburg, 190000, Russia
Subscription Index П5513, Russian Post Catalog

Price not fixed