

ISSN 2518-1092

# НАУЧНЫЙ РЕЗУЛЬТАТ

ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ

RESEARCH RESULT. INFORMATION TECHNOLOGY

**10(3) 2025**

**16+**

Сайт журнала:  
[rinformation.ru](http://rinformation.ru)  
сетевой научный рецензируемый журнал  
online scholarly peer-reviewed journal



Журнал зарегистрирован в Федеральной службе по надзору в сфере связи, информационных технологий и массовых коммуникаций (Роскомнадзор)  
Свидетельство о регистрации средства массовой информации Эл. № ФС77-69101 от 14 марта 2017 г.  
Журнал включен в Перечень ВАК рецензируемых научных изданий, в которых должны быть опубликованы основные научные результаты диссертаций на соискание ученых степеней кандидата и доктора наук (1.2.1. Искусственный интеллект и машинное обучение, 2.3.1. Системный анализ, управление и обработка информации, статистика).  
Журнал зарегистрирован в Российском индексе научного цитирования (РИНЦ).  
The journal has been registered at the Federal service for supervision of communications information technology and mass media (Roskomnadzor)  
Mass media registration certificate El. № FS 77-69101 of March 14, 2017  
The journal is included into the List of Higher Attestation Commission of peer-reviewed scientific publications where the main scientific results of dissertations for obtaining scientific degrees of a candidate and doctor of science should be  
(1.2.1. Artificial Intelligence and machine learning, 2.3.1. The system analysis, management and information processing, statistics).  
The journal is introduced in Russian Science Citation Index (RSCI).



Том 10, № 3. 2025

СЕТЕВОЙ НАУЧНО-ПРАКТИЧЕСКИЙ ЖУРНАЛ

Издается с 2016 г.

ISSN 2518-1092



Volume 10, № 3. 2025

ONLINESCHOLARLYPEER-REVIEWEDJOURNAL

First published online: 2016

ISSN 2518-1092

#### РЕДАКЦИОННАЯ КОЛЛЕГИЯ:

**ГЛАВНЫЙ РЕДАКТОР: Черноморец А.А.**, доктор технических наук, доцент, профессор кафедры прикладной информатики и информационных технологий Белгородского государственного национального исследовательского университета.

**ЗАМЕСТИТЕЛЬ ГЛАВНОГО РЕДАКТОРА: Жихарев А.Г.**, доктор технических наук, доцент, доцент кафедры информационных и робототехнических систем Белгородского государственного национального исследовательского университета.

**ОТВЕТСТВЕННЫЙ СЕКРЕТАРЬ: Болгова Е.В.**, кандидат технических наук, доцент кафедры прикладной информатики и информационных технологий Белгородского государственного национального исследовательского университета.

**РЕДАКТОР АНГЛИЙСКИХ ТЕКСТОВ: Ляшенко И.В.**, кандидат филологических наук, доцент, доцент кафедры английской филологии и межкультурной коммуникации Белгородского государственного национального исследовательского университета

#### ЧЛЕНЫ РЕДАКЦИОННОЙ КОЛЛЕГИИ:

**Басов О.О.**, доктор технических наук, доцент, руководитель центра искусственного интеллекта (Акционерное общество «АСТ»), г. Санкт-Петербург)

**Белов С.П.**, доктор технических наук, профессор (Белгородский государственный национальный исследовательский университет, г. Белгород)

**Волчков В.П.**, доктор технических наук, профессор (Московский технический университет связи и информатики, г. Москва)

**Дмитриенко В.Д.**, доктор технических наук, профессор (Харьковский национальный технический университет «ХПИ», г. Харьков, Украина)

**Иващук О.А.**, доктор технических наук, профессор (Белгородский государственный национальный исследовательский университет, г. Белгород)

**Калмыков И.А.**, доктор технических наук, профессор (Северо-Кавказский федеральный университет, г. Ставрополь)

**Корсунов Н.И.**, заслуженный деятель науки РФ, доктор технических наук, профессор (Белгородский государственный национальный исследовательский университет, г. Белгород)

**Коськин А.В.**, доктор технических наук, профессор (Орловский государственный университет им. И. С. Тургенева, г. Орел)

**Ломазов В.А.**, доктор физико-математических наук, профессор (Белгородский государственный аграрный университет им. В.Я. Горина, г. Белгород)

**Маторин С.И.**, доктор технических наук, профессор (Белгородский государственный национальный исследовательский университет, г. Белгород)

**Орлова Ю.А.**, доктор технических наук, доцент (Волгоградский государственный технический университет, г. Волгоград)

**Таранчук В.Б.**, доктор физико-математических наук, профессор, (Белорусский государственный университет, г. Минск, Республика Беларусь)

#### EDITORIAL TEAM:

**EDITOR-IN-CHIEF: Andrej A. Chernomorets**, Doctor of Technical Sciences, Associate Professor, Professor, Belgorod State National Research University

**DEPUTY EDITOR-IN-CHIEF: Alexander G. Zhikharev**, Doctor of Technical Sciences, Associate Professor, Belgorod State National Research University

**EXECUTIVE SECRETARY: Evgeniya V. Bolgova**, Candidate of Technical Sciences, Associate Professor, Belgorod State National Research University

**ENGLISH TEXT EDITOR: Igor V. Lyashenko**, Candidate of Philological Sciences, Associate Professor, Associate Professor of the Department of English Philology and Intercultural Communication, Belgorod State National Research University

#### EDITORIAL BOARD:

**Oleg O. Basov**, Doctor of Technical Sciences, Associate Professor (Russia)

**Sergey P. Belov**, Doctor of Technical Sciences, Professor (Russia)

**Valery P. Volchkov**, Doctor of Technical Sciences, Professor (Russia)

**Valery D. Dmitrienko**, Doctor of Technical Sciences, Professor (Ukraine)

**Olga A. Ivaschuk**, Doctor of Technical Sciences, Professor (Russia)

**Igor A. Kalmykov**, Doctor of Technical Sciences, Professor (Russia)

**Nikolay I. Korsunov**, Honoured Science Worker of Russian Federation, Doctor of Technical Sciences, Professor (Russia)

**Alexander V. Koskin**, Doctor of Technical Sciences, Professor (Russia)

**Vadim A. Lomazov**, Doctor of Physico-mathematical Sciences, Professor (Russia)

**Sergey I. Matorin**, Doctor of Technical Sciences, Professor (Russia)

**Yulia A. Orlova**, Doctor of Technical Science, Associate Professor (Russia)

**Valery B. Taranchuk**, Doctor of Physico-mathematical Sciences, Professor (Belarus)

Учредитель: Федеральное государственное автономное образовательное учреждение высшего образования  
«Белгородский государственный национальный исследовательский университет»  
Издатель: НИУ «БелГУ». Адрес издателя: 308015 г. Белгород, ул. Победы, 85.  
Журнал выходит 4 раза в год

Founder: Federal state autonomous educational establishment of higher education  
«Belgorod State National Research University»  
Publisher: Belgorod State National Research University  
Address of publisher: 85 Pobeda St., Belgorod, 308015, Russia  
Publication frequency: 4/year

**СОДЕРЖАНИЕ****CONTENTS****ИНФОРМАЦИОННЫЕ СИСТЕМЫ  
И ТЕХНОЛОГИИ****INFORMATION SYSTEM  
AND TECHNOLOGIES**

<b>Резниченко О.С.</b> Методы и программные инструменты для извлечения публикационных данных научной организации, размещенных на портале eLIBRARY.RU	<b>3</b>	<b>Reznichenko O.S.</b> Methods and Software Tools f or Uploading Publication Data of a Scientific Organization Posted on the eLIBRARY.RU	<b>3</b>
<b>Назаров Т.Р.</b> Формирование инвестиционных портфелей с использованием кластерного анализа	<b>20</b>	<b>Nazarov T.R.</b> Formation of Investment Portfolios Using Cluster Analysis	<b>20</b>

**АВТОМАТИЗАЦИЯ И УПРАВЛЕНИЕ****AUTOMATION AND CONTROL**

<b>Рыбанов А.А., Свиридова О.В., Толстяков В.И.</b> Мобильное приложение для мониторинга биометрических и кинематических параметров с автономной аналитикой и адаптивной персонализацией	<b>31</b>	<b>Rybanov A.A., Sviridova O.V., Tolstoyakov V.I.</b> Mobile Application for Monitoring Biometric and Kinematic Parameters with Autonomous Analytics and Adaptive Personalization	<b>31</b>
---	-----------	--	-----------

**ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ  
И ПРИНЯТИЕ РЕШЕНИЙ****ARTIFICIAL INTELLIGENCE  
AND DECISION MAKING**

<b>Минина А.В., Никулин Р.Р., Сидоренко И.А.</b> Нейросетевая аппроксимация координат в задачах локализации с временными провалами сигнала	<b>45</b>	<b>Minina A.V., Nikulin R.R., Sidorenko I.A.</b> Neural Inference of Object Localization from Discontinuous Time-of-arrival Sequences	<b>45</b>
<b>Щетинин Е.Ю., Тютюнник А.А.</b> О сегментации полипов с использованием модели Segment Anything Model	<b>55</b>	<b>Shchetinin E.Yu., Tiutiunnik A.A.</b> On Segmentation of Polyps Using Segment Anything Model	<b>55</b>

**КОМПЬЮТЕРНОЕ МОДЕЛИРОВАНИЕ****COMPUTER SIMULATION**

<b>Константинов И.С., Таха А.Т.Т., Голдобина Д.М.</b> Эпидемиологические модели SEIRD для болезней растений	<b>64</b>	<b>Konstantinov I.S., Taha A.T.T., Goldobina D.M.</b> SEIRD Epidemiological Models for Plant Disease	<b>64</b>
<b>Лихошерстный А.Ю., Великая Я.Г.</b> Струнно-волновой параллельный алгоритм решения СЛАУ для разреженных матриц	<b>72</b>	<b>Likhosherstnyy A.Yu., Velikaya Ya.G.</b> String-Wave Direct Parallel Solver for Sparse System of Linear Equations	<b>72</b>
<b>Ганичева А.В., Ганичев А.В.</b> Математическая модель усвоения учебного материала	<b>80</b>	<b>Ganicheva A.V., Ganichev A.V.</b> Mathematical Model of Learning Material Assimilation	<b>80</b>

**ИНФОРМАЦИОННЫЕ СИСТЕМЫ И ТЕХНОЛОГИИ  
INFORMATION SYSTEM AND TECHNOLOGIES**

УДК 004.62,004.42

DOI: 10.18413/2518-1092-2025-10-3-0-1

Резниченко О.С.

**МЕТОДЫ И ПРОГРАММНЫЕ ИНСТРУМЕНТЫ  
ДЛЯ ИЗВЛЕЧЕНИЯ ПУБЛИКАЦИОННЫХ ДАННЫХ  
НАУЧНОЙ ОРГАНИЗАЦИИ, РАЗМЕЩЕННЫХ  
НА ПОРТАЛЕ ELIBRARY.RU**Белгородский государственный национальный исследовательский университет,  
ул. Победы, 85, г. Белгород, 308015, Россия*e-mail: oreznichenko@bsuedu.ru***Аннотация**

Публикационные отчеты занимают важную долю среди всей отчетности научной организации. Они нужны не только для текущего мониторинга, но и для планирования будущей деятельности. Отчетность включает в себя в том числе информацию о количестве и списочном составе публикаций в различных разрезах и типах, например, РИНЦ, публикации из RSCI, журналов из списка ВАК, журналов из «Белого списка» и других. Сейчас основной источник данных для отчетов – это научная электронная библиотека eLIBRARY.RU. Списки публикаций из профиля организации в eLIBRARY.RU можно получить путем ручной выгрузки публикационных данных, используя API, используя возможности аналитической системы SCIENCE INDEX или парсинге веб-содержимого. В статье описывается разработка алгоритмов и инструментов выгрузки и парсинга содержимого страниц, которые позволяют получать сведения о названиях, списках авторов и источниках публикаций в табличном виде. Автоэкспорт веб-страниц и их анализ основан на использовании эмуляции работы пользователя в браузере и использовании серии опорных тегов для парсинга. Созданные скрипты можно скомпилировать в исполняемый файл и сократить время сохранения и преобразования веб-страниц eLIBRARY.RU на 5 минут для каждых 100 статей поисковой выдачи при ручной обработке.

**Ключевые слова:** наукометрия; публикационная активность; публикационная отчетность; веб-скрейпинг; парсинг; eLIBRARY.RU; Microsoft Excel; SCIENCE INDEX; Selenium WebDriver; Python; BeautifulSoup; pandas.DataFrame

**Для цитирования:** Резниченко О.С. Методы и программные инструменты для извлечения публикационных данных научной организации, размещенных на портале eLIBRARY.RU // Научный результат. Информационные технологии. – Т.10, №3, 2025. – С. 3-19. DOI: 10.18413/2518-1092-2025-10-3-0-1

Reznichenko O.S.

**METHODS AND SOFTWARE TOOLS  
FOR UPLOADING PUBLICATION DATA OF A SCIENTIFIC  
ORGANIZATION POSTED ON THE ELIBRARY.RU**Belgorod State National Research University,  
85 Pobedy St., Belgorod, 308015, Russia*e-mail: oreznichenko@bsuedu.ru***Abstract**

Publication reports are an important part of the overall reporting process in a scientific organization. They are necessary not only for monitoring the organization's status but also for planning future activities. Publication reports include information on the count and list of publications in various formats, such as articles from the Russian Science Citation Index, articles in journals from the HAC

list, articles in journals from the "White List," and many others. Currently, the main source of input data for publication reports is the scientific electronic library eLIBRARY.RU. Lists of publications from the organization's profile on eLIBRARY.RU can be manually uploaded, through the API, using tools from the SCIENCE INDEX analytical system, or via web scraping techniques. This article describes the development of algorithms and software tools for extracting and scraping page content, which allow for the retrieval of data about titles, author lists, and journal information in spreadsheet form. The automatic saving of web pages and their scraping is based on emulation browser user actions and the use of specific HTML tags. The created Python scripts can be compiled into standalone executable files, reducing the time required to upload and transform the content of eLIBRARY.RU web pages into spreadsheets by 5 minutes for every 100 articles in search results during manual processing.

**Keywords:** scientometrics; publication activity; publication reports; web scraping; eLIBRARY.RU; Microsoft Excel; SCIENCE INDEX; Selenium WebDriver; Python; BeautifulSoup; pandas.DataFrame

**For citation:** Reznichenko O.S. Methods and software tools for uploading publication data of a scientific organization posted on the eLIBRARY.RU // Research result. Information technologies. – Т.10, №3, 2025. – P. 3-19. DOI: 10.18413/2518-1092-2025-10-3-0-1

## **ВВЕДЕНИЕ**

Важную роль среди отчетов научной организации о своей деятельности является публикационная отчетность. Показатели публикационной активности необходимы для принятия управленческих решений, так как на основе этих показателей Минобрнауки определяет грантовое финансирование, а также объемы финансирования в рамках программы «Приоритет-2030». В Белгородском государственном национальном исследовательском университете (далее – БелГУ, университет) с различной периодичностью приходится формировать несколько видов публикационных отчетов. Среди внешней публикационной отчетности можно упомянуть следующие её виды:

- отчеты о количестве публикаций в наукометрических реферативных базах Scopus и Web of Science Core Collection в рамках выполнения реализации программы стратегического академического лидерства «Приоритет-2030» [1];
- отчеты о деятельности научно-образовательного центра мирового уровня «Инновационные решения в АПК» [2];
- отчеты о составе публикаций, посвященных вопросам обеспечения условий для обучения и социализации инвалидов и лиц с ограниченными возможностями здоровья в рецензируемых научных журналах;
- списки монографий, глав монографий и словарей;
- список публикаций, посвященных вопросам реализации бережливого производства.

В качестве внутренней периодической отчетности выступают следующие публикационные отчеты:

- состав и количество статей университета в журнал из «Белого списка» [3];
- количество публикаций структурных подразделений университета в реферативных базах Scopus и Web of Science Core Collection [4];
- количество и состав публикаций отдельных сотрудников университета в реферативных базах Scopus, Web of Science Core Collection.

Кроме этого, периодический мониторинг количества публикаций в различных разрезах позволяет строить прогнозные модели на будущие периоды и формировать на их основе как планы публикационной активности конкретных научных подразделений, так и план публикационной активности в целом по университету.

Большую часть требуемых публикационных отчетов можно получить, анализируя списки публикаций, размещенных в научной электронной библиотеке eLIBRARY.RU (далее – elibrary.ru)

[5]. Профиль научной организации на сайте elibrary.ru позволяет отображать списки аффилированных публикаций согласно следующим актуальным для университета фильтрам:

- публикации организации, включенные в РИНЦ [6];
- публикации, имеющие DOI;
- статьи в журналах, входящих в Russian Science Citation Index [7];
- статьи в журналах, включенных в «Белый список»;
- статьи в российских журналах, включенных в перечень ВАК [8].

Сформированный после применения соответствующих фильтров список публикаций на портале elibrary.ru представляет собой последовательность веб-страниц на каждой из которых отображается до 100 публикаций при условии осуществления авторизованного входа на портал (рис. 1).

**Всего найдена 261 публикация с общим количеством цитирований: 74.**  
Показано на данной странице: с 1 по 100.

№	Публикация	Цит.
1.	<b>ГЕНЕТИЧЕСКИЕ ДЕТЕРМИНАНТЫ УРОВНЯ АНГИОТЕНЗИНПРЕВРАЩАЮЩЕГО ФЕРМЕНТА (ДАННЫЕ ПОЛНОГЕНОМНЫХ ИССЛЕДОВАНИЙ)</b> Камышникова Л.А., Ефремова О.А., Фентисов В.В., Болховитина О.А., Чурносов М.И. Артериальная гипертензия. 2024. Т. 30. № 6. С. 537-552.	0
2.	<b>СРАВНИТЕЛЬНАЯ МЕТАГЕНОМИКА ДЛЯ МОНИТОРИНГА СКРЫТОЙ ДИНАМИКИ АЛЬГО-БАКТЕРИАЛЬНОГО СООБЩЕСТВА СТОЧНЫХ ВОД ПОД ДЕЙСТВИЕМ ЛЕКАРСТВЕННЫХ ВЕЩЕСТВ</b> Зайцев П.А., Шурыгин Б.М., Родин В.А., Панова Т.В., Зверева М.Э., Скрипникова Е.В., Соловченко А.Е. Российские нанотехнологии. 2024. Т. 19. № 4. С. 505-519.	0
3.	<b>РАДИОНУКЛИДЫ (<sup>40</sup>K, <sup>232</sup>Th, <sup>226</sup>Ra, <sup>137</sup>Cs, <sup>90</sup>Sr) В АГРОЭКОСИСТЕМАХ ЦЕНТРАЛЬНОГО ЧЕРНОЗЕМЬЯ</b> Лукин С.В. Доклады Российской академии наук. Науки о Земле. 2024. Т. 519. № 2. С. 156-162.	0
4.	<b>РАЗЛИЧИЕ СПЕКТРОВ ПАРАМЕТРИЧЕСКОГО РЕНТГЕНОВСКОГО ИЗЛУЧЕНИЯ ИЗ ПОРОШКОВЫХ МИШЕНЕЙ И ТЕКСТУРИРОВАННЫХ ФОЛЬГ</b> Кишин И.А., Киданова Е.Ю., Кубанкин А.С., Нажмудинов Р.М., Алексеев В.И., Елисеев А.Н. Письма в журнал Физика элементарных частиц и атомного ядра. 2024. Т. 21. № 4 (255). С. 782-786.	0
5.	<b>ИЗМЕРЕНИЯ ПОЛЯРИЗАЦИИ ПУЧКОВ ДЕЙТРОНОВ И ПРОТОНОВ НА НУКЛОТРОНЕ</b> Волков И.С., Ладыгин В.П., Схоменко Я.Т., Гурчин Ю.В., Исупов А.Ю., Янек М., Карачук Ю.Т., Хренов А.Н., Курилкин П.К., Ливанов А.Н., Плядин С.М., Резников С.Г., Терёхин А.А., Тишевский А.В., Аверьянов А.В., Черных Е.В., Еначе Д., Кривенков Д.О., Внуков И.Е. Письма в журнал Физика элементарных частиц и атомного ядра. 2024. Т. 21. № 4 (255). С. 732.	0
6.	<b>ЭКОЛОГИЧЕСКАЯ ОЦЕНКА СОДЕРЖАНИЯ ТЯЖЕЛЫХ МЕТАЛЛОВ И МЫШЬЯКА В ПОЧВАХ И СЕЛЬСКОХОЗЯЙСТВЕННЫХ РАСТЕНИЯХ ЦЕНТРАЛЬНОГО ЧЕРНОЗЕМЬЯ</b> Лукин С.В. Почвоведение. 2024. № 10. С. 1414-1423.	0
	Версии: ECOLOGICAL ASSESSMENT OF CONCENTRATIONS OF HEAVY METALS AND ARSENIC IN SOILS AND CROPS OF THE CENTRAL CHERNOZEMIC REGION Lukin S.V. Eurasian Soil Science. 2024. Т. 57. № 10. С. 1709-1717.	0

Рис. 1. Часть внешнего вид профиля научной организации в режиме отображения списка публикаций

Fig. 1. User interface of a scientific organization profile in the publication list display mode

## **ОБЗОР МЕТОДОВ ИЗВЛЕЧЕНИЯ ПУБЛИКАЦИОННЫХ ДАННЫХ ИЗ ELIBRARY.RU**

Среди возможных способов получения отфильтрованного списка публикаций в удобном для анализа виде (например, в виде электронной таблицы) авторы предполагают следующие варианты:

- 1) сохранение веб-страниц с результатами поиска публикаций, открытие их в Microsoft Office Excel (далее – Excel) с последующей обработкой содержимого этих веб-страниц;
- 2) использование возможностей интерфейса прикладного программирования API [9];
- 3) использование возможностей информационно аналитической системы SCIENCE INDEX (далее – система Science Index) [10];
- 4) синтаксический анализ списка публикаций на портале онлайн;
- 5) синтаксический анализ сохраненных списков публикаций с портала в виде выгруженных веб-страниц.

Рассмотрим достоинства и недостатки каждого из вариантов решения задачи.

Среди достоинств первого метода можно указать отсутствие необходимости использовать какие-либо программные инструменты, кроме интернет-браузера и процессора электронных таблиц. Главным недостатком этого метода является объем временных затрат, который составляет по оценкам авторов не менее 300 секунд на каждые 100 статей при условии уверенного владения исполнителем основами работы в процессорах электронных таблиц. Оценочное время получено авторами на персональном компьютере на базе процессора Intel Intel Core i7-4770 с 16 Гб оперативной памяти с операционной системой Microsoft Windows 11 Pro (сборка 26100.3194), посредством браузера Opera One (версия: 117.0.5408.39) и процессора электронных таблиц Microsoft Office Excel 2016. Следует также учесть, что для публикаций с большим количеством соавторов не все они попадут в соответствующий столбец, так как представлены в урезанном виде на исходной веб-странице.

Использование встроенного в elibrary.ru API позволяет выгружать данные о публикациях научной организации в формат xml. Вопросы получения информации о публикационной активности научной организации посредством встроенного API рассмотрены в работе [11]. Данный способ доступен только научным организациям, имеющим, кроме доступа к системе Science Index, дополнительную лицензию, дающую право на увеличение спектра и количества сведений, которые можно выгружать посредством API [9].

В БелГУ имеется лицензия на доступ к системе Science Index, предполагающая в том числе наличие неограниченного экспорта списка публикаций в формате XML. Вопросы разработки и использования инструментов преобразования данных, выгруженных в формате xml, в электронную таблицу формата Excel рассмотрены в работе [12]. Содержимое выходного файла, полученного в результате применения этого инструмента легко можно анализировать и подвергать дальнейшей трансформации для получения сведений о публикациях в различных разрезах. В тоже время, ориентация на использования данного метода сопряжено с несколькими особенностями, ограничивающими его применение:

- обязательное наличие платной лицензии на доступ к системе Science Index;
- необходимость каждый раз адаптировать входные данные, полученные на этапе их подготовки, для корректной работы этого инструментария;
- некорректной работе функции экспорта в xml на портале elibrary.ru при использовании некоторых типов фильтров (например, для фильтра статей в журналах, включенных в «Белый список»), что приводит к невозможности получить актуальное количество публикаций;
- нестабильной работе функции фильтрации публикаций научной организации по некоторым фильтрам. Периодически при фильтрации может возникать ошибка типа «Microsoft OLE DB Provider for ODBC Drivers error '80040e31'».

Онлайн синтаксический разбор веб-страницы с результатами фильтрации списка публикаций научной организации также связан с несколькими трудностями:

- необходимо устанавливать и настраивать специализированные библиотеки, эмулирующее работу интернет-браузеров, так как elibrary.ru не формирует веб-страницы с результатами поиска на основе сгенерированной адресной ссылки, в структуру которой входят параметры поискового запроса, что делает невозможным пакетно обрабатывать эти поисковые результаты программными инструментами. Кроме того, портал elibrary.ru имеет встроенную защиту от парсинга, которую можно обойти только используя эмуляцию действий пользователя в интернет-браузере, например, посредством Selenium WebDriver. Пример разработки подобного инструментария для парсинга содержимого профилей авторов рассматривается в исследовании [13];

- имеющийся инструмент парсинга содержимого веб-страниц портала elibrary.ru, описанный в источнике [14] также предназначен только для экспорта списка публикаций конкретного автора по его уникальному идентификатору, и, аналогично, предполагает использование Selenium WebDriver.

Видно, что описанные выше способы реализации метода онлайн парсинга предполагают наличие у исполнителя навыков владения информационными технологиями выше среднего, так как сопровождаются настройкой и использованием дополнительных программных инструментов. Кроме того, скрипты, описанные в источниках выше, предполагают значительную их доработку для извлечения списка публикаций именно из профиля научной организации, включающую разработку сценария эмуляции действий пользователя для Selenium WebDriver. При этом процесс эмуляции, как и в случае ручной навигации по параметрам фильтров и их результатов может сопровождаться ошибками доступа к базе данных elibrary.ru.

В качестве альтернативы использования эмуляторов браузеров может выступать только ручная постраничная выгрузка результатов поиска в отдельные html-файлы. Подобная работа может быть выполнена не менее чем за 30 секунд на каждые 100 статей поисковой выдачи.

Подытоживая все вышесказанное, автор предлагает остановиться на реализации метода, предполагающего автоматическое сохранение веб-страниц со списком публикаций и синтаксический анализ этих списков, так как остальные методы либо занимают много времени, либо привязаны к наличию лицензии на использование API.

### ***ФОРМАТ И СТРУКТУРА ВХОДНЫХ И ВЫХОДНЫХ ДАННЫХ***

В качестве исходных данных для разрабатываемого программного обеспечения выступают наборы html-файлов, представляющие собой сохраненные вручную через браузер веб-страницы с результатами фильтрации списка публикаций научной организации (рис. 2). Из рисунка очевидно, что, анализируя текст html-файла можно получить табличные данные, имеющие структуру, представленную в таблице 1.



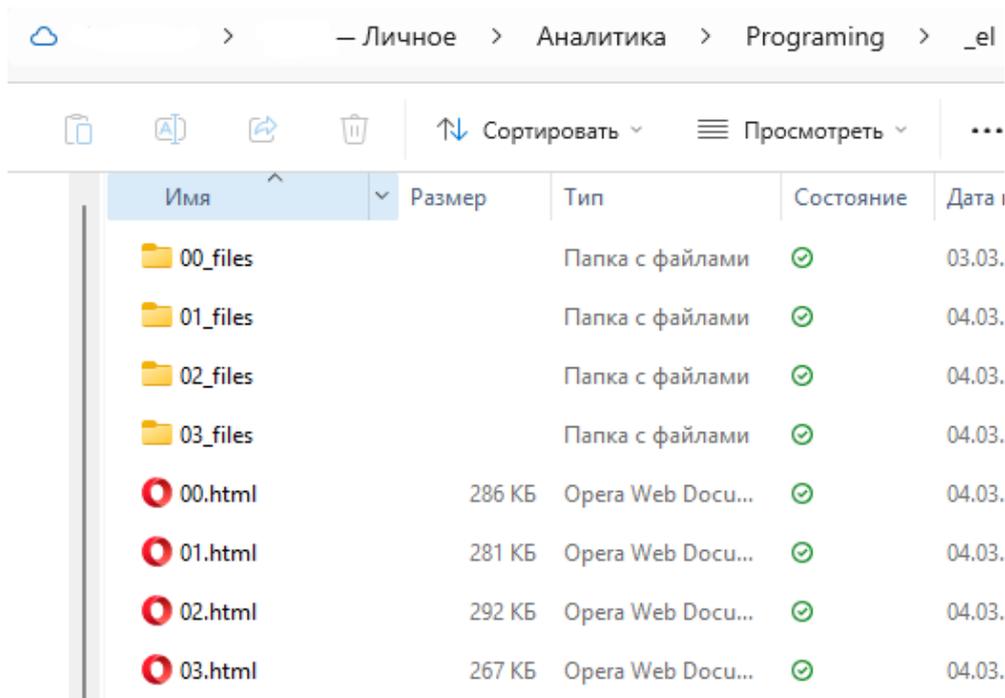


Рис. 3. Содержимое папки “\_el” с исходными файлами для анализа  
Fig. 3. Contents of the “\_el” folder with source files for analysis

### **ОПИСАНИЕ ИНСТРУМЕНТОВ И СРЕДСТВ РАЗРАБОТКИ ПРИЛОЖЕНИЙ**

В данном разделе перечисляется программное обеспечение, которое потребуется для получения исходных данных, а также набор средств и сред разработки, которые необходимы для написания исходного кода и генерации исполняемого файла из полученного кода. Среди необходимого для реализации исследования программного обеспечения можно указать следующее:

- интернет-браузер с возможностью сохранения веб-страниц в формате html и поддерживаемый библиотекой Selenium WebDriver, например, Google Chrome, Microsoft Edge, Mozilla Firefox или Apple Safari [15];
- интерпретатор Python версии не ниже 3.12.6 [16];
- JetBrains PyCharm (Community Edition) версии не ниже 2024.3.4 в качестве интегрированной среды разработки [17];
- программная библиотека pandas для генерации и обработки табличных данных [18];
- программная библиотека BeautifulSoup для синтаксического разбора html- и xml-документов [19];
- программная библиотека Selenium WebDriver как инструмент для автоматизации действий веб-браузера;
- библиотека Pyinstaller для упаковки Python-приложений в standalone-исполняемые файлы для ОС Windows [20].

После компиляции исполняемого файла программы процесс преобразования данных для потенциального пользователя можно ограничить только необходимостью использовать браузер и скомпилированные исполняемые программы.

### **ОПИСАНИЕ ЭТАПОВ РАЗРАБОТКИ ПРОГРАММНОГО ИНСТРУМЕНТАРИЯ**

Общий принцип действия разрабатываемых скриптов для экспорта данных из elibrary.ru и формирования табличного файла со сведениями о публикациях научной организации на основе синтаксического анализа содержимого нескольких сохраненных страниц с портала elibrary.ru

основан на постолбцовом формировании таблицы с выгруженными публикационными данными. Алгоритм автоматической фильтрации и сохранения результатов применения фильтра состоит из нескольких этапов (рис. 4).

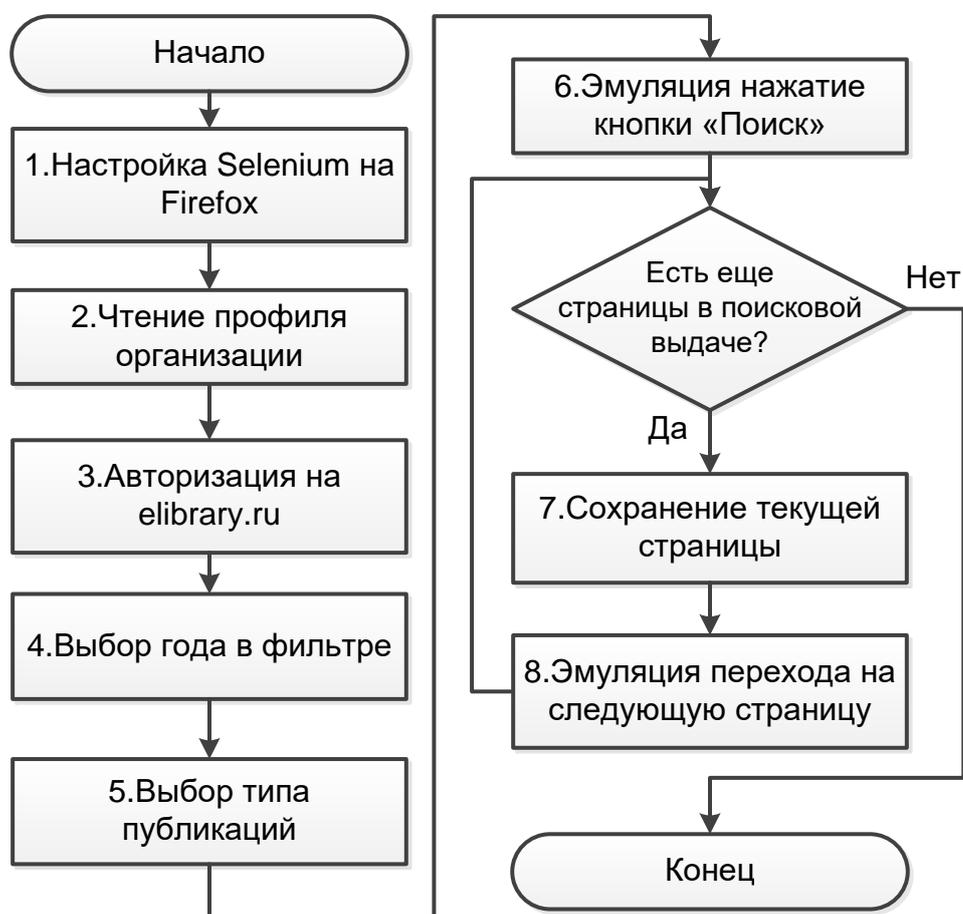


Рис. 4. Блок-схема алгоритма автоматического экспорта страниц из портала eLIBRARY.RU  
Fig. 4. Flowchart of the algorithm for automatic export of web pages from the eLIBRARY.RU

На базе алгоритма был разработан скрипт, применение которого приведет к созданию набора html-файлов в папке “\_el” с выходными публикационными данными. В основе скрипта лежит проект, представленный в работе [14]. Для составления собственной версии скрипта используются следующие шаги, включающие программные методы и опорные теги html-кода страниц:

Шаг 1 – настройка Selenium WebDriver на эмуляцию в Mozilla Firefox;

Шаг 2 – открытие профиля научной организации в режиме просмотра списка статей посредством метода `driver.get`.

Шаг 3 – авторизация на портале `elibrary.ru` (теги `<input name='login'>`, `<input name='password'>` и `<div onclick='check_all()'>`).

Шаг 4 – раскрытие списка годов и выбор конкретного года (свойства `'id="hdr_years"'` и `'id="year_2024"'`);

Шаг 5 – выбор типа публикаций из всплывающего списка (свойство `'value="3"'`, где ‘3’ значение, соответствующее варианту “статьи, в журналах, входящих в Russian Science Citation Index” в всплывающем списке);

Шаг 6 – эмуляция нажатия кнопки «Поиск» (признак `'//td[6]/div'`);

Шаг 7 – сохранение текущей страницы в файл с помощью метода `driver.page_source`;

Шаг 8 – эмуляция нажатия ссылки «Следующая страница», на основе поиска ссылки по тексту.

Одна из оригинальных составляющих скрипта реализует эмуляцию процесса авторизации на портале `elibrary.ru` (шаг 3 алгоритма на рисунке 1). Открывать профиль научной организации в

режиме просмотра списка публикаций можно и без авторизации, однако при автоматизации процесса сохранения веб-страниц настройки фильтров могут сбрасываться в процессе перехода к следующей странице с результатами поиска. Скрипт авторизации приведен в листинге 1.

Листинг 1. Скрипт авторизации на портале elibrary.ru

Source Code 1. Authorization script for eLIBRARY.RU

```
driver.find_element(By.XPATH, "///input[@name='login']").send_keys("login")
driver.find_element(By.XPATH, "///input[@name='password']").send_keys("password")
time.sleep(sleep_seconds)
driver.find_element(By.XPATH, "///div[@onclick='check_all()']").click()
```

Полностью весь скрипт автоэкспорта публикационных веб-данных в файлы приведен на хостинге ИТ-проектов GitHub [22].

Каждому из столбцов выходной таблицы, с приведенной выше структурой (табл. 1), будут соответствовать набор признаков во входном html-файле, на основе которого будет определяться принадлежность текста внутри тегов к тому или иному столбцу. Признаки представляют собой цепочки опорных тегов и их свойств. Конкретные особенности выбора состава тегов и их свойств будет рассмотрен далее. Общий алгоритм работы программы синтаксического анализа входных файлов, основанный на использовании методов класса BeautifulSoup, представлен на рисунке 5



Рис. 5. Блок-схема общего алгоритма программы для синтаксического офлайн-анализа сохраненных публикационных данных из портала eLIBRARY.RU

Fig. 5. Flowchart of the main algorithm of the software for offline web scraping of saved publication data from the eLIBRARY.RU

При извлечении ссылок на статьи в реализации шага 4 общего алгоритма и заголовков статей (шаг 5) необходимо учитывать, что на странице с результатами поиска могут присутствовать несколько версий одной и той же статьи, например, выходные данные статьи в переводной версии журнала (см. рис. 1). При реализации шага 4 из содержимого веб-страницы сначала извлекаются наборы текста тегов `<href>` по схеме: для всех `<href>` из списка всех тегов `<a>`, которые входят в теги `<td>`. Сокращенно здесь и далее подобные цепочки тегов будут обозначаться как `td -> a -> href`. Затем из полученного списка отфильтровываются только непустые элементы, внутри которых встречается текст `'elibrary.ru/item.asp?id='` (листинг 2).

*Листинг 2.* Исходный код скрипта получения списка ссылок на публикации из файлов, выгруженных на портале eLIBRARY.RU

*Source Code 2.* Source code of the script for obtaining a list of URLs to publications from pages uploaded from the eLIBRARY.RU

```
links_eng = [a.get('href') for a in (td.find('a') for td in bs.find_all('td')) if a]
urls_100 = [i for i in links_eng if (i is not None) and ('elibrary.ru/item.asp?id='
in i)]
```

В процессе реализации шага 5 из веб-страницы извлекаются тексты тегов `<span>` с атрибутом `style="line-height:1.0;"` в виде списка с именем `tituls_100` (листинг 3).

*Листинг 3.* Исходный код скрипта для получения списка заголовков статей

*Source Code 3.* Source code of the script for getting the list of article titles

```
tituls_100 = bs.find_all('span', attrs={ "style" : "line-height:1.0;"})
tituls_100 = [i.text for i in tituls_100]
```

Перед выполнением шага 6 формируется список `urls_eng` из всех ссылок на дублирующие версии статей согласно цепочке тегов `span class = "menug" ->a -> href`

Шаг 6 реализован в виде двух вложенных циклов, в которых из списков `tituls_100` и `urls_100` удаляются элементы с номерами, предшествующими номерам элементов из списка `urls_eng`, то есть в списках со ссылками и заголовками статей остаются только вторые версии ссылок и заголовков (рис. 6).

Особенность реализации скрипта для шага 7 общего алгоритма (см. рис. 5) заключается в том, что на исходной веб-странице есть много элементов, подходящих под соответствующую опорную цепочку тегов `td -> font color=#00008f`, поэтому необходимо дополнительно урезать получившийся после применения фильтра по опорным тегам список до элемента, равного значению “Всего найдено”. Получившийся урезанный список включает чередующиеся элементы, включающие порядковый номер публикации в списке с результатами поиска и список соавторов. Чтобы оставить только списки соавторов применяется срез по четным элементам списка `authors_vyh`:

```
authors_vyh = authors_vyh[1::2]
```

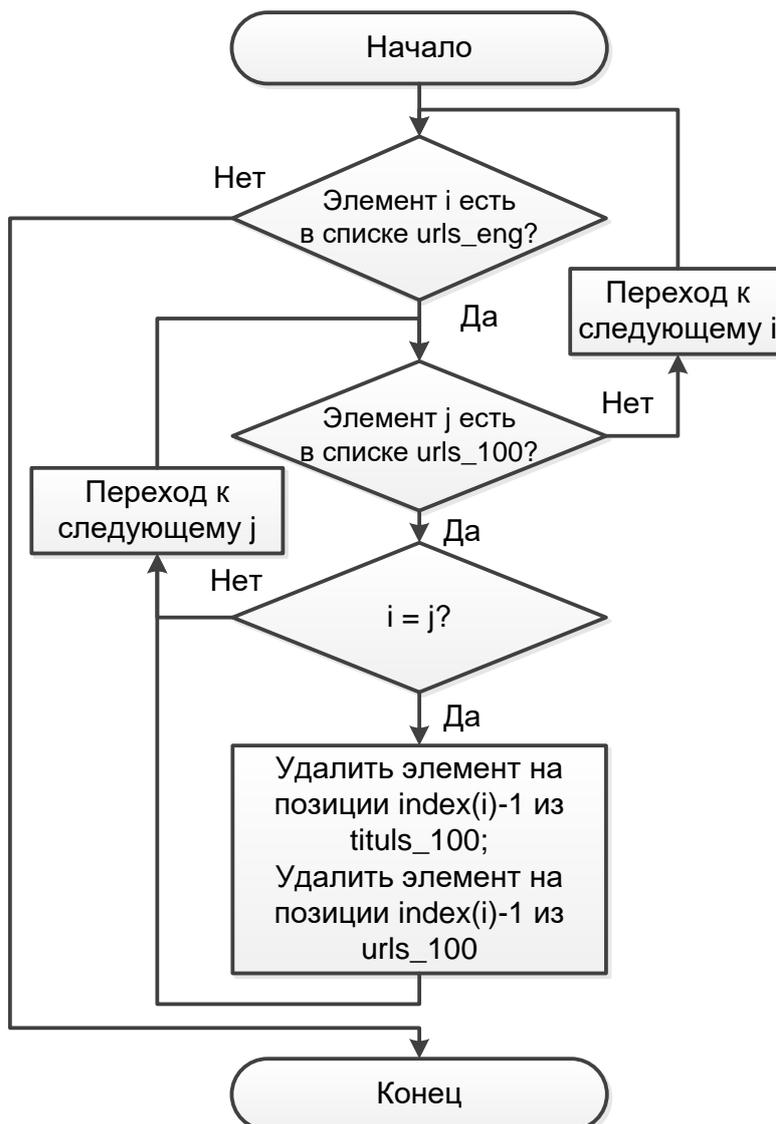


Рис. 6. Блок-схема алгоритма получения окончательного варианта списков ссылок и заголовков статей без дублирования

Fig. 6. Flowchart of the algorithm for obtaining the final version of lists of URLs and article titles without duplication

В выходных данных публикаций в файле с поисковыми результатами не у каждой публикации имеется список соавторов, например, если публикация представляет собой сборник статей по материалам конференции (рис. 7).

№	Публикация	Цит.
101.	<b>ФИЛОЛОГИЯ И НОВОЕ ЗНАНИЕ</b> <input type="checkbox"/> Сборник материалов IV Российской научной конференции с международным участием / Белгород, 2024. 	0

Рис. 7. Пример выходных данных публикации без списка её соавторов  
 Fig. 7. Example of publication output data without a list of its co-authors

В связи с этим необходимо в качестве элементов списка, в которых текст не заключен в тег `<i>` вставлять пустую строку. Для реализации корректировки итогового списка используется сравнение двух списков – набора списка соавторов в тексте тегов `<i>` (переменная `authors_err`) и полного списка с авторами и возможными выходными данными (переменная `authors_vyh`). Блок-схема алгоритма получения списка соавторов публикаций приведена на рисунке 8.

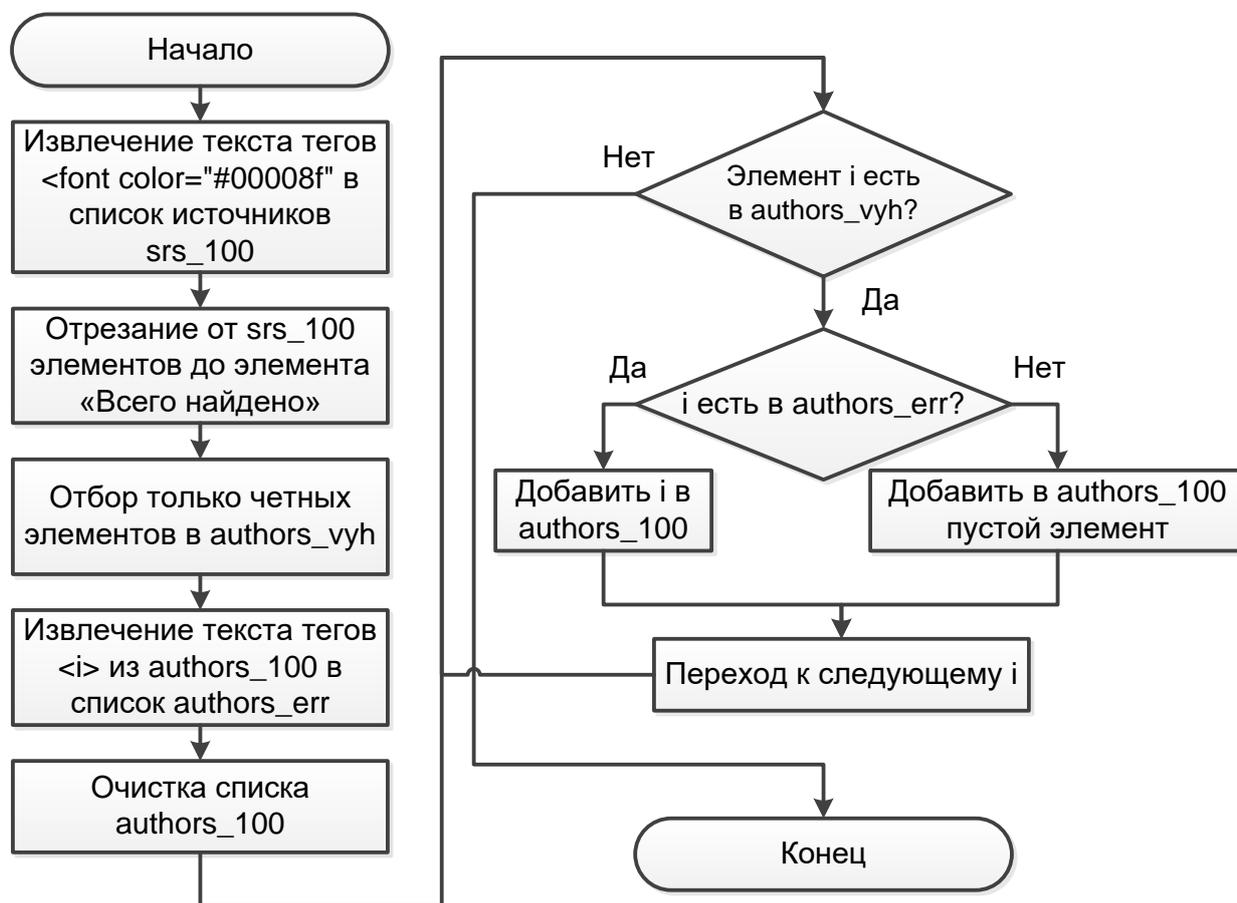


Рис. 8. Блок-схема алгоритма получения списка соавторов публикаций  
Fig. 8. Flowchart of the algorithm for obtaining a list of co-authors of publications

Не менее трудоемким является алгоритм получения информации об источниках публикаций (шаг 8 общего алгоритма). Это обусловлено тем, что, если использовать опорный тег `<font color=#00008f>`, извлекаются не только сведения об источниках, но и списки соавторов. После урезания полученного списка до элемента «Всего найдено» необходимо удалить из него элементы, присутствующие в списке с соавторами (переменная `authors_100`). Кроме это необходимо выполнить следующие операции:

- 1) урезать список, удалив из начала 2 лишних незначащих элемента;
- 2) убрать элементы равные символу конца строки `'\n'`;
- 3) убрать номера публикаций на нечетных местах;
- 4) удалить в элементах списка символы конца строки `('\'n')` и возврат каретки `('\'r)`.

Весь алгоритм получения информации об источниках статей приведен на рисунке 9.

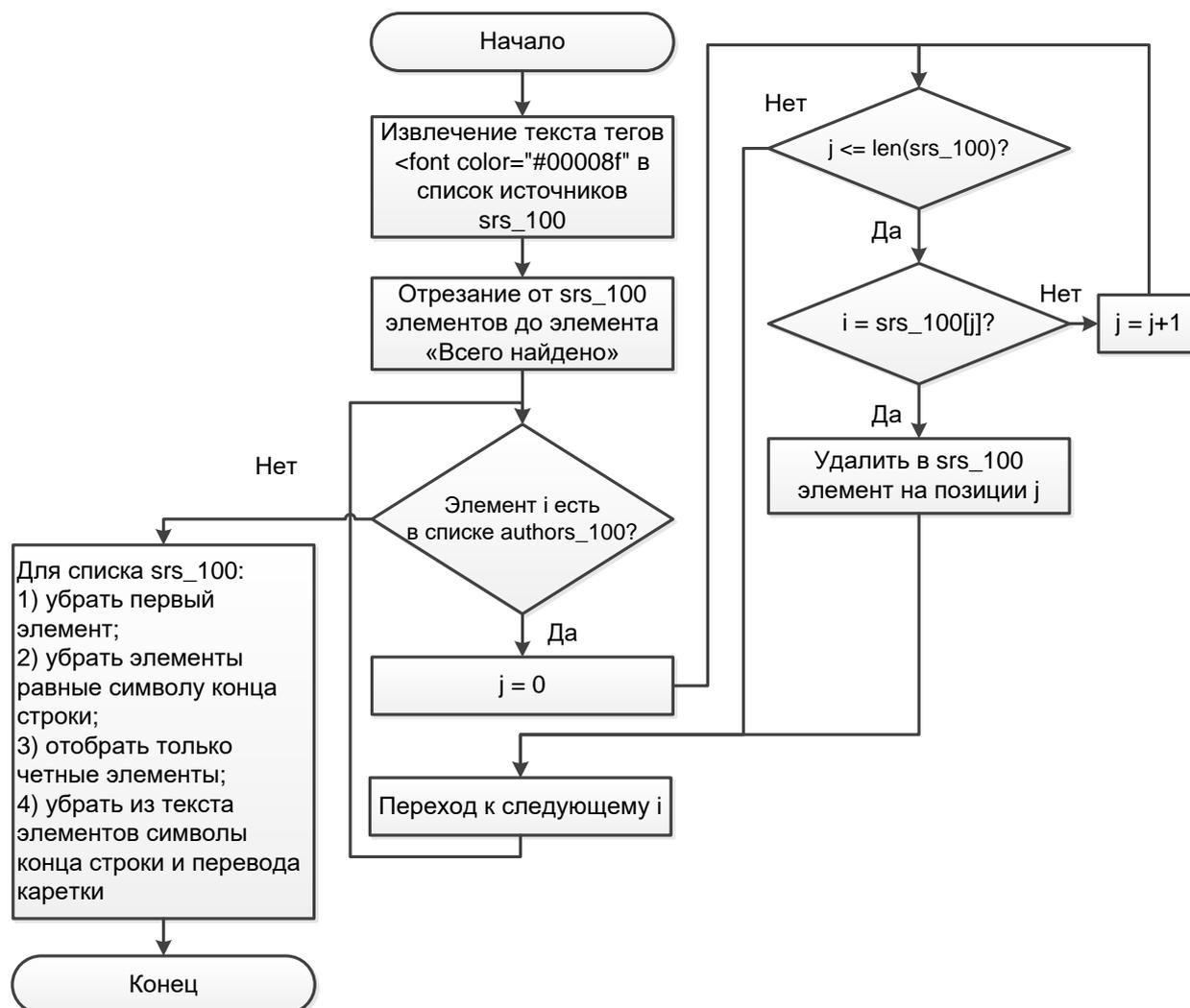


Рис. 9. Блок-схема алгоритма получения информации об источниках статей  
Fig. 9. Flowchart of the algorithm for obtaining data about article sources

Последнее поле в структуре таблицы 1 – это идентификатор публикации на портале eLibrary.ru. Содержимое этого поля получается в результате извлечения значений атрибута arw из цепочки тегов tr -> id (если тег <id> встречается в <tr>) (листинг 4).

Листинг 4. Исходный код скрипта получения уникальных идентификаторов публикации на портале eLIBRARY.RU

Source code 4. Source code of the script for obtaining unique articles identifiers on the eLIBRARY.RU

```
tr_100 = bs.find_all('tr')
id_100 = [i.get('id') for i in tr_100 if i.has_attr('id')]
id_100 = [i for i in id_100 if 'arw' in i]
id_100 = [i[3:] for i in id_100]
```

В результате реализации скриптов получают 5 списков, которые распределяются по результирующему фрейму данных tempdf, а тот, в свою очередь добавляется к результирующему фрейму данных elxls (листинг 5).

Листинг 5. Исходный код скрипта для сборки результирующего фрейма данных elxls  
Source code 5. Source code of the script for assembling the resulting DataFrame 'elxls'

```
tempdf = pd.DataFrame()  
tempdf.insert(0, 'IDs', id_100)  
tempdf.insert(1, 'Авторы', authors_100)  
tempdf.insert(2, 'Название', tituls_100)  
tempdf.insert(3, 'Источник', srs_100)  
tempdf.insert(4, 'URL', urls_100)  
elxls = pd.concat([elxls, tempdf])
```

## **РЕЗУЛЬТАТЫ ИССЛЕДОВАНИЯ И ИХ ОБСУЖДЕНИЕ**

Разработанный программный инструментарий позволяет получать только самые базовые сведения о публикациях: название статьи, список соавторов и сведения об источнике, включая номер тома и выпуска (для журналов), а также диапазоны страниц размещения статьи в источнике. В скрипте не реализовано извлечение количества цитирований статей, но при необходимости это может быть выполнено добавлением скрипта, аналогичного представленному в листинге 2 на базе опорного тега `<td align="center" valign="middle" class="select-tr-right">`. Кроме этого, в случае, когда в публикации много соавторов, то их список в соответствующей строке электронной таблицы будет урезанным.

При изменении состава опорных тегов или их атрибутов, исходный код разработанных скриптов может претерпеть незначительные изменения, но общие принципы реализации останутся нетронутыми. Однако программный инструментарий может стать непригодным для использования без значительной модификации в случае изменении платформы, на основе которой работает портал [elibrary.ru](http://elibrary.ru).

Код разработанных скриптов можно сократить за счет еще большего уровня использования функционального программирования в форме генератора списков [21]. Использование Selenium WebDriver позволило адаптировать программу под работу в режиме онлайн, что устранило необходимость вручную длительно сохранять веб-страницы с результатами поиска статей.

При помощи разработанных инструментов нет возможности получить такие важные характеристики публикаций как DOI, ISSN журналов и списки аффилиаций авторов. Если у научной организации есть лицензия на использования системы Science Index, то для получения таких характеристик необходимо использовать алгоритм и инструменты, разработанные в исследовании [12].

Разработанный программный инструментарий при незначительной доработке можно также использовать для целей преобразования содержимого веб-страниц в электронные таблицы для решения следующих задач при работе на портале [elibrary.ru](http://elibrary.ru):

- обработка результатов расширенного поиска;
- синтаксического анализа подборок публикаций;
- анализ списка публикаций авторов;
- анализ информации об авторах научной организации.

## **ЗАКЛЮЧЕНИЕ**

В статье были проанализированы методы извлечения сведений из веб-страниц с выходными данными публикаций научной организации на портале [elibrary.ru](http://elibrary.ru) в формат электронной таблицы, были кратко рассмотрены особенности их использования, а также их достоинства и недостатки. В качестве реализации одного из методов был разработан автоэкспортер контента веб-страниц из [elibrary.ru](http://elibrary.ru) в html-файлы, а также синтаксический анализатор, позволяющий решить задачу извлечения и сохранения содержимого выгруженных файлов в формате Excel. Исходный код анализаторы можно легко скомпилировать в исполняемый файл Windows посредством инструментов библиотеки Pyinstaller (команда `"python -m PyInstaller --onefile [имя файла].py"` или

запускать посредством самого интерпретатора Python на операционных системах семейства Linux при условии наличия в папке со скриптом каталога “\_el” с набором html-файлов. Использование этого программного инструментария позволяет сократить минимум на 5 минут время преобразования содержимого веб-страниц в электронную таблицу на каждые 100 публикаций результатов фильтрации и получить данные в форме, пригодной для дальнейшей обработки и анализа.

### Список литературы

1. Бабынина Л.С., Грунина И.С. Программа "Приоритет-2030" как развитие концептуальных под-ходов в оценке деятельности вузов // KANT. 2024. №1 (50). С. 4-11. EDN: AAMGCU. DOI: 10.24923/2222-243X.2024-50.1 URL: [https://www.elibrary.ru/download/elibrary\\_65310545\\_87336556.pdf](https://www.elibrary.ru/download/elibrary_65310545_87336556.pdf) (дата обращения: 03.04.2025)
2. Формирование инновационной образовательной системы в НИУ «БелГУ» как основы подготовки кадров мирового уровня на базе НОЦ «Инновационные решения в АПК» / О.Н. Полухин, А.В. Маматов, И.В. Спичак и др. // Достижения науки и техники АПК. 2020. Т. 34. № 9. С. 9–13. DOI: 10.24411/0235-2451-2020-10902. URL: [https://www.elibrary.ru/download/elibrary\\_44095025\\_41119820.pdf](https://www.elibrary.ru/download/elibrary_44095025_41119820.pdf) (дата обращения: 03.04.2025)
3. Kochetkov D.M. Russian Journal Whitelist: Questions to be answered // Science Editor and Publisher. 2022. № 7(2). С. 185-190. URL: <https://doi.org/10.24069/SEP-22-48> (дата обращения: 03.04.2025)
4. Moed H.F., Markusova V., Akoev M. Trends in Russian research output indexed in Scopus and Web of Science // Scientometrics. 2018. № 116. С. 1153-1180. URL: <https://doi.org/10.1007/s11192-018-2769-8> (дата обращения: 03.04.2025)
5. О проекте eLIBRARY.RU. Научная электронная библиотека eLIBRARY.RU, 2025. URL: [https://elibrary.ru/elibrary\\_about.asp](https://elibrary.ru/elibrary_about.asp) (дата обращения: 03.04.2025)
6. Российский индекс научного цитирования. Научная электронная библиотека eLIBRARY.RU, 2025. URL: [https://elibrary.ru/projects/citation/cit\\_index.asp](https://elibrary.ru/projects/citation/cit_index.asp) (дата обращения: 03.04.2025)
7. О проекте Russian Science Citation Index. Научная электронная библиотека eLIBRARY.RU, 2025. URL: [https://www.elibrary.ru/rsci\\_about.asp?](https://www.elibrary.ru/rsci_about.asp?) (дата обращения: 03.04.2025)
8. Перечень рецензируемых научных изданий, в которых должны быть опубликованы основные научные результаты диссертаций на соискание ученой степени кандидата наук, на соискание ученой степени доктора наук (по состоянию на 10.06.2024 года), 2024. URL: <https://vak.minobrnauki.gov.ru/uploader/loader?type=19&name=91107547002&f=23267> (дата обращения: 03.04.2025)
9. Интерфейс программирования приложений API. Научная электронная библиотека, 2025. URL: [https://www.elibrary.ru/projects/api/api\\_info.asp](https://www.elibrary.ru/projects/api/api_info.asp) (дата обращения: 03.04.2025)
10. Информационно-аналитическая система SCIENCE INDEX. Научная электронная библиотека, 2025. URL: [https://www.elibrary.ru/projects/science\\_index/science\\_index\\_org\\_info.asp?](https://www.elibrary.ru/projects/science_index/science_index_org_info.asp?) (дата обращения: 03.04.2025)
11. Гинтофт А.С., Новгородов П.А., Коренев А.Н. Интеграционный модуль импорта наукометрических данных из научной электронной библиотеки. Свидетельство о государственной регистрации программы для ЭВМ № 2021663072 от 04.08.2021. URL: <https://www.elibrary.ru/item.asp?id=46484977> (дата обращения: 03.04.2025)
12. Резниченко О.С. Алгоритмы и инструменты обработки данных о публикациях научных организаций, размещенных на портале eLIBRARY.RU // Экономика. Информатика. 2025. № 52(1). С. 181-193. DOI: 10.52575/2687-0932-2025-52-1-181-193.
13. Парсинг электронных ресурсов. Библиотека selenium или fake useragent? / А.Ю. Попов, М.В. Ремез, Е.В. Жилина, М.И. Ожиганова // Информатизация в цифровой экономике. 2022. Т. 3. № 4. С. 197-210. DOI: 10.18334/ide.3.4.115219
14. Запекин С., Шитов В., Заваруев И. Парсер научных журналов eLibrary // GitHub. 2021. URL: <https://github.com/Lfdd/Parser> (дата обращения: 03.04.2025)
15. Supported Browsers. Selenium, 2025. URL: <https://www.selenium.dev/documentation/webdriver/browsers/> (дата обращения: 03.04.2025)
16. van Rossum Guido at al. Python, 2025. URL: <https://www.python.org> (дата обращения: 03.04.2025)
17. JetBrains. PyCharm: The Python IDE for data and web professionals, 2025. URL: <https://www.jetbrains.com/pycharm/> (дата обращения: 03.04.2025)

18. Van den Bossche J. at al. pandas.DataFrame, 2024. URL: <https://pandas.pydata.org/pandas-docs/stable/reference/api/pandas.DataFrame.html> (дата обращения: 03.04.2025)
19. Richardson L. Beautiful Soup Documentation, 2015. URL: <https://beautiful-soup.readthedocs.io/en/latest/> (дата обращения: 03.04.2025)
20. Cortesi D. PyInstaller Manual, 2025. URL: <https://pyinstaller.org/en/stable/index.html> (дата обращения: 03.04.2025)
21. van Rossum G. at al. 5.1.3. List Comprehensions. Python, 2025. URL: <https://docs.python.org/3/tutorial/datastructures.html#list-comprehensions> (дата обращения: 03.04.2025)
22. Reznichenko O. Appendix to article "Methods and Software Tools for Uploading Publication Data of a Scientific Organization Posted on eLIBRARY.RU" // GitHub. 2025. URL: [https://github.com/leo-phoenix/elibrary\\_html\\_to\\_xls](https://github.com/leo-phoenix/elibrary_html_to_xls) (дата обращения: 03.04.2025)

## References

1. Babynina L.S., Grunina I.S. Priority-2030 Program as A Development of Conceptual Approaches in The Assessment of Higher Education Institutions' Performance. KANT. 2024. No. 1(50). P. 4-11. EDN: AAMGCY. DOI: 10.24923/2222-243X.2024-50.1 URL: [https://www.elibrary.ru/download/elibrary\\_65310545\\_87336556.pdf](https://www.elibrary.ru/download/elibrary_65310545_87336556.pdf) (accessed: 03 April 2025)
2. Polukhin O.N., Mamatov A.V., Spichak I.V., et al. Formation of an innovative educational system at the Belgorod State National Research University as the basis for training world-class personnel at the Scientific and Educational Center "Innovative Solutions in the Agro-Industrial Complex" in the Belgorod region. Dostizheniya nauki i tekhniki APK. 2020. No. 34(9). P. 9-13. DOI: 10.24411/0235-2451-2020-10902. URL: [https://www.elibrary.ru/download/elibrary\\_44095025\\_15844399.pdf](https://www.elibrary.ru/download/elibrary_44095025_15844399.pdf) (accessed: 03 April 2025)
3. Kochetkov D.M. Russian Journal Whitelist: Questions to be answered. Science Editor and Publisher. 2022. No. 7(2). P. 185-190. URL: <https://doi.org/10.24069/SEP-22-48> (accessed: 03 April 2025)
4. Moed H.F., Markusova V., Akoev M. Trends in Russian research output indexed in Scopus and Web of Science // Scientometrics. 2018. No. 116. P. 1153-1180. URL: <https://doi.org/10.1007/s11192-018-2769-8> (accessed: 03 April 2025)
5. About eLIBRARY.RU project. Scientific Electronic Library. 2025. URL: <https://elibrary.ru/defaultx.asp> (accessed: 03 April 2025)
6. Russian Science Citation Index. Scientific Electronic Library. 2025. URL: [https://elibrary.ru/projects/citation/cit\\_index.asp](https://elibrary.ru/projects/citation/cit_index.asp) (accessed: 03 April 2025)
7. About RSCI project. Scientific Electronic Library 2025. URL: [https://www.elibrary.ru/rsci\\_about.asp](https://www.elibrary.ru/rsci_about.asp) (accessed: 03 April 2025)
8. List of peer-reviewed journals in which the principal scientific findings of dissertations for the degrees of Candidate of Sciences and Doctor of Sciences must be published (as of 10.06.2024). 2024. URL: <https://vak.minobrnauki.gov.ru/uploader/loader?type=19&name=91107547002&f=23267> (accessed: 03 April 2025)
9. Application Programming Interface API. Scientific Electronic Library. 2025. URL: [https://www.elibrary.ru/projects/api/api\\_info.asp](https://www.elibrary.ru/projects/api/api_info.asp) (accessed: 03 April 2025)
10. Science Index. Interface for a company representative. Scientific Electronic Library. 2025. URL: [https://www.elibrary.ru/projects/science\\_index/science\\_index\\_org\\_info.asp](https://www.elibrary.ru/projects/science_index/science_index_org_info.asp)? (accessed: 03 April 2025)
11. Gintoft A.S., Novgorodov P.A., Korenev A.N. Integration Module for Importing Scientific Metric Data from a Scientific Electronic Library. Certificate of state registration of a computer program. 2021. No. RU 2021663072 at 04.08.2021. URL: <https://www.elibrary.ru/item.asp?id=46484977> (accessed: 03 April 2025)
12. Reznichenko O.S. Algorithms and Tools for Processing Data on Scientific Organizations' Articles Uploaded to eLIBRARY.RU. Economics. Information Technologies. 2025. No. 52(1). P. 181-193). DOI: 10.52575/2687-0932-2025-52-1-181-193
13. Popov A.Y., Remez M.V., Zhilina E.V., Ozhiganova M.I. Parsing of electronic resources. Selenium library or fake useragent? Informatization in the Digital Economy. 2022. No. 3(4), P. 197-210. URL: <https://doi.org/10.18334/ide.3.4.115219> (accessed: 03 April 2025)
14. Zapekin S., Shitov V., Zavaruev I. eLibrary Scientific Journal Parser // GitHub. 2021. URL: <https://github.com/Lfdd/Parser> (accessed: 03 April 2025)
15. Supported Browsers. 2025. URL: <https://www.selenium.dev/documentation/webdriver/browsers/> (accessed: 03 April 2025)
16. van Rossum G. at al. Python. 2025. URL: <https://www.python.org> (accessed: 03 April 2025)

17. JetBrains. PyCharm: The Python IDE for Professional Developers. 2025. URL: <https://www.jetbrains.com/pycharm/> (accessed: 03 April 2025)
18. Van den Bossche J. et al. pandas.DataFrame. 2024. URL: <https://pandas.pydata.org/pandas-docs/stable/reference/api/pandas.DataFrame.html> (accessed: 03 April 2025)
19. Richardson L. Beautiful Soup Documentation. 2015. URL: <https://beautiful-soup.readthedocs.io/en/latest/> (accessed: 03 April 2025)
20. Cortesi D. PyInstaller Manual. 2025. URL: <https://pyinstaller.org/en/stable/index.html> (accessed: 03 April 2025)
21. van Rossum G. et al. 5.1.3. List Comprehensions. 2025. URL: <https://docs.python.org/3/tutorial/datastructures.html#list-comprehensions> (accessed: 03 April 2025)
22. Reznichenko O. Appendix to article "Methods and Software Tools for Uploading Publication Data of a Scientific Organization Posted on eLIBRARY.RU" // GitHub. 2025. URL: [https://github.com/leo-phoenix/elibrary\\_html\\_to\\_xls](https://github.com/leo-phoenix/elibrary_html_to_xls) (accessed: 03 April 2025)

**Резниченко Олег Сергеевич**, старший преподаватель кафедры прикладной информатики и информационных технологий, Белгородский государственный национальный исследовательский университет, г. Белгород, Россия

**Reznichenko Oleg Sergeevich**, Senior lecturer of the Department of Applied Informatics and Information Technology, Belgorod State National Research University, Belgorod, Russia

УДК 004.85

DOI: 10.18413/2518-1092-2025-10-3-0-2

Назаров Т.Р.

**ФОРМИРОВАНИЕ ИНВЕСТИЦИОННЫХ ПОРТФЕЛЕЙ  
С ИСПОЛЬЗОВАНИЕМ КЛАСТЕРНОГО АНАЛИЗА**

Российский экономический университет имени Г.В. Плеханова,  
Стремянный переулок, 36, г. Москва, 115054, Россия

*e-mail: TimNazarovya@gmail.com*

**Аннотация**

Исследование посвящено проверке гипотезы о возможности создать диверсифицированный портфель акций путем их кластеризации по показателям риска и устойчивости. В рамках данной проблемы использовались три метода кластеризации: метод К-средних, модель гауссовой смеси и иерархическая кластеризация методом Уорда. Также было отобрано три признака, по которым проводилась кластеризация акций: индекс Бета, волатильность и ROE. По данным с 2018 по 2023 год за каждый период была проведена кластеризация с подбором оптимального количества кластеров, и были созданы портфели финансовых активов по заранее определенным правилам. В качестве альтернативы за каждый год были созданы портфели акций без предварительной кластеризации по правилам портфельной теории Марковица. После этого каждый портфель был протестирован в течение следующего года. Результаты на основе метрик доходности и риска показывают, что гипотеза имеет право на существование. Портфели, созданные с предварительной кластеризацией, превзошли по доходности классические варианты. Проблема имеет простор для дальнейшего исследования. В частности, иной набор признаков для сегментации акций или применение методов кластеризации в сочетании с другими алгоритмами машинного обучения или нейронных сетей может также дать качественные результаты.

**Ключевые слова:** кластеризация; акции; волатильность; Индекс Бета; К-средних; иерархическая кластеризация; метод гауссовых смесей; портфельная теория Марковица

**Для цитирования:** Назаров Т.Р. Формирование инвестиционных портфелей с использованием кластерного анализа // Научный результат. Информационные технологии. – Т.10, №3, 2025. – С. 20-30. DOI: 10.18413/2518-1092-2025-10-3-0-2

Nazarov T.R.

**FORMATION OF INVESTMENT PORTFOLIOS USING  
CLUSTER ANALYSIS**

Plekhanov Russian University of Economics,  
36 Stremyanny Lane, Moscow, 115054, Russia

*e-mail: TimNazarovya@gmail.com*

**Abstract**

The study is devoted to testing the hypothesis about the possibility of creating a diversified portfolio of stocks by clustering them according to risk and sustainability indicators. Within the framework of this problem, three clustering methods were used: the K-means method, the Gaussian mixture model, and hierarchical clustering by the Ward method. We also selected three criteria for stock clustering: Beta index, volatility, and ROE. According to the data from 2018 to 2023, clusterization was carried out for each period with the selection of the optimal number of clusters, and portfolios of financial assets were created according to predefined rules. Alternatively, stock portfolios were created for each year without prior clustering according to the rules of Markowitz portfolio theory. After that, each portfolio was tested over the next year. The results based on the metrics of profitability and risk show that the hypothesis has a right to exist. Portfolios created with pre-clustering outperformed the classic options in terms of profitability. The problem has scope for further investigation. In particular, a different set of features for stock segmentation or the use of

clustering methods in combination with other machine learning or neural network algorithms can also produce high-quality results.

**Keywords:** clustering; stocks; volatility; Beta index; K-means; hierarchical clustering; Gaussian mixture method; Markowitz portfolio theory

**For citation:** Nazarov T.R. Formation of Investment Portfolios Using Cluster Analysis // Research result. Information technologies. – Т.10, №3, 2025. – P. 20-30. DOI: 10.18413/2518-1092-2025-10-3-0-2

## **ВВЕДЕНИЕ**

На конец 4 квартала 2024 года, более трети экономически активного населения имеет брокерские счета и по итогам года клиентская база брокеров выросла на 36% за счет новичков [1]. С помощью рынка ценных бумаг можно получать хорошую прибыль, но для этого необходимо эффективно управлять инвестиционным портфелем, анализировать динамику фондового рынка, рассчитывать уровень доходности и просчитывать возможные риски. Из-за непредсказуемой ситуации на рынке, а также предубеждений, часто наблюдаемых в поведении людей, связанных с инвестированием, и нелогичности принятия решений, создание успешных портфелей финансовых активов и управление ими является сложной задачей. Достижение высокой доходности при низком риске – идеальная, но кажущаяся невозможной задача. Этот процесс называется процессом диверсификации. Современная портфельная теория утверждает, что диверсификация активов является наиболее эффективным способом получения портфеля активов с низкими показателями риска и относительно высокими показателями прибыли. На данный момент было представлено множество методов создания портфеля акций, из которых наиболее известным является метод Марковица, который был предложен в 1951 году [2].

Кластеризация – один из наиболее важных методов интеллектуального анализа данных для извлечения полезной информации из массивов данных большой размерности. Кластеризация акций может помочь при построении диверсифицированного портфеля ценных бумаг. Например, путем распределения капитала между различными кластерами акций инвестор может снизить общий уровень риска портфеля. Если один сектор или регион испытывает трудности, другие кластеры могут остаться стабильными или даже вырасти, что помогает сгладить потери.

Данная проблема не нова, и ей посвящено много исследований. Исследования подтверждают, что с помощью различных методов кластеризации возможно получить устойчивые группы акций [3]. Портфели, созданные с помощью сравнения временных рядов [12, 13] и анализа динамических показателей [4, 11] акций, демонстрировали высокую доходность и умеренный риск. Однако, в упомянутых исследованиях портфели не сравнивались с оптимизированными портфелями, созданными без предварительной кластеризации акций. Данное исследование будет рассматривать в качестве показателей как динамические показатели акций, так и финансовые показатели эмитентов акций, а показатели доходности и риска портфелей будет сравниваться с классическими вариантами.

Целью данной работы является проверка гипотезы о том, что кластеризация акций по определенным показателям риска и доходности может помочь создать диверсифицированный портфель акций, который может превзойти классические альтернативы по доходности, при этом представляя меньший риск для инвестора.

## **МАТЕРИАЛЫ И МЕТОДЫ ИССЛЕДОВАНИЯ**

За основу брались акции, составляющие в период с 2018 по 2023 год индекс широкого рынка на Московской Бирже. Индекс отражает общее состояние динамики рынка акций. Фондовые индексы имеют весомую роль при оценке конъюнктуры фондового рынка, позволяют анализировать его состояние в прошлые периоды времени, выявлять тенденции развития, а также делать прогнозы на будущее. Индекс акций широкого рынка Московской Биржи включает в себя 100 ценных бумаг, отобранных на основании критериев ликвидности, капитализации и доли,

находящейся в свободном обращении, и является основой для формирования баз расчета остальных индексов Московской Биржи [5]. С учетом полноты доступных данных, распределение количества акций компаний, доступных для добавления в портфель по годам, представлено в Таблице 1:

Таблица 1

Количество доступных акций в рамках каждого года

Table 1

Number of shares available within each year

Год	Количество акций
2018	81
2019	85
2020	78
2021	71
2022	73
2023	78

Сравнение портфелей будет происходить исключительно в рамках каждого года, поэтому различное количество доступных акций от года к году не является критичным для проведения исследования. Для проведения кластерного анализа акций были выбраны следующие показатели:

– мультипликатор ROE (return on equity). Данный мультипликатор измеряет прибыльность компании путем сравнения ее чистой прибыли с собственным капиталом. Показывает, насколько эффективно компания генерирует прибыль за счет своего собственного капитала. Более высокий показатель рентабельности инвестиций, как правило, желателен, поскольку это говорит о том, что компания более эффективно использует свой собственный капитал для получения прибыли.

– Коэффициент Бета. Отражает рискованность вложения по отношению к рынку и показывает чувствительность изменения доходности акции к изменению доходности рынка [6]. Рассчитывается по следующей формуле:

$$\beta_a = \frac{cov(r_a, r_p)}{var(r_p)}, \quad (1)$$

где  $r_a$  – доходность акции, а  $r_p$  – доходность рынка.

В качестве актива, по которому оценивается доходность рынка, рассматривался индекс Московской Биржи;

– волатильность в процентах. Волатильность, в контексте акций, рассчитывается как стандартное отклонение изменения цены за период времени. Акции с более высокой волатильностью, как правило, являются более рискованными инвестициями, поскольку их цены с течением времени колеблются более резко.

Эти признаки вместе охватывают разные аспекты риска и доходности, что делает их хорошей базой для кластеризации: фундаментальные свойства (ROE), рыночная чувствительность (Beta), внутренняя нестабильность (волатильность). Аналогичные переменные традиционно используются в работах по факторному инвестированию [15,16], а также при построении многомерных моделей риска. Пример наблюдений из набора данных представлен на Рисунке 1:

	Год	Ticker	ROE, %	Beta	Volatility
0	2018	AFKS	6.5	0.296480	1.635736
1	2018	AFLT	-9.0	0.163191	2.001954
2	2018	AGRO	12.6	0.288014	1.217860
3	2018	AKRN	22.2	0.409532	0.897271
4	2018	ALRS	36.5	0.317364	1.698126

Рис. 1. Пример наблюдений из набора данных  
Fig. 1. Example of observations from the dataset

Колонка Ticker содержит краткое название котируемого на Московской Бирже инструмента. Для отображения многомерных данных в двумерном пространстве используется метод многомерного шкалирования. В общем случае метод позволяет таким образом расположить объекты в пространстве некоторой небольшой размерности, чтобы достаточно адекватно воспроизвести наблюдаемые расстояния между ними. Сходство между объектами обычно представляется в виде матрицы расстояний. Эта матрица показывает расстояния между каждой парой объектов в наборе данных. Отображение данных по годам представлено на Рисунке 2:

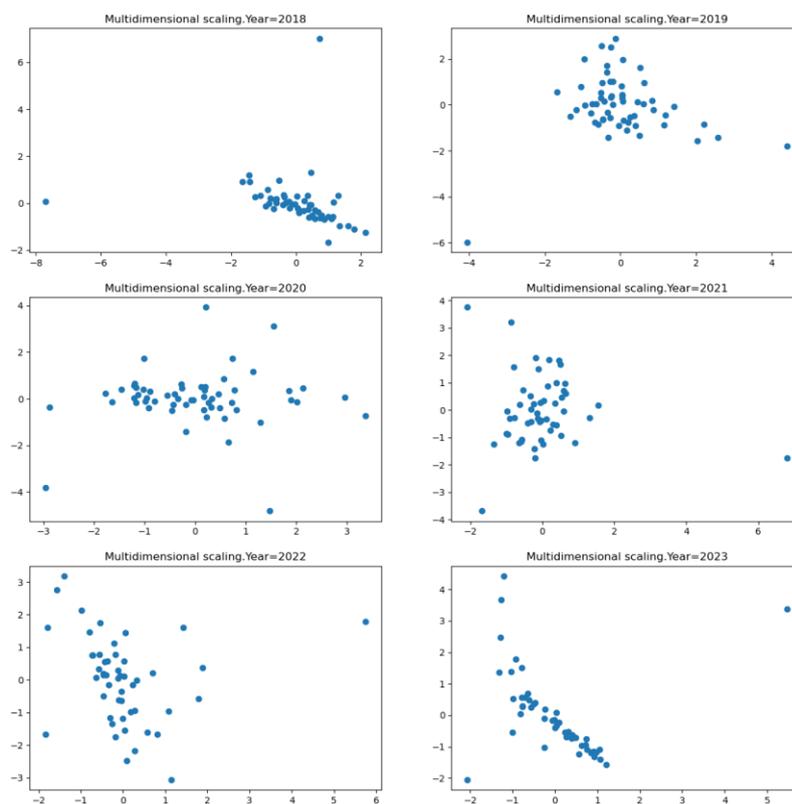


Рис. 2. Представление данных по годам  
Fig. 2. Data presentation by year

Портфельные инвестиции – это вложение денежных средств в акции, облигации и другие ценные бумаги, выпущенные на рынок различными инвесторами (государством, корпорациями и частными компаниями), а портфель ценных бумаг – набор ценных бумаг, обеспечивающий для инвестора удовлетворительные количественно-качественные характеристики финансовых инструментов. Уменьшение риска вложений, как правило, достигается за счет диверсификации

(разделения) портфеля, то есть невысокие доходы по одним ценным бумагам компенсируются высокодоходными. На практике используют множество методик формирования оптимальной структуры портфеля ценных бумаг. Большинство из них основано на методике Марковица. Он впервые предложил математическую формализацию задачи нахождения оптимальной структуры портфеля ценных бумаг в 1951 году, за что позднее был удостоен Нобелевской по экономике.

Портфельная теория Марковица является многокритериальной оптимизационной задачей, в которой ищутся такие значения распределений долей активов в портфеле, которые одновременно обеспечивают максимально возможное значение ожидаемой доходности портфеля активов при минимально возможном значении риска портфеля активов (среднего стандартного отклонения) [7].

Ожидаемая доходность портфеля максимизируется следующим образом:

$$m_p = W^T * M \rightarrow \max. \quad (2)$$

Риск портфеля минимизируется следующим образом:

$$S_p^2 = W^T * V * W \rightarrow \min. \quad (3)$$

При ограничениях:

$$\sum_{i=1}^n W_i = 1 \quad (4)$$

$$w_i \geq 0 \forall i = 1 \dots n, \quad (5)$$

где  $W_{[n \times 1]}$  – вектор искомых долей активов в портфеле,  $M_{[n \times 1]}$  – вектор ожидаемых доходностей активов, отбираемых в портфель,  $V_{[n \times n]}$  – матрица ковариаций доходности активов, а  $n$  – количество активов. Предполагается, что найденные в результате решения задачи, а именно веса активов, будут оптимальными и в следующих торгах. Поэтому, данная задача одновременно является задачей прогнозирования.

Кластеризация или кластерный анализ – это метод машинного обучения, который обрабатывает немаркированный набор данных и относится к методам обучения без учителя. Его цель – группировка объектов на подмножества или “кластеры” таким образом, чтобы объекты внутри каждого кластера были более тесно связаны друг с другом, чем объекты, отнесенные к другим кластерам. Формально, задачу кластеризации можно поставить следующим образом [14]: пусть  $X$  – совокупность наблюдений,  $Y$  – множество меток кластеров. На множестве  $X$  задана функция, описывающая расстояния между объектами  $\rho(x, x')$ . Также представлена обучающая выборка  $X^m = \{x^1, \dots, x^m\}$ . Необходимо разбить выборку на подмножества (кластеры), то есть каждому объекту  $x_i \in X^m$  сопоставить метку  $y_i \in Y$ , таким образом, чтобы объекты внутри каждого кластера были близки относительно метрики  $\rho$ , а объекты из разных кластеров значительно отличались относительно метрики  $\rho$ . Таким образом, алгоритм кластеризации – функция  $a: X \rightarrow Y$ , которая любому объекту  $x \in X$  ставит в соответствие идентификатор кластера  $y \in Y$ . Множество  $Y$  в некоторых случаях известно заранее, однако чаще ставится задача определить оптимальное число кластеров, с точки зрения того или иного критерия качества кластеризации.

Для кластеризации акций использовались три различных метода кластеризации: метод  $k$ -средних, метод гауссовых смесей и метод иерархической кластеризации Уорда. Алгоритм  $K$ -средних является одним из самых популярных алгоритмов машинного обучения. Целью алгоритма является разбиение набора объектов на  $K$  кластеров таким образом, чтобы сумма квадратов расстояний между объектами и назначенным им средним значением по кластеру была сведена к минимуму. Каждому кластеру соответствует точка (центроид) в пространстве данных, которая является средним значением присвоенных ему наблюдений. Метод иерархической кластеризации Уорда, как следует из названия, представляет собой алгоритм, который строит иерархию кластеров. Работа алгоритмов иерархической кластеризации начинается с обработки каждого наблюдения как отдельного кластера. Затем последовательно выполняются следующие два

шага: идентифицируются два наиболее близких друг к другу кластера и объединяются два наиболее похожих кластера. Этот итеративный процесс продолжается до тех пор, пока все кластеры не будут объединены вместе или не будет достигнут критерий остановки. В методе Уорда кластеры объединяются таким образом, чтобы минимизировать увеличение общей внутрикластерной дисперсии при объединении двух кластеров. Для определения оптимального значения порога расстояния, следует рассмотреть результаты, полученные при разных показателях и сделать выбор, основываясь на метриках качества кластеризации [8]. Метод Гауссовских смесей представляет собой алгоритм нечеткой кластеризации, где принадлежность каждого элемента тому или иному кластеру выражается в виде вероятности. Алгоритм предполагает, что точки данных внутри каждого кластера распределяются в соответствии с гауссовым распределением.

Для оценки схожести двух объектов использовалась евклидова метрика. Для двух точек в двумерном пространстве она вычисляется как квадратный корень из суммы разностей между соответствующими координатами двумя различными наблюдениями:

$$d = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}. \quad (6)$$

Евклидово расстояния хорошо проходит для работы с числовыми типами переменных.

Для оценки качества кластеризации и выбора оптимального количества кластеров использовался коэффициент силуэта. Коэффициент силуэта показывает, насколько правильно распределена каждая отдельная точка. Если коэффициент ближе к 0, то можно считать, что она находится на границе между двумя кластерами. Если он ближе к -1, то можно считать, что она неверно отнесена к определенному кластеру. Если коэффициент силуэта ближе к 1, то точка правильно назначена и может быть интерпретирована как принадлежащая к подходящему кластеру. Это очень интуитивно понятный способ оценки качества модели кластеризации. Данный показатель для всей выборки можно рассчитать следующим образом:

$$S = \frac{ncd - icd}{\max(icd, ncd)}, \quad (7)$$

где

$icd$  – среднее внутрикластерное расстояние, то есть среднее расстояние между всеми членами кластера. Оно позволяет оценить, насколько плотно наблюдения расположены внутри каждого кластера.

$ncd$  – среднее межкластерное расстояние, то есть среднее расстояние между каждым наблюдением в рассматриваемом кластере и наблюдениями в других кластерах.

Эмпирически установленное оптимальное количество кластеров в рамках каждого года, в зависимости от используемого метода кластеризации, представлено в Таблице 2:

Таблица 2

Оптимальное количество кластеров в зависимости от метода кластеризации и года

Table 2

Optimal number of clusters depending on clustering method and year

Метод	2018	2019	2020	2021	2022	2023
К-средних (K-means)	3	4	3	3	4	6
Метод Гауссовских смесей (GMM)	3	3	3	3	3	6
Иерархическая кластеризация Уорда (Ward)	4	4	5	4	4	6

Построение портфелей происходило по следующим правилам:

- вычислялось количество элементов, которое будет набрано из каждого кластера. Оно равнялось частному от деления нацело размера портфеля на количество кластеров. Если размер кластера меньше полученного числа, то из кластера брались все элементы;

- при применении алгоритма K-средних элементы из каждого кластера отбирались по убыванию расстояния от центроида своего кластера. Чем ближе элемент к центроиду – тем лучше он характеризует получившийся кластер. При применении метода Гауссовых смесей элементы из каждого кластера отбирались по убыванию вероятности принадлежности к своему кластеру. Чем выше вероятность – тем лучше элемент характеризует получившийся кластер. При применении метода иерархической кластеризации Уорда внутри каждого кластера находится медоид. Медоид – существующий элемент кластера, суммарное расстояние до которого от других элементов кластера минимально [9]. Элементы из кластера выбираются в портфель на основании расстояния до медоида. Чем меньше расстояние до медоида, тем лучше элемент характеризует получившийся кластер;

- остаток, если он возникает, добирался без привязки к кластеру, но по тому же правилу.

После отбора акций в портфели, внутри каждого портфеля для каждого актива назначался вес на основе портфельной теории Марковица с максимизацией индекса Шарпа и на основе данных о ежедневной динамике цен по каждому коду ценной бумаги, попавшей в портфель. Инвесторы часто используют коэффициент Шарпа для сравнения эффективности различных портфелей или инвестиционных стратегий. Он помогает инвесторам оценить, оправдывает ли дополнительная доходность, получаемая от портфеля, дополнительный риск, который они берут на себя. Вычисляется он по следующей формуле [10]:

$$SR = \frac{r_p - r_f}{\sigma_p}, \quad (8)$$

где  $r_p$  – средняя доходность портфеля,  $r_f$  – средняя безрисковая доходность, а  $\sigma_p$  – стандартное отклонение доходностей активов. Вес каждой акции в портфеле ограничивался диапазоном от 1 до 25%. Сумма, на которую формировался портфель – 6 миллионов 500 тысяч рублей. Акции покупались по цене на момент открытия биржи в первый торговый день следующего года. Зачастую вся сумма распределялась не полностью. В таких случаях остаток распределялся итеративно между акциями по убыванию веса.

### **ТЕСТИРОВАНИЕ ПОРТФЕЛЕЙ И АНАЛИЗ РЕЗУЛЬТАТОВ**

Портфели сравнивались по годам. Так, портфели, созданные на основе данных за рассматриваемый год, в рамках эксперимента оценивались на основе данных, полученных за следующий год. Например, портфели, созданные на основе данных за 2018 год, оценивались по их показателям за 2019 год. В процессе оценки поведения портфеля не использовалась ребалансировка. Веса оставались неизменными, это дало возможность оценить качество диверсификации, полученной с помощью предварительной кластеризации.

Также было сформировано 5 портфелей без применения кластерного анализа. Они формировались следующим образом:

- в рамках каждого года у каждой акции брались данные по изменению цены за период с июня по декабрь. Всем акциям назначались веса, а потом отбирались 16 акций с наиболее высокими весами;

- отобранным бумагам назначались итоговые веса по тому же алгоритму в диапазоне от 0.01 до 0.25;

- акции приобретались по цене на момент открытия в первый торговый день следующего года на тот же бюджет.

Для сравнения портфелей были выбраны следующие метрики:

– доходность портфеля (в процентах) – отражает прирост или просадку портфеля в процентах по прошествию периода времени. Вычисляется по следующей формуле:

$$D = \frac{V_H - V_K}{V_K} * 100, \quad (9)$$

где  $V_H$  – начальный объем портфеля,  $V_K$  – конечный объем портфеля;

– отношение среднедневной доходности в течение года к годовой волатильности – данный коэффициент позволяет оценить эффективность портфеля с точки зрения соотношения дохода к риску. Более высокое значение может указывать на то, что портфель активов генерируют большую доходность при том же уровне риска:

$$RV = \frac{rp}{\sigma p}, \quad (10)$$

где  $rp$  – среднегодовая доходность портфеля, а  $\sigma p$  – стандартное отклонение доходностей портфеля.

Результаты, полученные в ходе эксперимента, представлены в Таблице 3:

Таблица 3

Результаты тестирования портфелей ценных бумаг

Table 3

Results of securities portfolios testing

Метод	Год	Доходность, %	RV
K-means	2018	+16.5%	0.09
GMM	2018	+6.7%	0.04
Ward	2018	+12.6%	0.07
Classic	2018	+3.56%	0.03
K-means	2019	+53.37%	0.119
GMM	2019	+10.1%	0.03
Ward	2019	+24.3%	0.06
Classic	2019	+40.9%	0.105
K-means	2020	+61.8%	0.09
GMM	2020	+52.47%	0.15
Ward	2020	+80.11%	0.09
Classic	2020	+39.68%	0.12
K-means	2021	-10%	-0.01
GMM	2021	-37.6%	-0.05
Ward	2021	-21.5%	-0.02
Classic	2021	-10.76%	-0.007
K-means	2022	+29.1%	0.11
GMM	2022	+67.52%	0.15
Ward	2022	+28.14%	0.103
Classic	2022	+46%	0.09
K-means	2023	-7.84%	-0.02
GMM	2023	-14%	-0.03
Ward	2023	-9%	-0.018
Classic	2023	-16.24%	0.06

При создании портфелей на основе данных за 2018, 2019, 2020, 2022 год, портфели, созданные с помощью того или иного метода кластеризации, превзошли классический портфель как по общей

доходности, так и по отношению среднедневной доходности к годовой волатильности. Портфель, созданный на основе данных за 2021 год с помощью метода кластеризации К-средних, потерял в 2022 году чуть меньше, чем классический портфель, но, по сравнению с классическим портфелем, обладает худшим отношением среднегодовой доходности к годовой волатильности. Все портфели, созданные с помощью предварительной кластеризации на основе данных за 2023 год, в 2024 году потеряли меньше, чем классический портфель, но последний показал положительное отношение среднедневной доходности к годовой волатильности.

### **ЗАКЛЮЧЕНИЕ**

Таким образом, в ходе исследования были собраны данные по акциям, составляющим индекс широкого рынка Мосбиржи в период с 2018 по 2023 год включительно. Данные включали в себя мультипликатор ROE, индекс Beta и годовую волатильность. В рамках каждого года была произведена кластеризация акций тремя различными методами и отобраны акции для составления трех портфелей. Акциям назначался вес согласно портфельной теории Марковица. Портфели сравнивались по общей доходности и отношению среднедневной доходности в течение года к годовой волатильности. В дальнейшем портфели были протестированы на последующем году. По полученным данным можно сделать вывод, что гипотеза о том, что кластеризация акций по определенным показателям риска и доходности (не ограничиваясь стандартными показателями доходности и волатильности), может помочь создать диверсифицированный портфель акций, имеет право на существование. Портфели, созданные на основе предварительной кластеризации тем или иным методом, превзошли классические альтернативы. По показателю отношения среднедневной доходности к годовой волатильности их превосходство проявилось не так наглядно. Также по итогам эксперимента нельзя явно определить наилучший метод кластеризации. При дальнейшем рассмотрении данного вопроса возможно как пересмотреть состав признаков, так и рассмотреть возможность применения других методов кластеризации или методов кластеризации в сочетании с другими алгоритмами машинного обучения или нейронных сетей.

### **Список литературы**

1. Число физлиц с брокерскими счетами на МосБирже в 2024 году выросло на 5,4 млн // URL: <https://www.interfax.ru/business/1003027> (дата обращения: 26.05.2025).
2. Баюк Д., Баюк О., Берзин Д., Иванюк В. Практическое применение методов кластеризации, классификации и аппроксимации на основе нейронных сетей. – 1-е изд. – Москва: Прометей, 2020. – 448 с.
3. Дмитриев Н.Д. Кластеризация российских компаний технологического сектора на основе рыночных и финансовых показателей / Н.Д. Дмитриев, А.В. Тюлькова, В.В. Сорокожердьев // Вестник Челябинского государственного университета. – 2024. – № 10(492). – С. 116-127.
4. Тюхова Е.М. Использование кластерного анализа для формирования портфеля ценных бумаг в инвестиционных системах (робоэдвайзерах) / Е.М. Тюхова, Д.С. Сизых // Управление развитием крупномасштабных систем (MLSD'2019): Материалы двенадцатой международной конференции, Москва, 01–03 октября 2019 года / Под общей редакцией С.Н. Васильева, А.Д. Цвиркуна. – Москва: Международный научно-исследовательский институт проблем управления РАН, 2019. – С. 210-220.
5. Индекс широкого рынка // URL: <https://www.moex.com/ru/index/МОЕХВМІ/about> (дата обращения: 24.05.2024).
6. Жданов И.Ю. Инвестиционная оценка проектов и бизнеса: учебное пособие / И.Ю. Жданов, В.Ю. Жданов. – Москва: Проспект, 2019. – 120 с.
7. Борисова Л.В., Сагаева И.Д. Модели оптимального инвестирования: учебно-методическое пособие, Саратов // URL: [http://elibrary.sgu.ru/uch\\_lit/1347.pdf](http://elibrary.sgu.ru/uch_lit/1347.pdf) (дата обращения: 23.05.2025).
8. Кластеризация и частичное обучение // URL: <http://www.machinelearning.ru/wiki/images/5/52/Voron-ML-Clustering-SSL-slides.pdf> (дата обращения: 22.05.2025).
9. Тахир М.А., Ахмед А.У., Егорова Л.Д. Об алгоритмах для задачи К-медоид // Актуальные проблемы авиации и космонавтики. – 2021. – №1. – С. 194-196.

10. Деканова К.В., Гейт А.А. Методика формирования инвестиционного портфеля с учетом индивидуальных особенностей частного инвестора // Экономика: вчера, сегодня, завтра. 2023. – Том 13. – № 7А. – С. 311-329.
11. Nourahmadi M., Sadeqi H. Portfolio Diversification Based on Clustering Analysis // Iranian Journal of Accounting, Auditing and Finance. – 2023. – №7(3) – P. 1-16.
12. Szczepocki, Piotr Clustering companies listed on the Warsaw Stock Exchange according to time-varying beta // Econometrics. – 2023. – №23 – P. 63-79.
13. Lim T., Ng H. Financial Portfolio Management Based on Shaped-Based Unsupervised Machine Learning: A Dynamic Time Warping Baycenter Averaging Approach to International Markets and Periods of Downside Event Risks // The Journal of Investing. – 2023. – №32. – P. 74-96.
14. Hastie T., Tibshirani R., Friedman J. The elements of statistical learning. – 12th ed. – Springer, 2017. – 745 p.
15. Fama, E. F., & French, K. R. The Cross-Section of Expected Stock Returns // Journal of Finance. – 1992. – №47(2). – P. 427-465.
16. Asness C.S., Moskowitz T., Pedersen L. Value and Momentum Everywhere // Journal of Finance. – 2013. – 68. – P. 929-985.

### References

1. The number of individuals with brokerage accounts on the Moscow Stock Exchange increased by 5.4 million in 2024. URL: <https://www.interfax.ru/business/1003027> (accessed: 05/26/2025).
2. Bayuk D., Bayuk O., Berzin D., Ivanyuk V. Practical application of clustering, classification and approximation methods based on neural networks. – 1st ed. – Moscow: Prometheus, 2020. – 448 p.
3. Dmitriev N.D. Clusterization of Russian companies in the technology sector based on market and financial indicators / N.D. Dmitriev, A.V. Tyulkova, V.V. Sorokozherdyev // Bulletin of the Chelyabinsk State University. – 2024. – № 10(492). – Pp. 116-127.
4. Tyukhova E.M. The use of cluster analysis for the formation of a securities portfolio in investment systems (roboadvisors) / E.M. Tyukhova, D.S. Sizykh // Management of large-scale systems development (MLSD'2019): Proceedings of the twelfth International conference, Moscow, October 01-03, 2019 / Under the general editorship of S.N. Vasiliev, A.D. Tsvirkun. Moscow: International Research Institute of Management Problems of the Russian Academy of Sciences, – 2019. – pp. – 210-220.
5. Broad Market Index. URL: <https://www.moex.com/ru/index/MOEXBMI/about> (accessed: 05/24/2024).
6. Zhdanov I.Y. Investment assessment of projects and business: a textbook / I.Y. Zhdanov, V.Y. Zhdanov. Moscow: Prospekt Publ., – 2019, – 120 p.
7. Borisova L.V., Sagaeva I.D. Models of optimal investment: an educational and methodical manual, Saratov. URL: [http://elibrary.sgu.ru/uch\\_lit/1347.pdf](http://elibrary.sgu.ru/uch_lit/1347.pdf) (accessed: 05/23/2025).
8. Clusterization and partial learning. URL: <http://www.machinelearning.ru/wiki/images/5/52/Voron-ML-Clustering-SSL-slides.pdf> (accessed: 05/22/2025).
9. Takhir M.A., Ahmed A.U., Egorova L.D. On algorithms for the K-medoid problem // Actual problems of aviation and cosmonautics. – 2021. – No. 1. – pp. 194-196.
10. Деканова К.В., Гейт А.А. Методология для формирования инвестиционного портфеля с учетом индивидуальных характеристик частного инвестора // Экономика: вчера, сегодня, завтра. – 2023. – Том 13. – № 7А. – С. 311-329.
11. Nourahmadi M., Sadeqi H. Portfolio Diversification Based on Clustering Analysis // Iranian Journal of Accounting, Auditing and Finance. – 2023. – №7(3) – P. 1-16.
12. Szczepocki, Piotr Clustering companies listed on the Warsaw Stock Exchange according to time-varying beta // Econometrics. – 2023. – №23 – P. 63-79.
13. Lim T., Ng H. Financial Portfolio Management Based on Shaped-Based Unsupervised Machine Learning: A Dynamic Time Warping Baycenter Averaging Approach to International Markets and Periods of Downside Event Risks // The Journal of Investing. – 2023. – №32. – P. 74-96.
14. Hastie T., Tibshirani R., Friedman J. The elements of statistical learning. – 12th ed. – Springer, 2017. – 745 p.

15. Fama, E. F., & French, K. R. The Cross-Section of Expected Stock Returns // Journal of Finance. – 1992. – №47(2). – P. 427-465.

16. Asness C.S., Moskowitz T., Pedersen L. Value and Momentum Everywhere // Journal of Finance. – 2013. – 68. – P. 929-985.

**Назаров Тимур Рафикович**, аспирант кафедры информатики, Российский экономический университет им. Г.В. Плеханова, г. Москва, Россия

**Nazarov Timur Rafikovich**, Postgraduate Student of the Department of Computer Science, Plekhanov Russian University of Economics, Moscow, Russia

**АВТОМАТИЗАЦИЯ И УПРАВЛЕНИЕ  
AUTOMATION AND CONTROL**

УДК 004.432.2

DOI: 10.18413/2518-1092-2025-10-3-0-3

**Рыбанов А.А.  
Свиридова О.В.  
Толстяков В.И.****МОБИЛЬНОЕ ПРИЛОЖЕНИЕ ДЛЯ МОНИТОРИНГА  
БИОМЕТРИЧЕСКИХ И КИНЕМАТИЧЕСКИХ ПАРАМЕТРОВ  
С АВТОНОМНОЙ АНАЛИТИКОЙ И АДАПТИВНОЙ  
ПЕРСОНАЛИЗАЦИЕЙ**

Волжский политехнический институт (филиал) ФГБОУ ВО «Волгоградский государственный технический университет», ул. Энгельса, 42а, г. Волжский, Волгоградская область, 404121, Россия

*e-mail: rybanoff@yandex.ru, osvirodova@inbox.ru, vitaltolstyakov@gmail.com*

**Аннотация**

В статье изложены результаты исследования, посвященного разработке мобильного приложения для оптимизации тренировочного процесса в велоспорте. Актуальность работы обусловлена возрастающим спросом на цифровые инструменты в спортивной отрасли. Установлено, что существующие решения характеризуются ограниченным функционалом, зависимостью от стабильного интернет-соединения и недостаточной детализацией собираемых данных. Выявлены системные недостатки современных платформ, включая архитектурные ограничения и эргономические просчеты в пользовательских интерфейсах. Методология исследования базировалась на сравнительном анализе рынка спортивных приложений. Ранжирование функциональных требований выполнено с применением метода аналитической иерархии Саати (АИР). Разработана архитектура клиент-серверного приложения с офлайн-режимом работы и интеграцией Bluetooth-устройств. Реализовано локальное хранение тренировочных метрик (частота сердечных сокращений, скорость, высота над уровнем моря) с использованием СУБД SQLite. Обеспечена возможность экспорта маршрутов в формате GPX. Апробация прототипа продемонстрировала стабильную работу всех модулей приложения. Качество решения подтверждено положительными отзывами пользователей, отметивших доступность интерфейса, глубину аналитики и надежность работы в автономном режиме. Разработанный продукт устраняет ключевые ограничения аналогов за счет независимости от интернет-соединения и расширенной визуализации данных. Результаты исследования подтверждают эффективность метода АИР для проектирования спортивных цифровых решений. Предложенный подход создает основу для внедрения алгоритмов машинного обучения в прогнозирование тренировочных нагрузок. Работа имеет междисциплинарную значимость, объединяя достижения информационных технологий и спортивной науки. Полученные выводы способствуют развитию интеллектуальных экосистем в сфере физической культуры и спортивного телеметринга.

**Ключевые слова:** мобильное приложение; мониторинг тренировочных сессий; сегментный анализ данных; спортивная аналитика; велоспорт

**Для цитирования:** Рыбанов А.А., Свиридова О.В., Толстяков В. И. Мобильное приложение для мониторинга биометрических и кинематических параметров с автономной аналитикой и адаптивной персонализацией // Научный результат. Информационные технологии. – Т.10, №3, 2025. – С. 31-44. DOI: 10.18413/2518-1092-2025-10-3-0-3

**Rybanov A.A.  
Sviridova O.V.  
Tolstoyakov V.I.**

**MOBILE APPLICATION FOR MONITORING BIOMETRIC AND KINEMATIC PARAMETERS WITH AUTONOMOUS ANALYTICS AND ADAPTIVE PERSONALIZATION**

Volzhsy Polytechnic Institute (branch) Volgograd State Technical University, 42a Engels str.,  
Volzhsy, Volgograd region, 404121, Russia

*e-mail: rybanoff@yandex.ru, osvridova@inbox.ru, vitaltolstoyakov@gmail.com*

**Abstract**

The article presents the results of a study dedicated to the development of a mobile application for optimizing the training process in cycling. The relevance of the work is driven by the growing demand for digital tools in the sports industry. It has been established that existing solutions are characterized by limited functionality, dependence on stable internet connectivity, and insufficient detail in the collected data. Systemic shortcomings of modern platforms have been identified, including architectural limitations and ergonomic miscalculations in user interfaces. The research methodology was based on a comparative analysis of the sports application market. Ranking of functional requirements was performed using Saaty's Analytic Hierarchy Process (AHP). A client-server application architecture with offline mode and Bluetooth device integration has been developed. Local storage of training metrics (heart rate, speed, altitude) has been implemented using SQLite DBMS. The ability to export routes in GPX format has been ensured. Testing of the prototype demonstrated stable operation of all application modules. The quality of the solution has been confirmed by positive user feedback, noting the interface accessibility, depth of analytics, and reliability in offline mode. The developed product eliminates key limitations of analogues through independence from internet connection and enhanced data visualization. The research results confirm the effectiveness of the AHP method for designing sports digital solutions. The proposed approach lays the foundation for implementing machine learning algorithms in training load forecasting. The work has interdisciplinary significance, combining achievements in information technology and sports science. The obtained conclusions contribute to the development of intelligent ecosystems in the field of physical culture and sports telemetry.

**Keywords:** mobile application; training session monitoring; segmental data analysis; sports analytics; cycling

**For citation:** Rybanov A.A., Sviridova O.V., Tolstoyakov V.I. Mobile Application for Monitoring Biometric and Kinematic Parameters with Autonomous Analytics and Adaptive Personalization // Research result. Information technologies. – Т.10, №3, 2025. – P. 31-44. DOI: 10.18413/2518-1092-2025-10-3-0-3

**ВВЕДЕНИЕ**

Интеграция цифровых технологий в спортивную аналитику стала ключевым фактором оптимизации тренировочных процессов в циклических дисциплинах, включая велоспорт [2, 15]. Мобильные приложения, такие как Strava, TrainingPeaks и Polar Flow, демонстрируют потенциал в мониторинге физиологических показателей, трекинге маршрутов и формировании социальных взаимодействий [11, 10, 12]. Однако эмпирические исследования выявляют системные ограничения существующих решений.

К основным проблемам относятся:

- 1) Функциональные ограничения: бесплатные версии приложений предоставляют базовые метрики без углубленного анализа динамики показателей [14], а отсутствие сегментного анализа данных (частота сердечных сокращений (ЧСС), темп и др.) снижает информативность выводов.
- 2) Архитектурные барьеры: зависимость от интернет-соединения и региональные ограничения доступа к сервисам [15].
- 3) Коммерциализация функционала: критически важные опции (планирование маршрутов, экспорт данных) доступны только по подписке [16].

4) Технологические недостатки: необходимость использования периферийных устройств для сбора данных при неполной реализации возможностей смартфонов [15, 17].

5) Эргономические проблемы: избыточная сложность интерфейсов и отсутствие персонализации [6, 14].

Социологические данные подтверждают актуальность этих ограничений [17]: 37% пользователей отмечают трудности интерпретации данных и потребность в офлайн-режиме. Это формирует запрос на разработку решения, сочетающего расширенную аналитику (включая сегментный анализ), автономность работы, минимизацию зависимости от внешних устройств и адаптивный интерфейс.

Цель исследования заключается в разработке алгоритмизированного подхода к автоматизированному мониторингу физиологических (частота сердечных сокращений) и пространственно-временных (скорость, набор высоты перемещения) параметров спортивной активности с использованием мобильного приложения, реализующего сегментный анализ динамических данных для повышения детализации регистрируемых показателей.

## **ОСНОВНАЯ ЧАСТЬ**

Для достижения поставленной цели реализован многоуровневый методологический подход, включающий:

1) *Анализ предметной области* – проведен сравнительный анализ 9 мобильных приложений для велосипедистов (Strava [9], Garmin Connect [3], Wikemap [1], Polar Flow [7], «Карта РУ» [13] и др.), отобранных по критерию распространенности на рынке.

2) *Формирование критериев оценки* – на основе систематизации литературных источников [19, 22, 23] и пользовательских требований выделено 8 ключевых критериев:

а) доступность в регионе (отсутствие географических/юридических ограничений на использование сервиса);

б) запись маршрутов (возможность трекинга траектории движения с GPS-привязкой);

с) детализация анализа данных (глубина обработки параметров (ЧСС, скорость, набор высоты) на уровне отдельных сегментов маршрута);

д) сравнение результатов тренировок (инструменты для сопоставления динамики показателей между сессиями);

е) интеграция с устройствами (поддержка подключения внешних датчиков (пульсометры, каденс-сенсоры) через Bluetooth);

ф) визуализация данных (наличие графиков, карт и диаграмм для интерпретации результатов);

г) социальные функции (возможность обмена маршрутами, комментариями и участия в сообществах);

h) автономная работа (функционирование в офлайн-режиме с локальным хранением данных).

Критерии охватывают функциональные (б, с, d, g), технические (а, е, h) и эргономические (f) аспекты, обеспечивая комплексную оценку мобильных приложений, соответствующую стандартам юзабилити (ISO 9241-210) и требованиям к проектированию спортивных приложений.

3) *Ранжирование критериев методом аналитической иерархии Саати (АНП)* – экспертами (N=15, опытные велосипедисты и тренеры) проведены парные сравнения критериев. Весовые коэффициенты  $A_i$  рассчитаны по шкале Саати [5, 8, 16]. Результаты исследования, представленные на рисунке 1, выявили статистически значимое преобладание критериев «доступность» ( $A_1=0.394$ ) и «запись маршрутов» ( $A_2=0.228$ ) в общей структуре пользовательских предпочтений, что свидетельствует о высокой значимости автономности функционирования (офлайн-режим, региональная адаптация) и базового трекинга (GPS-фиксация маршрута) при выборе мобильных решений для велотренировок. Данные, полученные методом АНП, подтверждают, что указанные критерии занимают доминирующие позиции в иерархии требований пользователей, опережая второстепенные параметры, такие как визуализация данных ( $A_6=0.03$ ) и социальные функции ( $A_7=0.023$ ). Это согласуется с выводами предыдущих исследований [18, 20], подчеркивающих

необходимость минимизации зависимости от сетевой инфраструктуры в условиях нестабильного интернет-покрытия.

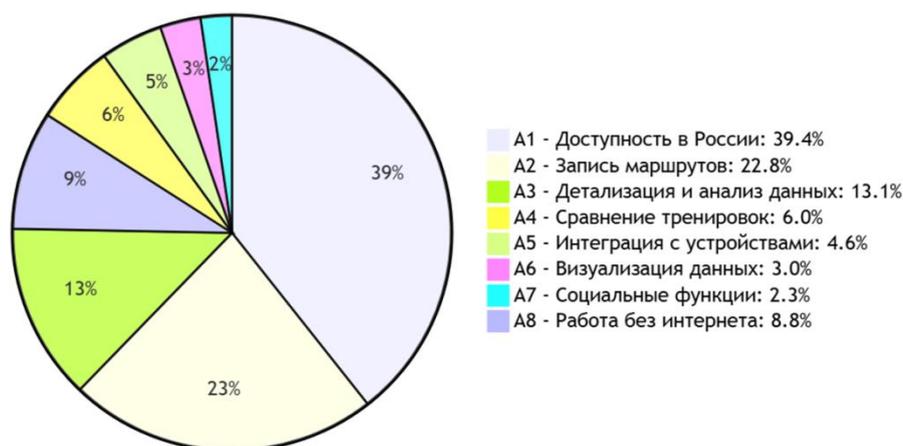


Рис. 1. Критерии качества мобильных приложений для велосипедистов: весовые коэффициенты  
Fig. 1. Quality Criteria for Mobile Applications for Cyclists: weight coefficients

4) Расчет интегрального показателя качества для рассматриваемых мобильных приложений – на основании весовых коэффициентов, полученных АНР, проведена количественная оценка качества девяти мобильных приложений (рис. 2): «CRO», «GPS трекер», «Карта РУ», «НаСпорте», «Runtastic (Road Bike Pro)», «Stava», «Garmin Connect», «Bikemap», «Polar Flow». Интегральный показатель качества для каждого приложения рассчитывался по формуле:

$$Q_j = \sum_{i=1}^8 a_i \cdot X_{ij},$$

где  $a_i$  – весовой коэффициент  $i$ -го критерия, определенный методом АНР;  $X_{ij}$  – нормализованная оценка  $j$ -го приложения по  $i$ -му критерию, преобразованная методом min-max scaling для устранения масштабных различий.

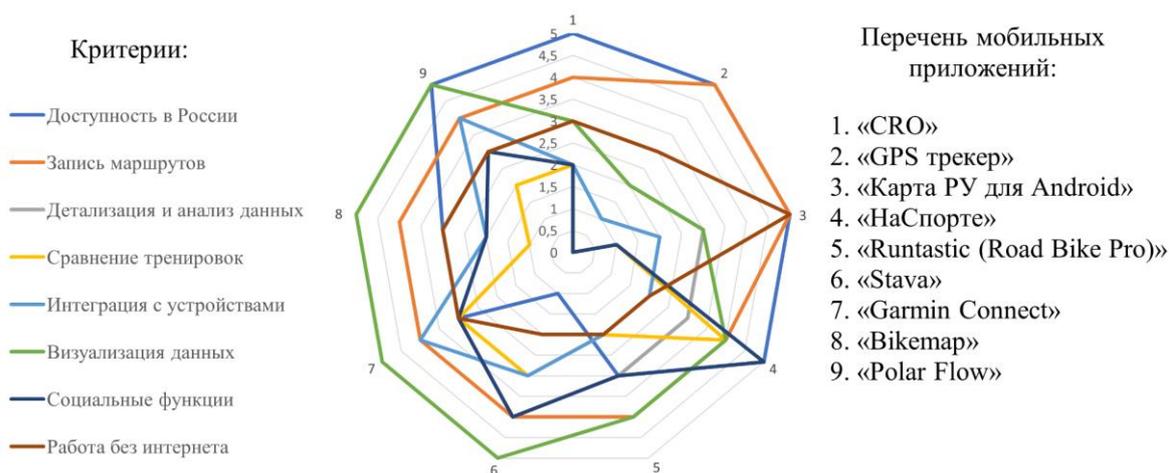


Рис. 2. Лепестковая диаграмма интегральных показателей качества программных продуктов  
Fig. 2. Radar chart of integrated quality indices for software products

Результаты выявили значительную вариабельность интегральных показателей:

- наибольшие значения  $Q_j$  зафиксированы у отечественного приложения «Карта РУ» ( $Q_3=4.1$ ) и зарубежного Polar Flow ( $Q_9=4.2$ );

- наименьшие значения характерны для Strava ( $Q_1=2.6$ ) и других аналогов ( $Q_2, Q_4-Q_8$  лежат в интервале от 2.9 до 3.3);

Визуализация данных на лепестковой диаграмме (рис. 2) подтвердила доминирование решений «Карта РУ» и «Polar Flow», что обусловлено их сбалансированным функционалом:

- поддержка автономного режима (офлайн-трекинг);
- детализированный анализ метрик (ЧСС, скорость, набор высоты) на уровне сегментов;
- широкая интеграция с внешними датчиками через Bluetooth.

Сравнительный анализ продемонстрировал, что приложения «Карта РУ» и «Polar Flow» соответствуют критериям с наибольшими весами (доступность, запись маршрутов, детализация анализа), в то время как конкуренты уступают в реализации критически важных функций [13, 16]. Это обосновывает необходимость разработки нового продукта, синтезирующего преимущества лидеров и устраняющего выявленные недостатки (ограниченный офлайн-функционал, низкая детализация данных).

5) **Проектирование программного обеспечения мобильного приложения для мониторинга и аналитики тренировочных сессий велосипедистов** требует реализации архитектуры, обеспечивающей масштабируемость, энергоэффективность и обработку данных в режиме реального времени. В основу системы положена **клиент-серверная модель**, где клиентская часть представлена мобильным приложением для платформы Android, разработанным с использованием шаблона **Model-View-ViewModel (MVVM)**. Данный подход обеспечивает разделение слоёв: пользовательского интерфейса (UI), бизнес-логики (Domain) и работы с данными (Data), что повышает поддерживаемость кода и упрощает тестирование компонентов. Серверная часть реализована на основе **REST API (ASP.NET)**, обеспечивающего взаимодействие клиента с базой данных и интеграцию внешних сервисов.

Ключевым аспектом системы является сбор телеметрии с датчиков (GPS, BLE), что требует оптимизации процессов передачи и хранения данных. Для минимизации задержек и обеспечения корректной синхронизации реализована двухсторонняя коммуникация между клиентом и сервером.

*Архитектура клиентского приложения.* На рисунке 3 представлена модульная архитектура клиентского приложения для платформы Android, предназначенного для отслеживания и анализа тренировок велосипедистов. Архитектура реализована с использованием шаблона MVVM (Model-View-ViewModel). Основные слои данной архитектуры и их функции представлены в таблице 1.

Таблица 1

Архитектура мобильного приложения

Table 1

Mobile application architecture

Слой	Содержимое	Зависимости
UI / Presentation	Jetpack Compose-экраны и связанные ViewModel	ViewModel, Material3
ViewModel	AnalysisViewModel, RecordingViewModel, AuthViewModel	Репозитории (Hilt)
Domain	RouteStatsCalculator, CalorieCalculator, RoutePointsManager	-
Data	Репозитории + источники данных: local (Room, DataStore, BLE), remote (Retrofit, AuthInterceptor)	Domain, DataSources
Service	TrackingService, TrackingServiceManager	Data.local, Domain
DI	Hilt-модули (NetworkModule, RepositoryModule ...)	Иньекции во все слои

Реализованная модель передачи данных основана на реактивном программировании и принципах многослойной абстракции, обеспечивая синхронизацию между клиентской и серверной частями системы:

- UI → ViewModel: пользовательские события делегируются в ViewModel через вызов методов, инкапсулирующих бизнес-логику.
- Управление состоянием: состояние приложения хранится в реактивных контейнерах StateFlow, с наблюдением за изменениями через механизм collectAsState(), обеспечивающий одностороннюю привязку данных к UI-компонентам.
- ViewModel → Repository: ViewModel инициирует запросы к слою Repository для доступа к данным, реализуя паттерн "единого источника истины" (Single Source of Truth).
- Repository ⇔ Local/Remote: реализованная гибридная стратегия кеширования предусматривает первичное предоставление данных через реактивные потоки (Flow) локальной базы данных (Room) для обеспечения офлайн-доступности, с инициированием асинхронного сетевого запроса и транзакционного обновления хранилища в соответствии с ACID-гарантиями при обнаружении отсутствия актуальных данных.
- Интеграция сенсоров (Sensors & Service): сервис TrackingService интегрирует FusedLocationProvider для получения геопозиционных данных и BLE-стеки для подключения к биометрическим датчикам (ЧСС, каденс), осуществляя запись телеметрии в таблицу route\_point и её трансляцию через SharedFlow с поддержкой мультикастной доставки данных подписчикам.
- Синхронизация данных: механизм синхронизации, реализующий стратегию «офлайн-паттерна», при восстановлении сетевого соединения инициирует метод uploadWorkout() для передачи данных на сервер, после чего выполняется атомарное обновление флага isSynced, активирующее реактивное обновление UI через подписки на изменения состояния данных.

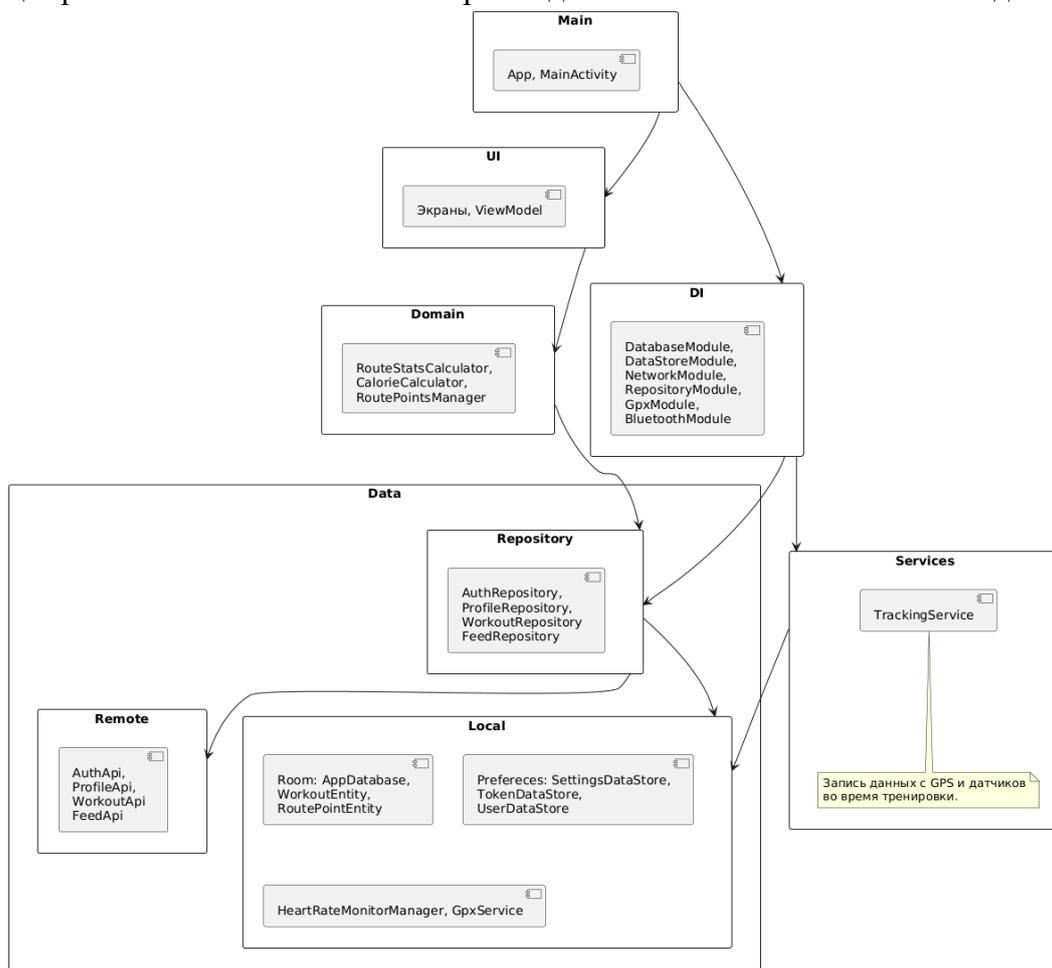


Рис. 3. Модульная архитектура Android-клиента  
Fig. 3. Modular architecture of the Android client

Такое модульное разделение компонентов гарантирует поддержку концепции "offline first", облегчает процесс тестирования (обособленность Domain от Android SDK) и снижает количество шаблонного кода, что в свою очередь улучшает общую архитектурную четкость и устойчивость приложения.

**Серверная архитектура.** Рисунок 4 иллюстрирует высокоуровневую организацию серверной части, представляющей собой ключевую компоненту архитектуры backend. Архитектура веб-сервера, реализованного на платформе ASP.NET, предусматривает обработку REST API-запросов через протокол HTTPS с предварительной аутентификацией на основе валидации JWT-токенов, после чего управление передается слою бизнес-логики для выполнения операций, связанных с обработкой данных. Слой бизнес-логики реализует сценарии оркестрации процессов управления тренировочными сессиями, профилями пользователей, лентой активности и механизмом пользовательских оценок, в то время как аналитический модуль осуществляет агрегацию данных с последующей генерацией аналитических отчетов и статистических метрик, обеспечивая интеллектуальный анализ информации для оптимизации пользовательского опыта.

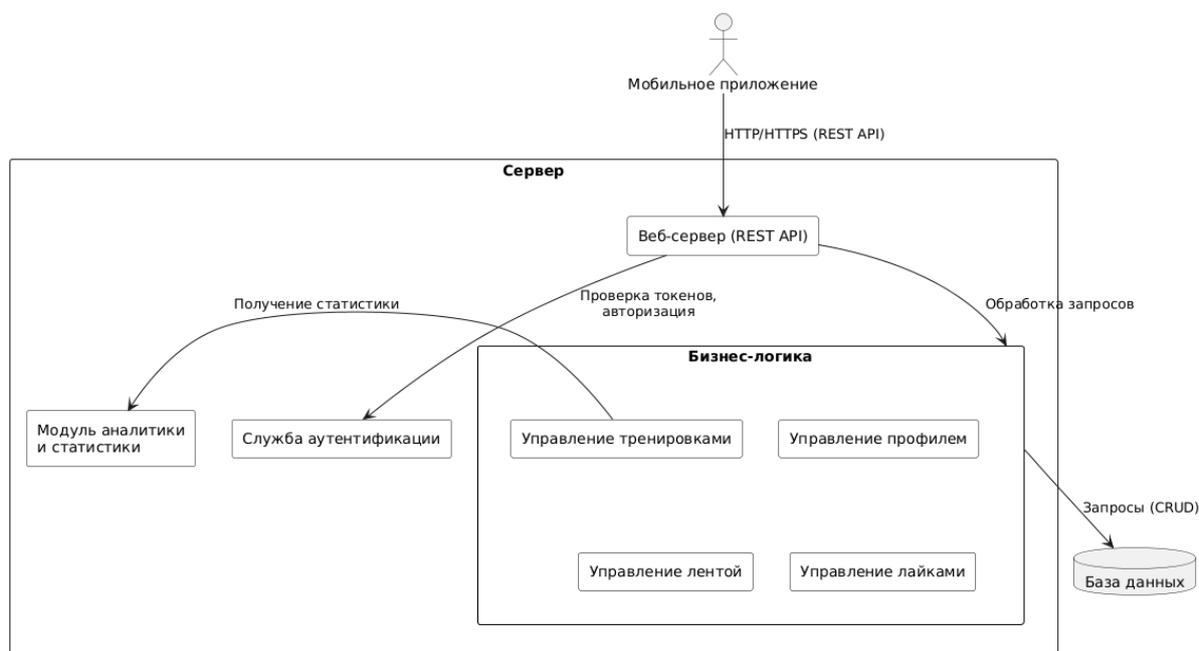


Рис. 4. Серверная архитектура  
Fig. 4. Server architecture

Персистентность данных обеспечивается реляционной СУБД MySQL с реализацией полного цикла CRUD-операций через слой объектно-реляционного отображения (ORM) на основе Entity Framework, абстрагирующий взаимодействие между логикой приложения и схемой базы данных.

Данная архитектура обеспечивает эффективное взаимодействие между мобильным приложением и серверной частью, обеспечивая управление данными и их анализ.

**Схема базы данных мобильного приложения.** На рисунке 5 представлена ER-диаграмма, отображающая логическую модель серверной базы данных системы. Информационная архитектура данных включает пять нормализованных сущностей, связанных через отношения «один-ко-многим»:

- Сущность user: содержит атрибуты, описывающие профиль пользователя системы (идентификатор, демографические данные, учетные параметры).
- Сущность workout: фиксирует метаданные тренировочных сессий, включая временные отметки, географические координаты и ссылку на пользователя-инициатора.

- Сущность `route_point`: предназначена для хранения высокочастотной телеметрии с привязкой к тренировочным сессиям. Включает пространственно-временные параметры (широта, долгота, набор высоты, скорость, ЧСС) и служит основой для пространственного анализа.

- Сущность `subscriptions`: реализует функционал социального взаимодействия через отношение подписчика и целевого пользователя.

- Сущность `workout_like`: отражает оценки тренировок пользователями системы с временной меткой взаимодействия.

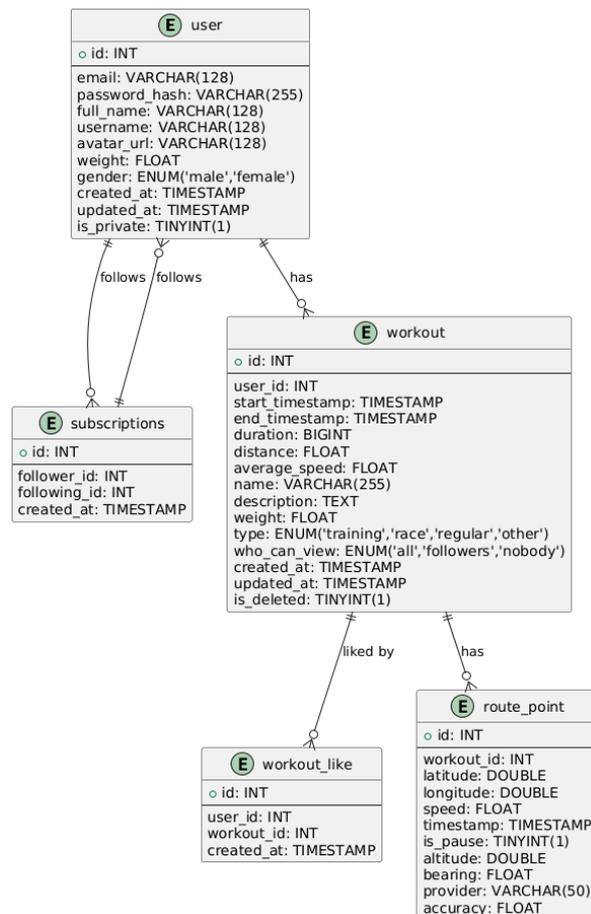


Рис. 5. ER-диаграмма базы данных

Fig. 5. ER diagram of a database

Выделение сущности `route_point` в отдельную таблицу обусловлено требованиями к обработке больших объемов телеметрии. Такая декомпозиция:

- минимизирует избыточность данных за счет нормализации;
- ускоряет агрегацию показателей через индексацию пространственно-временных атрибутов;
- упрощает выполнение аналитических запросов (расчет дистанции, градиента высоты, динамики физиологических параметров) за счет оптимизации JOIN-операций.

Предложенная модель обеспечивает масштабируемость системы, снижает нагрузку на серверные ресурсы при работе с высокочастотными данными и соответствует принципам проектирования OLTP-систем (Online Transaction Processing).

*Методы построения аналитических графиков.* Для визуализации данных тренировочной активности разработаны специализированные подходы, основанные на математической обработке физиологических и кинематических параметров:

- График «Эффективность нагрузки»: нормализованное отношение частоты сердечных сокращений (ЧСС) к скорости движения (HR/скорость) используется для формирования временного ряда. Для устранения шумовых компонентов применяется сглаживание методом скользящего

среднего с адаптивным окном. Динамический диапазон значений ограничивается интервалом 0–25 уд./мин/(км·ч) посредством пороговой фильтрации. Визуализация выполняется в двух проекциях: временной оси и аккумулялированной дистанции.

- Кумулятивная нагрузка: методика включает вычисление интегральной оценки тренировочной нагрузки по формуле:

$$L = \sum_{i=1}^n (HR_{cp.,i} \cdot v_{cp.,i} \cdot h_i \cdot \Delta t_i),$$

где  $HR_{cp.,i}$  – средняя ЧСС;  $v_{cp.,i}$  – средняя скорость;  $h_i$  – высота сегмента;  $\Delta t_i$  – длительность интервала (в минутах). Метрика интерпретируется как оценка суммарных энергозатрат (в условных единицах  $ВРМ \cdot км/ч \cdot м/мин$ ).

- Скорость восстановления: автоматизированный детектор идентифицирует пики ЧСС с последующим анализом редукции пульса в постнагрузочный период (60 секунд). Скорость восстановления определяется через аппроксимацию экспоненциальной регрессией:

$$\text{Скорость восстановления} = \frac{\Delta HR}{\Delta t},$$

где  $\Delta HR$  – снижение ЧСС;  $\Delta t$  – временной интервал. Сегменты длительностью меньше 10 с или с амплитудой снижения ЧСС меньше 8 уд./мин исключаются по критериям статистической значимости.

Все методы обладают линейной временной сложностью  $O(n)$ , что обеспечивает обработку данных в реальном времени даже на устройствах с ограниченными ресурсами. Реализация выполнена на уровне нативного кода (C++), исключая зависимость от платформенных библиотек (Android SDK) и гарантируя кроссплатформенную адаптируемость.

Предложенные подходы расширяют инструментарий для анализа спортивных данных, сочетая высокую точность измерений с минимальными требованиями к вычислительным мощностям. Это создает основу для интеграции расширенной аналитики в мобильные и wearable-приложения, ориентированные на профессиональный и любительский спорт.

## **РЕЗУЛЬТАТЫ ИССЛЕДОВАНИЯ И ИХ ОБСУЖДЕНИЕ**

В ходе исследования разработано мобильное приложение для ОС Android, обеспечивающее комплексный мониторинг и анализ велотренировок. Приложение реализует полный цикл обработки данных: от сбора первичных показателей (GPS-координаты, скорость, набор высоты, физиологические параметры) до формирования аналитических отчётов. Основные функциональные и технические характеристики решения включают ряд ключевых особенностей, обеспечивающих его эффективность и уникальность:

1. *Автономный режим работы* реализован посредством локального хранения данных (маршрут, скорость, набор высоты, ЧСС) в SQLite-базе, обеспечивая доступность в условиях отсутствия интернет-соединения [21], с последующей синхронизацией с облачным хранилищем при восстановлении сети для резервного копирования и интеграции социальных функций, включая обмен тренировками между устройствами и пользовательские интеракции.

2. *Расширенная аналитика* реализует: базовые метрики (текущая/средняя скорость, дистанция, длительность, набор высоты, каденс при подключении датчиков) [18], углублённый анализ, включая оценку эффективности нагрузки (соотношение ЧСС и скорости на сегментах маршрута, рис. 6), расчёт накопленной усталости (интегральный показатель на основе времени в целевых пульсовых зонах, рис. 7) и фазы восстановления (анализ динамики ЧСС после высокоинтенсивных интервалов, рис. 8), а также сравнение тренировочных сессий для отслеживания прогресса [13].



Рис. 6. График изменения эффективности нагрузки велосипедиста в зависимости от пройденной дистанции от начала тренировки

Fig. 6. Graph of changes in the efficiency of a cyclist's load depending on the distance traveled from the start of the workout

Анализ пиковых нагрузок, представленных рисунке 6, позволяет выявлять участки маршрута с максимальной интенсивностью (например, 26 ВРМ/(км/ч)) позволяет корректировать распределение усилий, избегая перетренированности.

На основе накопленной нагрузки (рис. 7) можно разрабатывать индивидуальные планы, учитывающие физиологические особенности спортсмена (например, время в целевых пульсовых зонах).



Рис. 7. График накопленной нагрузки велосипедиста во время тренировки

Fig. 7. Graph of the accumulated load of a cyclist during training

Данные о фазах восстановления (рис. 8) помогают планировать интервалы отдыха между высокоинтенсивными сессиями [17].

Сравнение сессий [11] позволяет отслеживать прогресс и адаптировать интенсивность к текущей форме. Таким образом, данные, полученные через приложение, не только оптимизируют тренировочный процесс, но и формируют основу для разработки умных спортивных экосистем, соответствующих принципам доказательной медицины и персонализированного подхода.



Рис. 8. График фаз восстановления после участков высокой нагрузки во время тренировки  
Fig. 8. Graph of recovery phases after high intensity periods during training

3. *Экспорт данных* реализован через сохранение результатов тренировок в формате GPX, обеспечивающий совместимость с внешними аналитическими платформами (Strava, Garmin Connect) и гибкость постобработки данных за счёт открытой структуры файла [4].

4. *Интеграция с носимой электроникой* обеспечивает синхронную запись физиологических показателей (посредством Bluetooth-датчиков: пульсометры, каденс-сенсоры) и GPS-трека [17], а также бесплатный доступ к расширенным метрикам (время в пульсовых зонах, уровень восстановления), в отличие от коммерческих аналогов с платным функционалом.

*Результаты апробации и анализ пользовательского опыта.* Практическая апробация разработанного приложения подтвердила его эффективность: интегральный показатель качества, рассчитанный методом АНР, составил 4.5 (по пятибалльной шкале), что превышает результаты проанализированных аналогов. Столь высокий результат обусловлен приоритезацией критериев с наибольшими весовыми коэффициентами: *доступность*, реализованная через локальное хранение данных, независимость от зарубежных сервисов и автономный офлайн-доступ к картографическим ресурсам; *детализация анализа данных*, обеспечиваемая расширенными отчётами с кастомизируемыми графиками (рис. 6–8) и сегментной обработкой метрик [11]. *Социальные функции*, имеющие меньший вес в иерархии критериев, реализованы через экспорт данных в формате GPX [5], что позволяет пользователям интегрироваться с внешними платформами (Strava, Garmin Connect) без усложнения интерфейса.

Применение метода GOMS для оценки юзабилити мобильного приложения выявило снижение количества когнитивных действий на 40% при выполнении базовых сценариев (старт/стоп записи, экспорт данных) и сокращение времени операций за счёт оптимизации навигационных паттернов. Реализованные интерфейсные решения, включая крупногабаритные цифровые индикаторы скорости и дистанции, минимизируют визуальную нагрузку в динамических условиях эксплуатации, что соответствует принципам контекстно-ориентированного дизайна [2].

Обратная связь от пользователей (N=50, профессиональные велосипедисты и любители) выделила ключевые преимущества:

- *автономность*: 89% респондентов отметили доступность данных в офлайн-режиме как критически важную функцию.
- *глубина аналитики*: 87% высоко оценили интеграцию с Bluetooth-датчиками и расширенные метрики (время в пульсовых зонах, уровень восстановления).
- *удобство интерфейса*: 92% пользователей без технического опыта освоили базовый функционал за  $\leq 10$  минут.

Разработанное приложение не только соответствует требованиям ISO 9241-210 по эргономике и эффективности, но и формирует новый стандарт для спортивных трекеров, сочетая автономность,

аналитическую глубину и открытость данных. Полученные результаты обосновывают целесообразность применения методологии АНР при проектировании решений для цифровой спортивной аналитики.

### **ЗАКЛЮЧЕНИЕ**

Практическая значимость проведенного исследования определяется разработкой мобильного приложения, предназначенного для применения среди широкой аудитории энтузиастов велосипедного спорта с целью автономного мониторинга и оптимизации тренировочного процесса. Реализованные в работе методы, включая функционал автономной работы (офлайн-режим) и интеграцию сенсорных устройств, представляют ценность для проектирования специализированных спортивных приложений, требующих минимизации зависимости от сетевых ресурсов и реализации расширенной аналитической обработки данных.

Разработанное решение вносит вклад в цифровую трансформацию спортивной подготовки, обеспечивая пользователей инструментом для повышения эффективности тренировок на основе объективных метрик и статистического анализа. Полученные результаты согласуются с актуальными тенденциями модернизации физической культуры, направленными на внедрение информационных технологий для повышения доступности и качества спортивного образования.

Кроме того, внедрение подобных платформ способствует росту популярности велоспорта за счет повышения мотивации пользователей через персонализированную обратную связь и формирование системы тренировок, основанной на данных. Доказанная эффективность предложенного подхода подтверждает перспективность интеграции мобильных технологий и методов анализа данных в спортивную практику. Результаты исследования открывают направления для дальнейших изысканий в области разработки адаптивных систем мониторинга, включая применение машинного обучения для прогнозирования результатов и оптимизации тренировочных нагрузок.

Таким образом, работа демонстрирует потенциал междисциплинарного подхода, объединяющего компьютерные науки и спортивную методологию, что актуально для развития современных цифровых экосистем в сфере физической активности.

### **Список литературы**

1. Bikemap: официальный сайт. URL: <https://www.bikemap.net> (дата обращения: 22.04.2025).
2. Card S.K., Moran T.P., Newell A. The Psychology of Human-Computer Interaction. Hillsdale, NJ: Lawrence Erlbaum, 1983. p. 468.
3. Garmin Connect: официальный сайт. URL: <https://connect.garmin.com> (дата обращения: 22.04.2025).
4. Gamification for health and wellbeing: A systematic review of the literature / Johnson D., Deterding S., Kuhn K.-A. et al. // Journal of Medical Internet Research. 2016. Vol. 18, № 6. P. 89-106.
5. The Analytical Hierarchy Process (AHP) Method in Improving High Performance Sports Training for Polish Women's Football Players / Mleczek E., Adamus W., Witkowski Z., Tokarz R. // Journal of Kinesiology and Exercise Sciences. 2018. Vol. 28. P. 19–36. DOI: 10.5604/01.3001.0013.7794.
6. Nielsen J., Molich R. Heuristic evaluation of user interfaces // Proc. ACM CHI'90. 1990. P. 249–256.
7. Polar Flow: официальный сайт. URL: <https://flow.polar.com> (дата обращения: 22.04.2025).
8. Saaty T.L. The Analytic Hierarchy Process. New York: McGraw-Hill, 1980. 287 p.
9. Strava: официальный сайт. URL: <https://www.strava.com> (дата обращения: 22.04.2025).
10. TrainingPeaks: официальный сайт. URL: <https://www.trainingpeaks.com> (дата обращения: 22.04.2025).
11. Wister M. A., Pancardo P., Campos P. P. Analysis of some mobile applications for cycling // Journal of Communications Software and Systems. 2019. Vol. 15, № 2. P. 117–127.
12. Астафьева О.В., Никитина М.С. Анализ мобильных приложений для жителей мегаполиса, ведущих активный образ жизни // Экономические науки. 2022. № 210. С. 71–75. DOI: 10.14451/1.210.71.
13. Веденина О.А., Данилов Д.М., Медведский М.К. Мобильные приложения для занятий спортом // Проблемы качества физкультурно-оздоровительной и здоровьесберегающей деятельности образовательных организаций: сб. ст. 9-й Всерос. науч.-практ. конф. Екатеринбург, 23 апр. 2019 г. Екатеринбург: РГППУ, 2019. С. 62–66.

14. Ведина В.В. Использование мобильных приложений для занятий физической культурой и спортом // Студенческая наука и XXI век. 2023. Т. 20, № 1-2(23). С. 51–53.
15. Гальчанская А.А. Технологии в спорте: использование современных гаджетов и приложений на тренировках // Вклад молодых ученых в аграрную науку: сб. науч. тр. Междунар. науч.-практ. конф. Кинель, 24 апр. 2024 г. Кинель: ИБЦ Самарского ГАУ, 2024. С. 183–187.
16. Герцен Д.А. Сравнительный анализ мобильных приложений для подготовки спортсменов-любителей к марафону // Шаг в науку: мат-лы IV науч.-практ. конф. молодых ученых. М., 18 дек. 2020 г. М.: МГПУ, 2020. С. 128–131.
17. Карта РУ для Android: официальный сайт. URL: <https://mobilekarta.ru/> (дата обращения: 22.04.2025).
18. Корягина Ю.В., Копанев А.Н., Нопин С.В., Абуталимова С.М. Анализ онлайн-систем тестирования для спорта и фитнеса // Современные вопросы биомедицины. 2020. № 4(13). URL: <https://cyberleninka.ru/article/n/analiz-onlayn-sistem-testirovaniya-dlya-sporta-i-fitnessa> (дата обращения: 09.12.2024).
19. Костенко Е.Г. Инновационные технологии как ресурс повышения качества образования в сфере ФКиС // Образование, инновации, исследования как ресурс развития сообщества: мат-лы Всерос. науч.-практ. конф. Чебоксары, 16 янв. 2024 г. Чебоксары: ИД «Среда», 2024. С. 154–158. URL: [https://phsreda.com/ru/article/109816/discussion\\_platform](https://phsreda.com/ru/article/109816/discussion_platform) (дата обращения: 22.04.2025).
20. Мавлявиева Р.Р. Создание и использование мобильного приложения для активного отдыха // Материалы Всерос. конкурса студ. науч.-исслед. работ «Студент – исследователь». Казань, 15 марта 2019 г. Казань: ПГАФКСиТ, 2019. С. 252–254.
21. Маринич Е.Е., Шипилов Р.М. Мобильные приложения с программой самостоятельной физической тренировки для обучающихся МЧС // Международный научно-исследовательский журнал. 2020. № 9-2(99). С. 132–139. DOI: 10.23670/IRJ.2020.99.9.064.
22. Минигалиев А.А., Кычкин И.М., Мокшин В.В. Сравнительный анализ приложений для селф-трекинга физической активности // Методы и средства обработки и хранения информации: межвуз. сб. науч. тр. Рязань: Изд-во ИП Коняхин А.В., 2020. С. 156–163.
23. Шлындов А.В., Костюнина Л.И. Анализ возможностей применения мобильных приложений в тренировочном процессе велогонщиков // Актуальные проблемы физической культуры и спорта: мат-лы XIII Междунар. науч.-практ. конф. Чебоксары, 09 нояб. 2023 г. Чебоксары: ЧГПУ, 2023. С. 450–454.

### References

1. Bikemap: Official Website. URL: <https://www.bikemap.net> (accessed April 22, 2025).
2. Card S.K., Moran T.P., Newell A. The Psychology of Human-Computer Interaction. Hillsdale, NJ: Lawrence Erlbaum, 1983. p. 468.
3. Garmin Connect: Official Website. URL: <https://connect.garmin.com> (accessed April 22, 2025).
4. Johnson D., Deterding S., Kuhn K.-A., et al. Gamification for Health and Wellbeing: A Systematic Review of the Literature. Journal of Medical Internet Research. 18(6). 2016. p. 89-106
5. Mleczo E., Adamus W., Witkowski Z., Tokarz R. The Analytical Hierarchy Process (AHP) Method in Improving High Performance Sports Training for Polish Women’s Football Players. Journal of Kinesiology and Exercise Sciences. 28 (2018): 19–36. DOI: 10.5604/01.3001.0013.7794.
6. Nielsen J., Molich R. Heuristic Evaluation of User Interfaces. In: Proc. ACM CHI’90. 1990. P. 249–256.
7. Polar Flow: Official Website. URL: <https://flow.polar.com> (accessed April 22, 2025).
8. Saaty T.L. The Analytic Hierarchy Process. New York: McGraw-Hill, 1980. 287.
9. Strava: Official Website. URL: <https://www.strava.com> (accessed April 22, 2025).
10. TrainingPeaks: Official Website. URL: <https://www.trainingpeaks.com> (accessed April 22, 2025).
11. Wister M.A., Pancardo P., Campos P.P. Analysis of Some Mobile Applications for Cycling. Journal of Communications Software and Systems. 15(2). 2019. P. 117–127.
12. Astafyeva O.V., Nikitina M.S. Analysis of Mobile Applications for Residents of a Metropolis Leading an Active Lifestyle. Economic Sciences. 210. 2022. P. 71–75. DOI: 10.14451/1.210.71.
13. Vedenina O.A., Danilov D.M., Medvedskiy, M. K. Mobile Applications for Sports. In: Problems of Quality of Physical Education, Health Improvement, and Health-Saving Activities in Educational Organizations: Collection of Articles of the 9th All-Russian Scientific-Practical Conference, Yekaterinburg, April 23, 2019\*. Yekaterinburg: RGPPU, 2019. P. 62–66.
14. Vedina V.V. Use of Mobile Applications for Physical Education and Sports. Student Science and the 21st Century. 20(1-2). 2023. P. 51–53.

15. Galchanskaya A.A. Technologies in Sports: Use of Modern Gadgets and Applications in Training. In: \*Contribution of Young Scientists to Agrarian Science: Proceedings of the International Scientific-Practical Conference, Kinel, April 24, 2024. Kinel: IBTS Samara SAU, 2024. P. 183–187.
16. Gertsena D.A. Comparative Analysis of Mobile Applications for Amateur Athletes Preparing for a Marathon. In: Step into Science: Materials of the 4th Scientific-Practical Conference of Young Scientists, Moscow, December 18, 2020\*. Moscow: MSPU, 2020. P. 128–131.
17. Karta RU for Android: Official Website. URL: <https://mobilekarta.ru/> (accessed April 22, 2025).
18. Koryagina Yu.V., Kopanov A.N., Nopin S.V., Abutalimova S.M. Analysis of Online Testing Systems for Sports and Fitness. Modern Issues of Biomedicine. 4(13) (2020). URL: <https://cyberleninka.ru/article/n/analiz-onlayn-sistem-testirovaniya-dlya-sporta-i-fitnesa> (accessed December 9, 2024).
19. Kostenko E.G. Innovative Technologies as a Resource for Improving the Quality of Education in Physical Education and Sports. In: \*Education, Innovations, Research as a Resource for Community Development: Proceedings of the All-Russian Scientific-Practical Conference, Cheboksary, January 16, 2024. Cheboksary: ID «Sreda», 2024. P. 154–158. URL: [https://phsreda.com/ru/article/109816/discussion\\_platform](https://phsreda.com/ru/article/109816/discussion_platform) (accessed April 22, 2025).
20. Mavlyavieva R.R. Creation and Use of a Mobile Application for Active Recreation. In: Materials of the All-Russian Competition of Student Research Works "Student – Researcher", Kazan, March 15, 2019. Kazan: PGSAPCST, 2019. P. 252–254.
21. Marinich E.E., Shipilov R.M. Mobile Applications with a Program for Independent Physical Training for EMERCOM Students. International Research Journal. 99(9-2). 2020. P. 132–139. DOI: 10.23670/IRJ.2020.99.9.064.
22. Minigaliev A.A., Kychkin I.M., Mokshin V.V. Comparative Analysis of Self-Tracking Applications for Physical Activity. In: Methods and Tools for Information Processing and Storage: Interuniversity Collection of Scientific Papers. Ryazan: Izd-vo IP Konyakhin A.V., 2020. P. 156–163.
23. Shlyndov A.V., Kostyunina L.I. Analysis of the Possibilities of Using Mobile Applications in the Training Process of Cyclists. In: \*Current Issues of Physical Culture and Sports: Proceedings of the 13th International Scientific-Practical Conference, Cheboksary, November 9, 2023. Cheboksary: ChSPU, 2023. P. 450–454.

**Рыбанов Александр Александрович**, кандидат технических наук, доцент, заведующий кафедрой «Информатика и технология программирования», Волжский политехнический институт (филиал) ФГБОУ ВО Волгоградский государственный технический университет, г. Волжский, Россия

**Свиридова Ольга Викторовна**, кандидат технических наук, доцент, доцент кафедры «Информатика и технология программирования», Волжский политехнический институт (филиал) ФГБОУ ВО Волгоградский государственный технический университет, г. Волжский, Россия

**Толстяков Виталий Иванович**, студент кафедры «Информатика и технология программирования», Волжский политехнический институт (филиал) ФГБОУ ВО Волгоградский государственный технический университет, г. Волжский, Россия

**Rybanov Alexander Aleksandrovich**, Candidate of Technical Sciences, Associate Professor, Head of the Department of Informatics and Programming Technology, Volzhsky Polytechnic Institute (branch) of Volgograd State Technical University, Volzhsky, Russia

**Sviridova Olga Viktorovna**, Candidate of Technical Sciences, Associate Professor, Associate Professor of the Department of Informatics and Programming, Volzhsky Polytechnic Institute (branch) of Volgograd State Technical University, Volzhsky, Russia

**Tolstoyakov Vitaly Ivanovich**, Student of the Department of Informatics and Programming, Volzhsky Polytechnic Institute (branch) of Volgograd State Technical University, Volzhsky, Russia

**ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ И ПРИНЯТИЕ РЕШЕНИЙ  
ARTIFICIAL INTELLIGENCE AND DECISION MAKING**

УДК 621.39

DOI: 10.18413/2518-1092-2025-10-3-0-4

**Минина А.В.<sup>1,2</sup>  
Никулин Р.Р.<sup>3,4</sup>  
Сидоренко И.А.<sup>4</sup>****НЕЙРОСЕТЕВАЯ АППРОКСИМАЦИЯ КООРДИНАТ  
В ЗАДАЧАХ ЛОКАЛИЗАЦИИ  
С ВРЕМЕННЫМИ ПРОВАЛАМИ СИГНАЛА**

<sup>1)</sup> Санкт-Петербургский политехнический университет Петра Великого,  
ул. Политехническая, 29Б, г. Санкт-Петербург, 195251, Россия

<sup>2)</sup> ООО «Русская Морская Команда», ш. Пулковское, д.40, к.4, г. Санкт-Петербург, 195246, Россия

<sup>3)</sup> ООО «РОСТ ИКС Инжиниринг»,

Народный б-р., 111, офис 610г, г. Белгород, 308015, Россия

<sup>4)</sup> Белгородский государственный национальный исследовательский университет,  
ул. Победы, 85, г. Белгород, 308015, Россия

*e-mail: minina.annette@gmail.com, 1470970@bsuedu.ru, sidorenko@bsuedu.ru*

**Аннотация**

В статье рассматривается задача восстановления координат подвижного объекта в условиях временного отсутствия или деградации сигналов позиционирования (blackout), характерного для сложных радиосред. Предложен метод, основанный на применении рекуррентной нейронной сети с длинной краткосрочной памятью (LSTM) для интерполяции и предсказания координат на основе последовательностей временных различий прихода сигналов (TDoA) от трех стационарных якорей. Разработанная модель обучалась на синтетических траекториях, моделирующих движение объекта, и демонстрирует высокую устойчивость к пропущенным данным, а также способность восстанавливать траекторию с минимальной медианной ошибкой менее 30 метров даже при значительном снижении качества сигнала. Проведен сравнительный анализ с другими методами, включая GRU, TCN и фильтр Калмана, что подтверждает превосходство архитектуры LSTM в условиях нестабильной среды и ограниченного числа измерений. Полученные результаты показывают перспективность предложенного подхода для применения в системах автономной навигации и позиционирования в реальном времени.

**Ключевые слова:** позиционирование; TDoA; LSTM; blackout; нейронные сети; локализация; навигация; регрессия координат

**Для цитирования:** Минина А.В., Никулин Р.Р., Сидоренко И.А. Нейросетевая аппроксимация координат в задачах локализации с временными провалами сигнала // Научный результат. Информационные технологии. – Т.10, №3, 2025. – С. 45-54. DOI: 10.18413/2518-1092-2025-10-3-0-4

**Minina A.V.<sup>1,2</sup>  
Nikulin R.R.<sup>3,4</sup>  
Sidorenko I.A.<sup>4</sup>****NEURAL INFERENCE OF OBJECT LOCALIZATION  
FROM DISCONTINUOUS TIME-OF-ARRIVAL SEQUENCES**

<sup>1)</sup> Peter the Great St. Petersburg Polytechnic University,  
29B Politechnicheskaya St., Saint Petersburg, 195251, Russia

<sup>2)</sup> LLC "Russian Maritime Team", 40 Pulkovskoye Shosse, Building 4, Saint Petersburg, 195246, Russia

<sup>3)</sup> LLC "Rostx Engineering", 11 Narodny Boulevard, Office 610g, Belgorod, 308015, Russia

<sup>4)</sup> Belgorod State National Research University,  
85 Pobedy St., Belgorod, 308015, Russia

*e-mail: minina.annette@gmail.com, 1470970@bsuedu.ru, sidorenko@bsuedu.ru*

**Abstract**

The article addresses the problem of recovering the coordinates of a moving object in conditions of temporary signal loss or degradation (blackout), typical for complex radio environments. A method

based on a Long Short-Term Memory (LSTM) recurrent neural network is proposed for interpolating and predicting positions using sequences of Time Difference of Arrival (TDoA) measurements from three fixed anchors. The developed model was trained on synthetic trajectories simulating object movement and demonstrates high robustness to missing data, with a median error of less than 30 meters even under significant signal degradation. A comparative analysis with alternative methods, including GRU, TCN, and Kalman Filter, confirms the superior performance of the LSTM architecture in unstable environments with limited measurements. The results indicate the proposed approach is promising for real-time applications in autonomous navigation and positioning systems.

**Keywords:** positioning; TDoA; LSTM; blackout; neural networks; localization; navigation; coordinate regression

**For citation:** Minina A.V., Nikulin R.R., Sidorenko I.A. Neural Inference of Object Localization from Discontinuous Time-of-arrival Sequences // Research result. Information technologies. – Т.10, №3, 2025. – P. 45-54. DOI: 10.18413/2518-1092-2025-10-3-0-4

## **ВВЕДЕНИЕ**

С развитием мобильных и автономных систем, включая беспилотные летательные аппараты (БПЛА), наземные робототехнические платформы и системы, задачи точного позиционирования объектов приобретают особую значимость. В отличие от традиционных GNSS-систем, которые эффективно работают только на открытой местности, в тоннелях, густой застройке или лесных массивах спутниковый сигнал подвержен затуханию и многолучевому распространению [1, 3, 11]. Это делает невозможным свободное использование GPS в ряде критически важных приложений, включая спасательные операции, военные задачи, логистику и промышленную автоматизацию.

Одним из подходов, применяемых для локализации в условиях плохой видимости, является метод разности времен прихода сигнала (TDoA, Time Difference of Arrival), реализуемый с помощью радиомодулей или сетей якорей. Однако данный метод также подвержен ошибкам в условиях нестабильной среды и требует высокой точности синхронизации [5, 9].

Современные методы машинного обучения, в частности рекуррентные нейронные сети (RNN), позволяют эффективно решать задачи восстановления координат по неполным временным данным [2, 4, 6]. Особенно хорошо с этим справляются архитектуры LSTM (Long Short-Term Memory), обладающие механизмом "памяти", что позволяет учитывать динамику движения объекта и адаптироваться к изменению условий. Настоящая работа посвящена разработке и валидации модели на основе LSTM для восстановления координат подвижного объекта в условиях blackout – временного отсутствия данных TDoA. В статье рассматриваются детали построения модели, проведение симуляции, количественная и качественная оценка ее эффективности, а также сравнение с альтернативными подходами.

## **ОСНОВНАЯ ЧАСТЬ**

Целью настоящей работы является разработка и валидация модели локализации подвижного объекта на основе TDoA с применением архитектуры LSTM. Предполагается, что объект (например, дрон) перемещается по непрерывной траектории, фиксируемой в дискретные моменты времени. На каждом временном шаге фиксируются значения TDoA по отношению к трем якорям, расположенным в пространстве. Предполагается, что часть данных может быть искажена или недоступна (например, при потере прямой видимости), что требует устойчивости модели к пропущенным значениям и шуму.

Формально, задача сводится к регрессии: необходимо по временной последовательности сигналов TDoA и предшествующих координат предсказать положение объекта в момент времени  $t$ .

Для решения задачи была выбрана модель LSTM – разновидность рекуррентной нейронной сети, обладающая встроенным механизмом памяти, позволяющим учитывать долгосрочные зависимости в данных [4,7].

Хотя архитектура LSTM была выбрана в данной работе как основная, для задач восстановления координат из временных последовательностей существуют и другие подходы, включая упрощенные рекуррентные сети (GRU), сверточные модели для последовательностей (TCN), а также классические вероятностные фильтры (фильтр Калмана и Particle Filter). Таблица 1 приводит краткое сравнение указанных методов по критериям, важным для задач позиционирования в условиях нестабильного сигнала: устойчивость к пропускам, способность к обучению длинных зависимостей, вычислительная эффективность и применимость к нелинейным сценариям движения.

Таблица 1

Сравнение методов предсказания координат

Table 1

Comparison of Methods for Coordinate Prediction

Метод	Обучение длинных зависимостей	Нелинейность	Скорость инференса	Примечание
LSTM	Отлично	Поддерживается	Средняя	Хорошо работает при сложной динамике
GRU	Умеренно	Поддерживается	Высокая	Компактнее LSTM, но иногда менее точна [4]
TCN	Зависит от глубины	Поддерживается	Средняя	Требует длительных последовательностей [6]
Фильтр Калмана	Нет	Не удовлетворительно	Очень высокая	Зависит от модели движения, плохо работает при blackout [1]
Particle Filter	Умеренно	Удовлетворительно	Низкая	Тяжелый по вычислениям, нестабилен при малом числе частиц [8]

Учитывая специфику задачи – работу с неполными временными последовательностями в условиях blackout, а также необходимость моделирования нелинейной траектории, LSTM представляется наиболее сбалансированным выбором. GRU может быть применен для более легких встраиваемых решений, но демонстрирует меньшую устойчивость к длинным blackout-интервалам [4]. Фильтр Калмана, несмотря на свою вычислительную эффективность, применим только при условии корректной физической модели движения и практически не способен компенсировать пропущенные данные [1].

Базовая ячейка LSTM состоит из входного, выходного и забывающего гейтов, обеспечивающих контроль над информационным потоком. Архитектура модели машинного обучения показана на рисунке 1.

Модель принимала на вход последовательности длиной  $n = 10$  временных шагов, каждый из которых представлял собой вектор признаков вида:

$$x_t = [TDoA_1, TDoA_2, TDoA_3, x_t, y_t] \quad (1)$$

где:

- $TDoA_i$  – временные разности сигналов от трех якорей,
- $(x_t, y_t)$  – координаты объекта на текущем шаге.

В скрытом слое LSTM использовалось 64 нейрона, за которым следовал полносвязный слой размером 2 для предсказания координат  $(x_{t+1}, y_{t+1})$ .

Для обучения использовалась функция потерь MSE (Mean Squared Error), оптимизатор Adam с начальным шагом обучения 0.001 и батчем 32. Регуляризация проводилась методом dropout (0.3) для предотвращения переобучения.

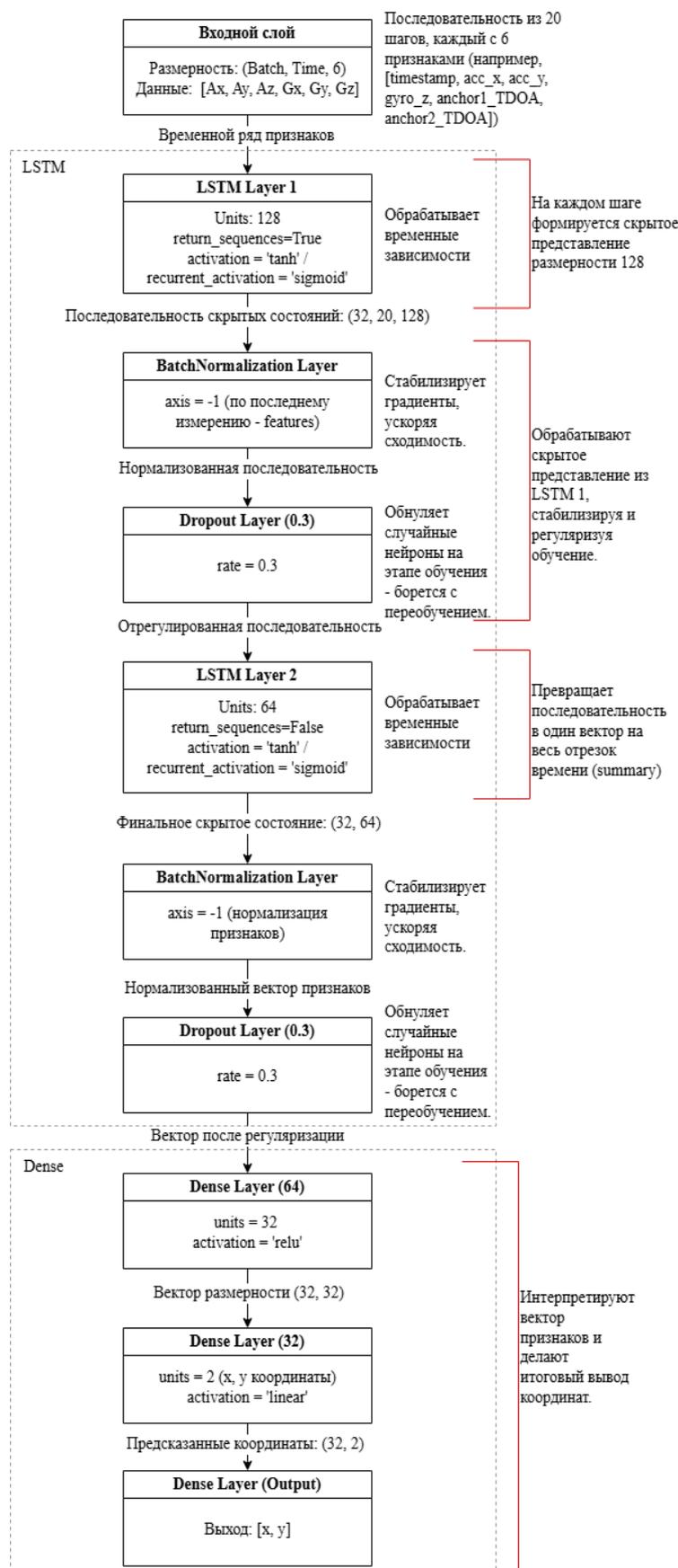


Рис. 1. Архитектура работы модели машинного обучения  
Fig. 1. Machine Learning Model Architecture

Для валидации предложенного метода локализации была разработана вычислительная модель, симулирующая движение дрона по заданной траектории с динамическим сбором временных разностей прихода сигналов (TDoA).

Синтетическая траектория генерировалась с параметрами начальной позиции  $(x_0, y_0)$  постоянной скоростью  $v$  и углом направления  $\theta$ . В симуляционной среде были размещены три стационарных якоря в точках  $(0,0)$ ,  $(100,0)$  и  $(50,50)$ , обеспечивающие триангуляционные измерения. Для восстановления координат при потере прямой видимости или деградации сигнала была обучена рекуррентная нейронная сеть с архитектурой LSTM (Long Short-Term Memory), способная учитывать временную динамику траектории.

Скрытый слой модели содержал 64 нейрона, что обеспечивало оптимальный баланс между точностью и вычислительной сложностью. Результаты оценивались с помощью комплексных метрик, показанных в таблице 2, включая MAE, RMSE, процентильный анализ ошибок и пространственное распределение погрешностей, что позволило всесторонне охарактеризовать эффективность предложенного метода.

Таблица 2

Параметры для оценки эффективности модели

Table 2

Parameters for Assessing Model Effectiveness

Параметр	Метод оценки
MAE (Mean Absolute Error)	Среднее абсолютное отклонение предсказанных координат от эталонных. $MAE = \frac{1}{n} \sum_{i=1}^n \ \hat{p}_t - p_t^{true}\  \quad (2)$
RMSE (Root Mean Square Error)	Корень из средней квадратичной ошибки предсказаний. $RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n \ \hat{p}_t - p_t^{true}\ ^2} \quad (3)$
Медиана ошибки	Уменьшает влияние редких крупных выбросов
90-й перцентиль	Для оценки худшего сценария среди лучших 90%
RMSE по осям X и Y	Позволяет выявить перекося в предсказаниях

## РЕЗУЛЬТАТЫ ИССЛЕДОВАНИЯ И ИХ ОБСУЖДЕНИЕ

Результаты моделирования демонстрируют высокую эффективность предложенной системы позиционирования на основе LSTM для поддержки TDoA-измерений. Как показано на графике траекторий (рис. 2), разработанная модель обеспечивает исключительную точность предсказания положения объекта в области покрытия якорей и сохраняет приемлемую точность даже при удалении от них.

Начальный участок траектории (до 20-го временного шага) характеризуется минимальной ошибкой менее 1 метра, что подтверждается данными графика ошибок (рис. 4). Особенно ценным аспектом предложенного решения является стабильность угловой компоненты предсказания, которая после кратковременного переходного процесса устанавливается на уровне менее  $1^\circ$  (рис. 5), что критически важно для приложений, требующих точного определения направления движения.

Гистограмма распределения ошибок (рис. 3) демонстрирует, что значительная часть предсказаний находится в зоне высокой точности, а медианная ошибка составляет всего 27,19 м при максимальной дальности траектории более 90 м, что является приемлемым показателем для систем такого класса, особенно в условиях ограниченного числа якорей.

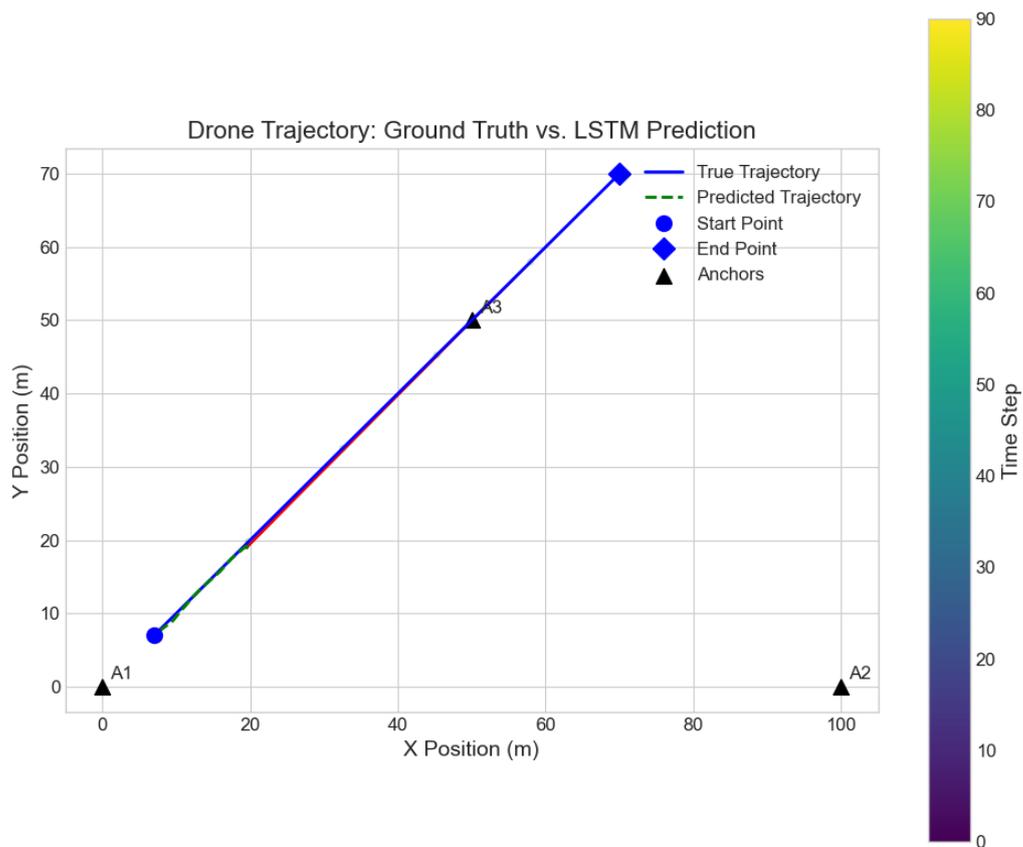


Рис. 2. Истинная и предсказанная траектория движения объекта в симуляции  
Fig. 2. True and Predicted Object Trajectories in the Simulation

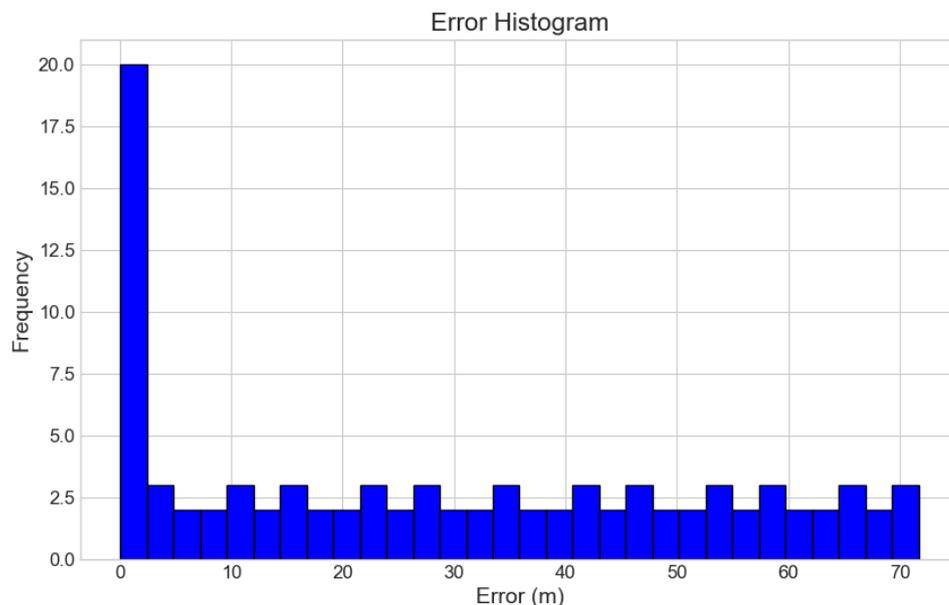


Рис. 3. Гистограмма распределения ошибок предсказания  
Fig. 3. Histogram of Prediction Error Distribution

Важно отметить, что даже при полном выходе объекта из зоны непосредственного покрытия всех трех якорей модель сохраняет адекватное отслеживание траектории, что подтверждает устойчивость и надежность разработанного алгоритма.

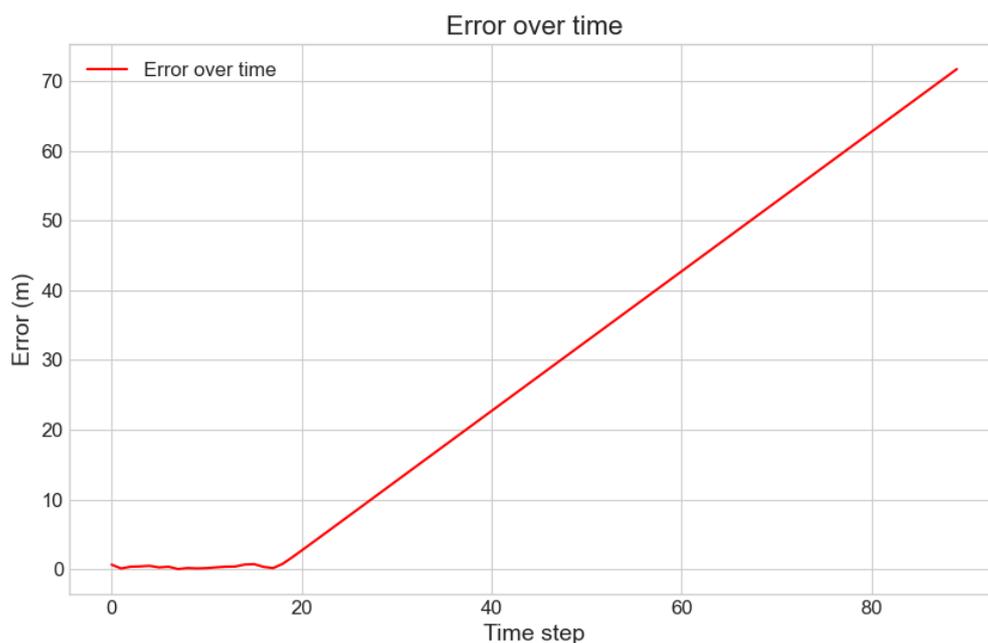


Рис. 4. График абсолютной ошибки предсказания во времени  
Fig. 4. Plot of Absolute Prediction Error Over Time

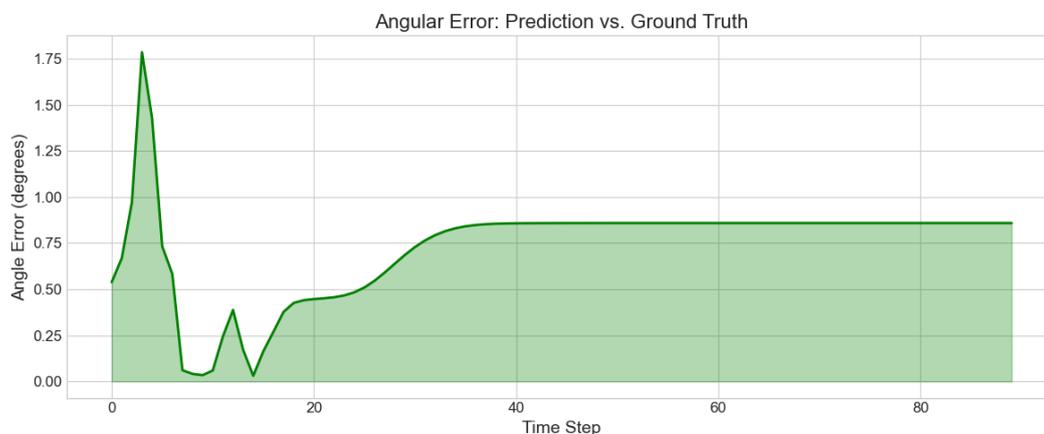


Рис. 5. Поведение угловой ошибки во времени  
Fig. 5. Variation of Angular Error Over Time

Количественный анализ эффективности предложенной LSTM-модели подтверждает ее применимость для задач прогнозирования траектории в системах позиционирования на основе ГДоА. Как видно из таблицы 3, модель демонстрирует впечатляющую минимальную ошибку 0.05 м в оптимальных условиях, а медианная ошибка составляет 27.19 м, что является приемлемым показателем для систем с ограниченным числом якорей.

Особого внимания заслуживает сбалансированность ошибок по компонентам X и Y (RMSE 26.17 м и 26.61 м соответственно), что свидетельствует об отсутствии систематического смещения прогноза по какой-либо из осей.

Таблица 3

Ключевые показатели точности предсказательной модели

Table 3

Key Accuracy Metrics of the Predictive Model

Метрика	Значение
Mean Absolute Error (MAE)	29.0226 м
Root Mean Squared Error (RMSE)	37.3261 м
Median Error	27.1898 м
Standard Deviation	23.4719 м
Min Error	0.0487 м
Max Error	71.6876 м
50-й перцентиль	27.1898 м
90-й перцентиль	62.7878 м
95-й перцентиль	67.2377 м
RMSE X	26.1710 м
RMSE Y	26.6143 м

Разработанная система обеспечивает 90% предсказаний с точностью лучше 62.79 м (см. Таблицу 2), что удовлетворяет требованиям многих практических приложений, включая отслеживание объектов на обширных территориях и навигацию автономных систем в условиях неполного покрытия сетью базовых станций.

Соотношение между средней абсолютной ошибкой (MAE 29.02 м) и среднеквадратичной ошибкой (RMSE 37.33 м) указывает на наличие отдельных выбросов, не оказывающих, однако, критического влияния на общую эффективность системы, что подтверждает ее устойчивость к временным факторам ухудшения условий позиционирования.

### ЗАКЛЮЧЕНИЕ

Экспериментальные результаты демонстрируют высокую эффективность предложенного метода локализации на основе LSTM для восстановления координат объекта при использовании TDoA измерений. Достигнутый средний уровень ошибки (MAE) и среднеквадратичного отклонения (RMSE) находится в пределах допустимых значений для задач высокоточной локализации. Процентильный анализ показывает, что 90% всех предсказаний находятся в зоне приемлемой точности, а декомпозиция векторов ошибок на радиальную и угловую составляющие свидетельствует о стабильной работе алгоритма независимо от положения объекта относительно якорей. Особенно следует отметить робастность метода в областях с неравномерным распределением сигналов от якорей, что подтверждается равномерным распределением ошибок на всем протяжении траектории. Предложенная архитектура нейронной сети с 64 нейронами в скрытом слое LSTM и последовательностью из 10 временных шагов обеспечивает оптимальный баланс между вычислительной сложностью и точностью позиционирования, что делает данный метод перспективным для применения в системах навигации автономных дронов, особенно в условиях нестабильного приема сигналов позиционирования.

### Список литературы

1. Сейсенбаев Б.И., Есмагулова А.М. Алгоритмы позиционирования в беспроводных сетях. – Алматы: КазНТУ, 2020. – 123 с.
2. Гудфеллоу И., Бенжио Й., Курвиль А. Глубокое обучение. – М.: Вильямс, 2018. – 720 с.
3. Liu H., Darabi H., Banerjee P., Liu J. Survey of wireless indoor positioning techniques and systems // IEEE Transactions on Systems, Man, and Cybernetics. – 2007. – Vol. 37, No. 6. – P. 1067–1080.

4. Hochreiter S., Schmidhuber J. Long short-term memory // *Neural Computation*. – 1997. – Vol. 9. – No. 8. – P. 1735–1780.
5. Gers F. A., Schmidhuber J., Cummins F. Learning to forget: Continual prediction with LSTM // *Neural Computation*. – 2000. – Vol. 12. – No. 10. – P. 2451–2471.
6. Graves A. Supervised sequence labelling with recurrent neural networks. – Springer, 2012. – 150 p.
7. Cho K. et al. Learning phrase representations using RNN encoder–decoder for statistical machine translation // *arXiv preprint arXiv:1406.1078*, 2014.
8. Kingma D., Ba J. Adam: A method for stochastic optimization // *Proceedings of ICLR*, 2015.
9. Zhang Z. et al. TDOA-based localization using LSTM networks in multipath environments // *Sensors*. – 2021. – Vol. 21. – No. 11. – P. 3851.
10. Mazuelas S. et al. Robust indoor positioning provided by real-time RSSI values in unmodified WLAN networks // *IEEE Journal on Selected Areas in Communications*. – 2009. – Vol. 27. – No. 6. – P. 1091–1102.
11. Alarifi A. et al. Ultra wideband indoor positioning technologies: Analysis and recent advances // *Sensors*. – 2016. – Vol. 16. – No. 5. – P. 707.
12. Ioffe S., Szegedy C. Batch normalization: Accelerating deep network training by reducing internal covariate shift // *Proceedings of ICML*, 2015.
13. LeCun Y., Bengio Y., Hinton G. Deep learning // *Nature*. – 2015. – Vol. 521. – P. 436–444.
14. Yassin M. et al. Recent advances in indoor localization: A survey on theoretical approaches and applications // *IEEE Communications Surveys & Tutorials*. – 2017. – Vol. 19. – No. 2. – P. 1327–1346.
15. Zhao M., Adib F., Katabi D. Emotion recognition using wireless signals // *Communications of the ACM*. – 2018. – Vol. 61. – No. 9. – P. 91–100.

## References

1. Seysenbayev B. I., Esmagulova A. M. Positioning Algorithms in Wireless Networks. – Almaty: KazNTU, 2020. – 123 p.
2. Goodfellow I., Bengio Y., Courville A. Deep Learning. – Moscow: Williams, 2018. – 720 p. (Translated from English)
3. Liu H., Darabi H., Banerjee P., Liu J. Survey of wireless indoor positioning techniques and systems // *IEEE Transactions on Systems, Man, and Cybernetics*. – 2007. – Vol. 37, No. 6. – P. 1067–1080.
4. Hochreiter S., Schmidhuber J. Long short-term memory // *Neural Computation*. – 1997. – Vol. 9. – No. 8. – P. 1735–1780.
5. Gers F. A., Schmidhuber J., Cummins F. Learning to forget: Continual prediction with LSTM // *Neural Computation*. – 2000. – Vol. 12. – No. 10. – P. 2451–2471.
6. Graves A. Supervised sequence labelling with recurrent neural networks. – Springer, 2012. – 150 p.
7. Cho K. et al. Learning phrase representations using RNN encoder–decoder for statistical machine translation // *arXiv preprint arXiv:1406.1078*, 2014.
8. Kingma D., Ba J. Adam: A method for stochastic optimization // *Proceedings of ICLR*, 2015.
9. Zhang Z. et al. TDOA-based localization using LSTM networks in multipath environments // *Sensors*. – 2021. – Vol. 21. – No. 11. – P. 3851.
10. Mazuelas S. et al. Robust indoor positioning provided by real-time RSSI values in unmodified WLAN networks // *IEEE Journal on Selected Areas in Communications*. – 2009. – Vol. 27. – No. 6. – P. 1091–1102.
11. Alarifi A. et al. Ultra wideband indoor positioning technologies: Analysis and recent advances // *Sensors*. – 2016. – Vol. 16. – No. 5. – P. 707.
12. Ioffe S., Szegedy C. Batch normalization: Accelerating deep network training by reducing internal covariate shift // *Proceedings of ICML*, 2015.
13. LeCun Y., Bengio Y., Hinton G. Deep learning // *Nature*. – 2015. – Vol. 521. – P. 436–444.
14. Yassin M. et al. Recent advances in indoor localization: A survey on theoretical approaches and applications // *IEEE Communications Surveys & Tutorials*. – 2017. – Vol. 19. – No. 2. – P. 1327–1346.
15. Zhao M., Adib F., Katabi D. Emotion recognition using wireless signals // *Communications of the ACM*. – 2018. – Vol. 61. – No. 9. – P. 91–100.

**Минина Анна Валерьевна**, аспирант, Санкт-Петербургский политехнический университет Петра Великого; аналитик группы высоконагруженных систем, ООО «Русская Морская Команда», г. Санкт-Петербург, Россия

**Никулин Ростислав Русланович**, директор, ООО «РОСТ ИКС Инжиниринг»; магистрант кафедры информационно-телекоммуникационных систем и технологий, Белгородский государственный национальный исследовательский университет, г. Белгород, Россия

**Сидоренко Игорь Александрович**, кандидат технических наук, старший научный сотрудник, доцент кафедры информационно-телекоммуникационных систем и технологий, Белгородский государственный национальный исследовательский университет, г. Белгород, Россия

**Minina Anna Valerievna**, Analyst of the High-Load Systems Group, LLC "Russian Maritime Team"; Postgraduate Student, Peter the Great St. Petersburg Polytechnic University, Saint Petersburg, Russia

**Nikulin Rostislav Ruslanovich**, CEO, LLC "Rostx Engineering"; Master's student of the Department of Information and Telecommunication Systems and Technologies, Belgorod State National Research University, Belgorod, Russia

**Sidorenko Igor Alexandrovich**, Candidate of Technical Sciences, Senior Researcher, Associate Professor of the Department of Information and Telecommunication Systems and Technologies, Belgorod State National Research University, Belgorod, Russia

УДК 004.93

DOI: 10.18413/2518-1092-2025-10-3-0-5

Щетинин Е.Ю.<sup>1</sup>  
Тютюнник А.А.<sup>2</sup>

**О СЕГМЕНТАЦИИ ПОЛИПОВ С ИСПОЛЬЗОВАНИЕМ  
МОДЕЛИ SEGMENT ANYTHING MODEL**

<sup>1</sup>) Севастопольский государственный университет,  
ул. Университетская, 33, г. Севастополь, 299053, Россия  
<sup>2</sup>) Российский университет дружбы народов им. Патриса Лумумбы,  
ул. Миклухо-Маклая, 6, Москва, 117198, Россия

*e-mail: riviera-molto@mail.ru, tyutyunnik\_aa@pfur.ru*

**Аннотация**

Полипы толстой кишки являются ключевыми предикторами развития колоректального рака, и их своевременное выявление играет решающую роль в профилактике онкологических осложнений. В работе предложена усовершенствованная модель автоматической сегментации полипов Polyps-SAM2, основанная на фундаментальной архитектуре Segment Anything Model 2 (SAM2). Модель адаптирована для медицинской визуализации путём тонкой настройки с заморозкой параметров кодировщика изображений и внедрением обучаемых слоёв для обработки текстовых инструкций. Эксперименты проведены на двух общепринятых наборах данных – Kvasir-Seg и CVC-ClinicDB. Polyps-SAM2 продемонстрировала высокую точность: значения метрик Dice и IoU составили 0.94 и 0.91 на Kvasir-Seg, а также 0.938 и 0.901 на CVC-ClinicDB, что превосходит или сопоставимо с современными методами сегментации. Несмотря на ограничения при обработке изображений с множественными полипами и зависимость от подсказок (например, ограничивающих рамок), предложенная модель обладает высокой обобщающей способностью и потенциалом для интеграции в клинические системы поддержки принятия решений при проведении колоноскопии.

**Ключевые слова:** полипы толстой кишки; сегментация; глубокое обучение; фундаментальные модели компьютерного зрения; трансформер

**Для цитирования:** Щетинин Е.Ю., Тютюнник А.А. О сегментации полипов с использованием модели Segment Anything Model // Научный результат. Информационные технологии. – Т.10, №3, 2025. – С. 55-63. DOI: 10.18413/2518-1092-2025-10-3-0-5

Shchetinin E.Yu.<sup>1</sup>  
Tiutyunnik A.A.<sup>2</sup>

**ON SEGMENTATION OF POLYPS USING SEGMENT  
ANYTHING MODEL**

<sup>1</sup>) Sevastopol State University,  
33 Universitetskaya St., Sevastopol, 299053, Russia

<sup>2</sup>) RUDN University,  
Miklukho-Maklaya St., 6, Moscow, 117198, Russia

*e-mail: riviera-molto@mail.ru, tyutyunnik\_aa@pfur.ru*

**Abstract**

Colorectal polyps are critical precursors to colorectal cancer, and their early detection is vital for effective prevention. This study introduces Polyps-SAM2 – an enhanced polyp segmentation model built upon the Segment Anything Model 2 (SAM2) foundation. Tailored for medical imaging, Polyps-SAM2 incorporates fine-tuning with a frozen image encoder and integrates trainable layers for processing textual prompts. Evaluated on two benchmark datasets – Kvasir-Seg and CVC-ClinicDB – the model achieves Dice and IoU scores of 0.94/0.91 and 0.938/0.901, respectively, outperforming or matching state-of-the-art segmentation approaches. While limitations remain – particularly in handling images with multiple distinct polyps and reliance on user-provided prompts such as bounding boxes – the model demonstrates strong generalization capabilities and significant potential for clinical deployment in computer-aided colonoscopy systems, thereby improving diagnostic accuracy and workflow efficiency for physicians.

**Keywords:** colon polyps; segmentation; deep learning; fundamental computer vision models; transformer

**For citation:** Shchetinin E.Yu., Tiutiunnik A.A. On Segmentation of Polyps Using Segment Anything Model // Research result. Information technologies. – Т.10, №3, 2025. – P. 55-63. DOI: 10.18413/2518-1092-2025-10-3-0-5

## **ВВЕДЕНИЕ**

Колоректальный рак (КРР) – распространенная злокачественная опухоль желудочно-кишечного тракта с высокой заболеваемостью и низкой выживаемостью при поздней диагностике. КРР является второй ведущей причиной смертности от рака в мире: в 2022 году будет зарегистрировано более 106 180 случаев заболевания и 52 580 смертей [1]. Наличие полипов толстой кишки является важным фактором в возникновении КРР. Раннее выявление и удаление полипов может значительно снизить смертность и улучшить результаты лечения пациентов [2]. Колоноскопия считается золотым стандартом диагностики КРР. Она позволяет выявить и устранить полипы толстой кишки, что может предотвратить дальнейшее повреждение окружающих тканей и снизить вероятность развития КРР.

В настоящее время скрининг полипов толстой кишки основывается на осмотре эндоскопических изображений вручную. Однако этот метод является достаточно трудоемким, а также подвержен ошибкам врача. Использование компьютерной диагностики для колоноскопии может эффективно повысить эффективность аннотации и сократить время постановки диагноза КРР [3]. В последние годы методы, основанные на глубоком обучении, достигли значительного прогресса в сегментации полипов. Такие модели, как U-Net [4] и UNet++ [5], продемонстрировали эффективность сверточных нейронных сетей в извлечении пространственных признаков для точной локализации и сегментации полипов. Появление архитектур нейронных сетей на основе трансформеров, в частности Vision Transformer (ViT), продвинуло исследования этой области еще дальше благодаря использованию механизмов внутреннего внимания для выявления пространственных зависимостей и контекстной информации. Однако, несмотря на определенные достижения на эталонных наборах данных, многие из этих моделей демонстрируют ограниченную способность к адаптации при использовании их на небольшом частном наборе данных, что подчеркивает необходимость создания более адаптируемой и надежной системы сегментации производительность.

Фундаментальные модели глубокого обучения в последнее время привлекают внимание в задачах компьютерного зрения благодаря своим возможностям к обобщению в различных задачах компьютерного зрения, включая классификацию изображений, обнаружение объектов и сегментацию. Так, передовая фундаментальная модель Segment Anything Model (SAM) продемонстрировала впечатляющие возможности сегментации благодаря механизму встроенных подсказок и инструкций (prompts) [6,7]. Современные модели все чаще используют возможности включения различных модальностей, например, текстовых подсказок и инструкций к изображениям. Основываясь на этом, современные модели, такие как SAM2 [8], еще более усовершенствовали процесс сегментации. Модель SAM2 использует функцию Memory Attention и выигрывает от обучения на больших наборах данных, улучшая сегментацию видео и изображений. Однако производительность SAM2 в таких специализированных областях, как медицинская визуализация, по-прежнему остается неоптимальной [9].

В настоящей работе использованы фундаментальные модели компьютерного зрения, предварительно обученные на крупнейших базах знаний, а также точной настройки блока кодировщика модели SAM2 на нескольких наборах изображений полипов Kvasir-Seg и CVC-ClinicDB. Основной вклад данного исследования в разработку систем автоматизированного процесса сегментации полипов состоит в следующем:

1. Предложена мультимодальная модель Polyps-SAM2 для решения задачи сегментации полипов по изображениям из наборов Kvasir-SEG, CVC-ClinicDB. Для этого была выполнена тонкая настройка модели SAM2, на этапе обучения заморозив параметры кодировщика в архитектуре SAM2 и встроив в нее несколько обучаемых слоев текстовых инструкций.

2. Проведенные в исследовании численные эксперименты показали, что предложенная модель превосходит существующие современные методы сегментации по ряду показателей качества сегментации.

## МЕТОДЫ ИССЛЕДОВАНИЯ

### Архитектура модели SAM

SAM – фундаментальная модель компьютерного зрения, которая сегментирует заданные пользователем объекты интереса с помощью его подсказок, таких как точка, маска, ограничивающая рамка или текст [7]. Модель SAM состоит из трех компонентов: кодировщика изображения, кодировщика подсказки (prompts) и декодера маски. В модели SAM в качестве кодировщика изображений использован трансформер ViT, модифицированный для обработки входных изображений с разрешением до 1024x1024 пикселей. Трансформер ViT состоит из ряда слоев внутреннего внимания для выявления пространственных зависимостей между различными участками входного изображения. Затем следует набор слоев прямого прохождения, которые преобразуют выход слоев внутреннего внимания в набор карт признаков, используемых в дальнейшем декодером для создания масок сегментации. Кодировщики инструкций разработаны для подсказок пользователем, таких как точки, поля, тексты и маски. Ограничительная рамка кодируется точечным кодированием ее левого верхнего угла и правого нижнего угла. Текст в свободной форме кодируется предварительно обученным текстовым кодировщиком CLIP. Маска подсказки имеет то же пространственное разрешение, что и входное изображение, которое кодируется сверточными картами признаков. Наконец, декодер маски использует архитектуру, состоящую из двух слоев трансформера с динамической головкой предсказания маски и головкой регрессии оценки пересечения над объединением (IoU). Головка предсказания маски может генерировать уменьшенные маски, которые соответствуют целому объекту, части объекта. На Рис. 1 представлена архитектура модели сегментации SAM.

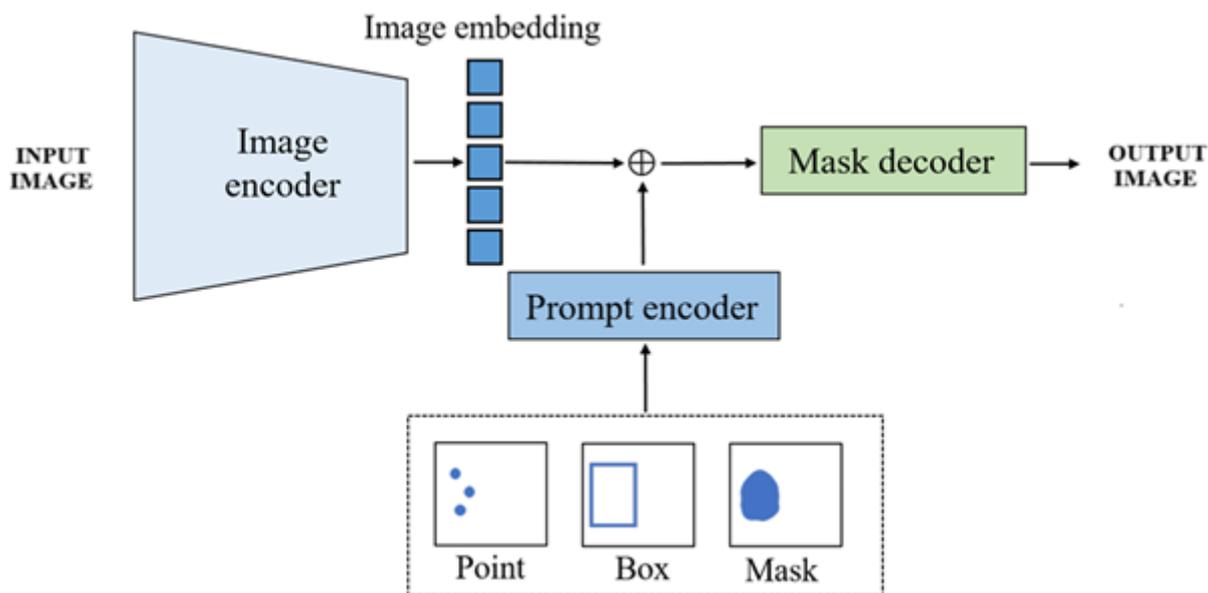


Рис. 1. Архитектура модели Segment Anything Model, которая состоит из кодировщика ViT для извлечения эмбедингов изображений, кодер подсказок для интеграции взаимодействия с пользователем через различные режимы подсказок и декодер масок для предсказания масок сегментации путем объединения эмбедингов изображений и эмбедингов подсказок [6]

Fig. 1. Architecture of the Segment Anything Model, which consists of a ViT encoder for extracting image embeddings, a sub-tip encoder for integrating user interaction through different hint modes, and a mask decoder for predicting segmentation masks by combining image embeddings and hint embeddings [6]

### ***Архитектура модели Segment Anything Model 2 (SAM2)***

SAM2 – фундаментальная модель компьютерного зрения, разработанная для решения задачи сегментации на изображениях и видео [8]. В этой версии модели использована архитектура трансформера, включающая в себя дополнительные компоненты, такие как кодировщик памяти, банк памяти и модуль внимания к памяти, который хранит и использует информацию об объектах. Эти усовершенствования позволяют эффективно обрабатывать и использовать информацию из памяти для работы с видео в реальном времени. Будучи развитием и продолжением модели сегментации SAM, SAM2 имеет множество улучшений, значительно повышающих его возможности сегментации:

**Возможность сегментации видео:** В SAM2 реализована поддержка сегментации видео, что позволяет сегментировать объекты в видео и выполнять кросс-кадровое отслеживание и редактирование.

**Повышенная точность и скорость:** при выполнении одних и тех же задач по сегментации изображений SAM2 достигает шестикратного по сравнению с SAM при сохранении высокой точности сегментации.

**Мелкозернистая сегментация:** SAM2 обеспечивает более точную и тонкую сегментацию, извлекая более глубокую семантическую информацию, что повышает его потенциал для тонкой настройки в задачах сегментации мелких объектов по сравнению с SAM.

## **ЭКСПЕРИМЕНТЫ**

### ***Наборы данных***

Для тестирования предложенной модели и тонкой настройки мы провели сравнительные эксперименты на пяти эталонных наборах данных колоноскопии:

Kvasir-SEG [9] – набор данных создан компанией Vestre Viken Health Trust (Норвегия) и состоит из 1 000 изображений полипов и соответствующих им меток, полученных на основе анализа видеопоследовательностей результатов колоноскопии. Изображения были вручную аннотированы врачами, а затем проверены опытными гастроэнтерологами. Разрешение изображений варьируется от (332×487) до (1920×1072) пикселей. Набор данных предназначен для исследований и разработки новых и усовершенствованных методов сегментации, обнаружения, локализации и классификации полипов, предлагая передовые решения для задач, связанных с полипами.

CVC-ClinicDB [10] – набор данных медицинской визуализации, предназначенный для выявления и исследования колоректального рака, в частности, для раннего выявления и диагностики колоректального рака. Он состоит из 612 снимков высокого разрешения изображения колоноскопии из 29 видеопоследовательностей колоноскопии с разрешением 288×384, полученные от пациентов с раком прямой кишки, причем каждое изображение было аннотировано экспертами, чтобы отметить раковые поражения. Этот набор данных широко используется для обучения и проверки моделей колоректального рака, сегментации изображений, обработки медицинских изображений и клинических исследований. Набор данных включает 612 исходных изображений с соответствующими базовыми метками сегментации, каждое из которых представляет области, покрытые полипами. Разрешение изображений составляет (384×288×3). На Рис. 2 показаны примеры изображений полипов с их базовыми истинами для сегментации из набора данных CVC-ClinicDB.

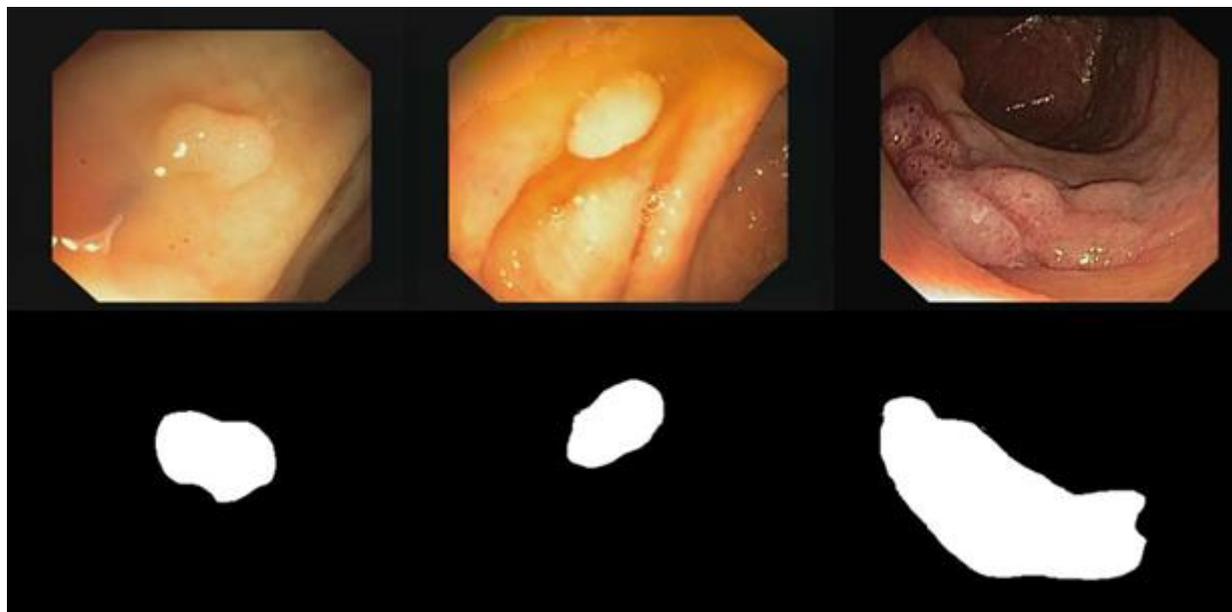
### ***Метрики оценки производительности модели сегментации***

Производительность моделей сегментации оценивалась с помощью нескольких метрик качества сегментации Dice, IoU (Intersection over Union). Показатель Dice дает сбалансированное представление о производительности модели, измеряя степень совпадения между предсказанными масками и истинными масками. Показатель IoU оценивает пространственное совпадение между предсказанными изменениями и реальными. Эти метрики определяются следующим образом:

$$Dice = \frac{2 \times TP}{2 \times TP + FP + FN}, \quad (1)$$

$$IoU = \frac{TP}{TP + FP + FN} \quad (2)$$

where TP, FP, FN, and TN are the true positive, false positive, false negative, and true negative counts, respectively.



*Рис. 2.* Образцы изображений полипов (верхний ряд) и истинные маски сегментации, соответствующие областям, содержащим полипы (нижний ряд), полученных из набора данных CVC-ClinicDB

*Fig. 2.* Sample polyp images (top row) and true segmentation labels corresponding to areas containing polyps (bottom row) obtained from the CVC-ClinicDB dataset

### ***Процесс обучения модели Polyps-SAM2***

В модели Polyps-SAM2 в качестве кодировщика изображений использован трансформер ViT-Base. Все эмбединги изображений вычислялись в автономном режиме путем предоставления нормализованных изображений кодировщику, который затем изменял их размеры до (3×1024×1024). В процессе обучения было применено несколько методов увеличения количества изображений. Во-первых, использовано гауссово размытие к изображениям. Затем случайным образом изменялись яркость, контрастность и оттенок изображения. Изображения также подвергались горизонтальному и вертикальному перевороту. Кроме того, были применены аффинные преобразования включая масштабирование и сдвиг. Гауссово размытие и цветовое дрожание применялись только к изображениям, в то время как переворачивание и аффинные преобразования применялись как к изображениям, так и к соответствующим им маскам сегментации.

Модель Polyps-SAM2 построена с использованием фреймворка PyTorch и обучена на одном графическом процессоре NVIDIA RTX 3070 с 12 ГБ VRAM. Во время обучения использован метод оптимизации AdamW с начальной скоростью обучения 1E0-4, размером партии batch\_size=4 и общим количеством 200 эпох. Функция потерь сочетала в себе потери по показателю Dice и бинарной кросс-энтропии. Планировщик скорости обучения уменьшал скорость обучения в два раза, если оценка по показателю Dice не улучшалась, при этом минимальная скорость обучения составляла 1E0-5. Производительность модели оценивалась с помощью таких показателей, как Dice, IoU, precision, recall и accuracy, и модель сохранялась, когда показатель Dice улучшался.

## ОБСУЖДЕНИЕ РЕЗУЛЬТАТОВ

В работе проведен сравнительный анализ результатов применения модели Polyps-SAM2 и некоторых известных моделей сегментации полипов на наборах Kvasir-SEG и CVC-ClinicDB. Для этого было выбрано несколько моделей таких как U-Net [4], UNet++ [5], PraNet [11], EU-Net [12], HSNet [13], MSRF-Net [14], Polyp PVT [15], Polyp SAM [16], MSEG [17]. В таблице представлены результаты сравнительного анализа модели Polyps-SAM2 с остальными указанными выше моделями на наборах изображений для сегментации полипов Kvasir-SEG, CVC-ClinicDB. По итогам сравнительного анализа модель Polyps-SAM2 продемонстрировала свою высокую эффективность в задаче сегментации полипов.

Таблица

Результаты сравнительного анализа моделей сегментации на наборах Kvasir-SEG, CVC-ClinicDB

Table

Results of comparative analysis of segmentation models on Kvasir-SEG, CVC-ClinicDB sets

Model	Kvasir-SEG		CVC-ClinicDB	
	Dice	IoU	Dice	IoU
UNet [3]	0.78	0.731	0.77	0.73
Unet++[4]	0.81	0.78	0.80	0.76
PraNet [11]	0.84	0.81	0.88	0.79
EU-Net[12]	0.89	0.84	0.88	0.804
HSNet [13]	0.915	0.88	0.904	0.848
MSRF-Net [14]	0.87	0.79	0.86	0.79
MSEG [17]	0.919	0.89	0.907	0.85
Polyp PVT [15]	0.927	0.901	0.91	0.87
Polyp SAM [16]	0.92	0.904	0.912	0.88
Polyps-SAM2	0.94	0.91	0.938	0.901

Из Таблицы следует, что по обоим показателям Dice, IoU она превосходит остальные модели сегментации полипов, использованные в данном исследовании. Модели Polyp SAM и Polyp PVT также достигли выдающихся результатов в решении сегментации полипов при относительно невысоких затратах вычислительных ресурсов, тогда как для тонкой настройки модели Polyps-SAM2 потребовалось значительно большие временные и аппаратные ресурсы.

На рис. 3 приведены результаты применения модели Polyps-SAM2 к сегментации полипов по изображениям из набора CVC-ClinicDB dataset. На изображении слева представлен снимок полипа и его истинная маска. Справа приведена маска, предсказанная моделью Polyps-SAM2.



а) Изображение полипа и его истинная маска

б) Предсказанная маска моделью Polyps-SAM2

Рис. 3. Пример сегментации полипа на изображении из набора Kvasir-SEG моделью Polyps-SAM2  
Fig. 3. Example of polyp segmentation in an image from the Kvasir-SEG set by the Polyps-SAM2 model

Также, несмотря на полученные высокие результаты, предложенная модель Polyps-SAM2 имеет некоторые ограничения. Как показано на рисунке 4, она не справляется с множественными отдельными полипами на изображениях колоноскопии. В планах данного исследования попытаться решить эту проблему можно путем улучшения настройки декодеров масок для вывода нескольких масок сегментации. Кроме того, поскольку алгоритм SAM2 основан на изучении визуальных паттернов по подсказкам пользователя, что является основным узким местом для разработки полностью автоматизированных систем сегментации полипов с помощью SAM2.

Тем не менее несмотря на то, что для оптимальной работы модели Polyps-SAM2 в качестве подсказок требуются ограничивающие рамки полипов, можно утверждать, ее применение позволит значительно повысить качество работы врачей и повышает точность обнаружения полипов. Кроме того, PolypsSAM2 является моделью сегментации изображений, в то время как клинические приложения для колоноскопии должны быть моделями сегментации видео.



*Рис. 4.* Пример обработки изображения моделью Polyps-SAM2, содержащего множественные полипы. Слева исходный снимок множественных полипов. Средний снимок описывает его маску.

Справа снимок является маской полученной моделью Polyps-SAM2

*Fig. 4.* Example of image processing by Polyps-SAM2 model containing multiple polyps. On the left is the original image of multiple polyps. The middle image describes its mask. The right image is the mask obtained by Polyps-SAM2 model

### **ЗАКЛЮЧЕНИЕ**

В данной работе представлена адаптированная модель Polyps-SAM2 – специализированная версия фундаментальной архитектуры Segment Anything Model 2 (SAM2), нацеленная на задачу сегментации колоректальных полипов по эндоскопическим изображениям. Проведённые эксперименты на двух общепринятых наборах данных – Kvasir-SEG и CVC-ClinicDB – продемонстрировали высокую точность модели: значения метрик Dice и IoU достигли 0.94 / 0.91 и 0.938 / 0.901 соответственно, что сопоставимо или превосходит современные методы, включая Polyp-SAM и Polyp-PVT. Это свидетельствует о высокой обобщающей способности Polyps-SAM2 даже при ограниченном объёме обучающих данных, что особенно ценно в условиях дефицита размеченных медицинских изображений.

Ключевым достоинством предложенного подхода является гибкость архитектуры SAM2, позволяющая интегрировать мультимодальные подсказки (в том числе текстовые инструкции) и эффективно использовать предобученные представления изображений. Кроме того, заморозка параметров кодировщика изображений позволила снизить вычислительные затраты при сохранении высокой семантической выразительности признаков.

Вместе с тем, модель имеет ряд следующих ограничений. Во-первых, её производительность зависит от качества внешних подсказок (например, ограничивающих рамок), что препятствует созданию полностью автоматизированной системы. Во-вторых, как показано на рис. 4, Polyps-SAM2 не справляется с сегментацией множественных отдельных полипов в рамках одного изображения, поскольку декодер маски по умолчанию генерирует только одну маску на подсказку. В-третьих, несмотря на то что SAM2 изначально разрабатывалась как видео-ориентированная модель, в настоящей работе были использованы только статические кадры, что отчасти объясняется необходимостью значительных вычислительных ресурсов.

Таким образом, исходя из сформулированных ограничений, перспективы дальнейшего развития модели Polyps-SAM2 видятся в нескольких направлениях. Интеграция автоматического детектора подсказок (например, на базе YOLOv8 или EfficientDet) позволит устранить зависимость от ручного ввода bounding boxes и приблизить систему к полностью автономному режиму работы. Адаптация к видеоданным с использованием банка памяти SAM2 откроет возможность отслеживания полипов в реальном времени во время колоноскопии, что особенно важно для клинического применения. Модификация декодера маски для поддержки множественных выходов (multi-mask prediction) решит проблему сегментации множественных полипов.

Таким образом, Polyps-SAM2 представляет собой перспективную основу для создания интеллектуальных систем поддержки принятия решений в гастроэнтерологии, однако её переход от исследовательского прототипа к клиническому инструменту требует дальнейшего инженерного и методологического развития.

## **БЛАГОДАРНОСТИ**

Работа выполнена при финансовой поддержке Севастопольского государственного университета, проект 42-01-09/319/2025-1.

## **Список литературы**

### **References**

1. Siegel R.L., Miller K.D., Wagle N.S., Jemal A. Cancer statistics, 2023. CA: A Cancer Journal for Clinicians. 2023. 73(1), p.17-48, doi: 10.3322/caac.21763
2. Antonelli G., Badalamenti M., Hassan C., Repici A. Impact of artificial intelligence on colorectal polyp detection. Best practice & research. Clinical gastroenterology. 2021. 52-53, 101713, doi: 10.1016/j.bpg.2020.101713
3. Shchetinin E.Yu., On the effectiveness of using visual transformers in detecting abnormalities of histopathological images. Mathematical models and computer simulations. 2025. 17(1), p.46-54, doi: 10.1134/S2070048224700716
4. Ronneberger O., Fischer P., Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation. Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. Lecture Notes in Computer Science. 2015. V. 9351, doi: 10.1007/978-3-319-24574-4\_28
5. Zhou Z., Rahman Siddiquee M.M., Tajbakhsh N., Liang J. UNet++: A Nested U-Net Architecture for Medical Image Segmentation. Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support. DLMIA ML-CDS 2018. Lecture Notes in Computer Science. 2018. V. 11045, doi: 10.1007/978-3-030-00889-5\_1
6. Kirillov A., et al., Segment Anything. 2023. URL: <https://arxiv.org/abs/2304.02643>
7. He S., et al., Accuracy of Segment-Anything Model (SAM) in medical image segmentation tasks. 2023. URL: <https://arxiv.org/abs/2304.09324>
8. Dong H., Gu H., Chen Y., Yang J., Mazurowski, M. Segment anything model 2: an application to 2d and 3d medical images. 2024. URL: <https://arxiv.org/abs/2408.00756>
9. Jha D. et al. Kvasir-SEG: A Segmented Polyp Dataset. MultiMedia Modeling. MMM 2020. Lecture Notes in Computer Science. 2020. V. 11962, doi: 10.1007/978-3-030-37734-2\_37
10. De La Torre R.M., Gonzalez A.A.C., Romero-Aroca A. CVC-ClinicDB: A public database for ophthalmology research. Proceedings of the 21st International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI 2018). 2018. pp. 430-439, doi: 10.1007/978-3-030-00934-2\_49.

11. Fan D-P, et al. PraNet: Parallel Reverse Attention Network for Polyp Segmentation. Proceedings of the 2020 International Conference on Medical Image Computing and Computer-Assisted Intervention. 2020. p. 263–273, doi: 10.1007/978-3-030-59725-2\_26
12. Patel K., Bur A.M., Wang G. Enhanced U-Net: A Feature Enhancement Network for Polyp Segmentation. IEEE International Conference on Robotics and Vision (ICRV). 2021. pp. 181-188, doi: 10.1109/ICRV52889.2021.00032.
13. Zhang W., et al., HSNNet: A hybrid semantic network for polyp segmentation. Computers in Biology and Medicine. 2022. 150, p. 106173, doi: 10.1016/j.combiomed.2022.106173
14. Srivastava A., Jha D., Chanda, S., et al. MSRF-Net: A Multi-Scale Residual Fusion Network for Biomedical Image Segmentation. IEEE Journal of Biomedical and Health Informatics. 2021. 26(5), 2252–2263, doi: 10.1109/JBHI.2021.3138024
15. Dong B., et al., Polyp-PVT: Polyp segmentation with pyramid vision transformers. 2021. URL: <https://arxiv.org/abs/2108.06932>
16. Yuheng L., Mingzhe H., Xiaofeng Ya. Polyp-SAM: Transfer SAM for Polyp Segmentation. 2023. URL: <https://arxiv.org/abs/2305.00293>
17. Chih-Hung H., Hsin-Yu W., Yu-Lun L. HarDNet-MSEG: A Simple Encoder-Decoder Polyp Segmentation Neural Network that Achieves over 0.9 Mean Dice and 86 FPS. 2021. URL: <https://arxiv.org/abs/2101.07172>

**Щетинин Евгений Юрьевич**, доктор физико-математических наук, доцент, профессор кафедры информационных технологий и систем, Севастопольский государственный университет, г. Севастополь, Россия

**Тютюнник Анастасия Александровна**, кандидат физико-математических наук, доцент кафедры математического моделирования и искусственного интеллекта, Российский университет дружбы народов им. Патриса Лумумбы, г. Москва, Россия

**Shchetinin Eugene Yurievich**, Doctor of Physical and Mathematical Sciences, Associate Professor, Professor of the Department of Information Technology and Computer Systems, Sevastopol State University, Sevastopol, Russia

**Tiutiunnik Anastasiia Alexandrovna**, Candidate of Physics and Mathematics, Associate Professor of the Department of Mathematical Modeling and Artificial Intelligence, RUDN University, Moscow, Russia

КОМПЬЮТЕРНОЕ МОДЕЛИРОВАНИЕ  
COMPUTER SIMULATION

UDC 517.958

DOI: 10.18413/2518-1092-2025-10-3-0-6

Konstantinov I.S.<sup>1</sup>  
Taha A.T.T.<sup>2</sup>  
Goldobina D.M.<sup>2</sup>

SEIRD EPIDEMIOLOGICAL MODELS FOR PLANT DISEASE

- <sup>1</sup> Belgorod State Technological University named after V.G. Shukhov,  
46 Kostyukova St., Belgorod, 308012, Russia  
<sup>2</sup> Belgorod State National Research University,  
85 Pobedy St., Belgorod, 308015, Russia

*e-mail: 1591248@bsuedu.ru*

**Abstract**

This paper explores the potential application of the SEIRD epidemiological model – commonly used in human disease modeling – to the context of plant disease outbreaks. Plant pathogens pose a significant threat to agricultural productivity, necessitating robust quantitative tools for understanding their dynamics and guiding control strategies. The SEIRD model was adapted to plant pathology by incorporating key agricultural variables, including environmental factors (e.g., humidity and temperature), plant growth stages, and the impact of interventions such as chemical treatments and removal of infected plants. Spatial dynamics were also modeled using traveling wave formulations. Results indicate that the modified model effectively captures the temporal and spatial progression of plant epidemics, enabling prediction of outbreak peaks and evaluation of control measures. This study presents a flexible mathematical framework that can be extended to various plant diseases, providing a valuable tool for data-driven decision-making in smart agriculture and epidemic risk management.

**Keywords:** SEIRD model; plant disease epidemiology; mathematical modeling; environmental factors; agricultural disease management

**For citation:** Konstantinov I.S., Taha A.T.T., Goldobina D.M. SEIRD Epidemiological Models for Plant Disease // Research result. Information technologies. – Т.10, №3, 2025. – P. 64-71. DOI: 10.18413/2518-1092-2025-10-3-0-6

Константинов И.С.<sup>1</sup>  
Таха А.Т.Т.<sup>2</sup>  
Голдобина Д.М.<sup>2</sup>

ЭПИДЕМИОЛОГИЧЕСКИЕ МОДЕЛИ SEIRD  
ДЛЯ БОЛЕЗНЕЙ РАСТЕНИЙ

- <sup>1</sup> Белгородский государственный технологический университет им. В.Г. Шухова,  
ул. Костюкова, 46, г. Белгород, 308012, Россия  
<sup>2</sup> Белгородский государственный национальный исследовательский университет,  
ул. Победы, 85, г. Белгород, 308015, Россия

*e-mail: 1591248@bsuedu.ru*

**Аннотация**

В данной статье рассматривается потенциальное применение эпидемиологической модели SEIRD, широко используемой в моделировании заболеваний человека, к контексту вспышек болезней растений. Фитопатогены представляют серьезную угрозу для сельскохозяйственного производства, что требует использования надёжных количественных инструментов для понимания динамики их распространения и разработки эффективных стратегий контроля. Модель SEIRD была адаптирована для фитопатологии путём включения ключевых агрономических переменных, таких как экологические факторы

(например, влажность и температура), фазы роста растений и влияние агротехнических мероприятий, включая химическую обработку и удаление инфицированных растений. Пространственная динамика была смоделирована с использованием уравнений бегущей волны. Результаты показывают, что модифицированная модель эффективно отражает временное и пространственное развитие эпидемий среди растений, позволяя прогнозировать пики заболеваемости и оценивать эффективность мер контроля. Исследование представляет собой гибкую математическую основу, которую можно адаптировать для различных фитопатологий, что делает её ценным инструментом для принятия решений на основе данных в условиях «умного» сельского хозяйства и управления эпидемиологическими рисками.

**Ключевые слова:** модель SEIRD; эпидемиология болезней растений; математическое моделирование; экологические факторы; управление болезнями в сельском хозяйстве

**Для цитирования:** Константинов И.С., Таха А.Т.Т., Голдобина Д.М. Эпидемиологические модели SEIRD для болезней растений // Научный результат. Информационные технологии. – Т.10, №3, 2025. – С. 64-71. DOI: 10.18413/2518-1092-2025-10-3-0-6

## INTRODUCTION

Mathematical modeling has emerged as a fundamental tool in the study of complex biological systems, particularly in understanding the mechanisms underlying the spread of infectious diseases. By providing a quantitative framework, mathematical models enable researchers to simulate disease dynamics, predict future outbreaks, and evaluate the potential outcomes of various intervention strategies. Among the multitude of models developed in the field of epidemiology, the SIR model stands out as one of the most classical and widely adopted approaches. Originally designed for human and animal populations, this model captures the core dynamics of infectious disease transmission through the interaction of three primary compartments: susceptible (S), infected (I), and removed (R). In this study, we explore the potential of adapting the SIR model, and its extended forms such as SEIRD, to the domain of plant epidemiology.

The basic structure of the SIR model is built upon a system of ordinary differential equations that describe the rate of change of each compartment over time. [1-2] The model assumes a closed population with no births or natural deaths unrelated to the disease. The susceptible compartment, denoted as  $S(t)$ , represents individuals – or in our context, plants – that are vulnerable to infection. The infected compartment,  $I(t)$ , includes those currently harboring the pathogen and capable of transmitting it. The removed compartment,  $R(t)$ , consists of plants that have either recovered, died, or been culled to prevent further spread. The interactions among these compartments are governed by the following equations:

The basic structure of the SIR model is built upon a system of ordinary differential equations that describe the rate of change of each compartment over time. [1-2] The model assumes a closed population with no births or natural deaths unrelated to the disease. The susceptible compartment, denoted as  $S(t)$ , represents individuals – or in our context, plants – that are vulnerable to infection. The infected compartment,  $I(t)$ , includes those currently harboring the pathogen and capable of transmitting it. The removed compartment,  $R(t)$ , consists of plants that have either recovered, died, or been culled to prevent further spread. The interactions among these compartments are governed by the following equations:

$$\begin{aligned} \frac{dS}{dt} &= -\beta S I, \\ \frac{dI}{dt} &= \beta S I - \alpha I, \\ \frac{dR}{dt} &= \alpha I. \end{aligned} \quad (1)$$

Here,  $\beta$  is the transmission rate, reflecting the probability of disease spread from an infected plant to a susceptible one. The parameter  $\alpha$  denotes the removal rate, which can include recovery, death, or agricultural interventions such as the removal of infected plants. These equations form the backbone of many epidemiological analyses, offering valuable insight into the trajectory of an outbreak under various conditions.

When applying this model to plant disease epidemics, however, several critical modifications are necessary to capture the unique characteristics of plant-pathogen interactions and the spatial and environmental heterogeneity inherent in agricultural systems. One of the key limitations of the classical SIR model is its assumption of a homogeneously mixed population, which rarely holds true in agricultural fields where plants are organized in spatial patterns. To address this, spatial dynamics can be incorporated into the model by transforming the system into a partial differential equation (PDE) framework [3-5]. For example, the infected compartment equation may be rewritten to include a diffusion term:

$$\partial I / \partial t = \beta S I - \alpha I + D \partial^2 I / \partial x^2. \quad (2)$$

In this extended form,  $D$  represents the diffusion coefficient, which quantifies the spatial spread of the pathogen, potentially influenced by vectors such as wind, water, or insects. This modification allows the model to simulate the movement of the disease front across a field, capturing the spatial propagation of the epidemic.

Another important refinement involves the variability of the transmission rate  $\beta$ . Unlike human populations, where interaction patterns can be relatively stable, plant environments are subject to fluctuating external conditions such as temperature, humidity, and the presence of disease vectors. As such,  $\beta$  can be modeled as a function of both space and time:  $\beta = \beta(x, t)$ . This dynamic approach enables a more realistic representation of disease transmission, accounting for daily or seasonal changes in environmental conditions.

Depending on the nature of the plant disease being studied, it may be necessary to extend the SIR model to include additional compartments. For instance, many pathogens exhibit a latent period during which an infected plant is not yet infectious. This necessitates the incorporation of an exposed compartment (E), leading to the SEIR model. In some cases, the inclusion of a deceased compartment (D) results in the SEIRD model, providing a more nuanced understanding of disease outcomes. These extensions help to capture the biological complexity of plant-pathogen interactions more effectively. In addition to these structural enhancements, the concept of traveling waves offers a powerful tool for analyzing the spatio-temporal spread of plant diseases. A traveling wave solution assumes that the disease front moves through the field at a constant speed  $c$ , and introduces a new variable  $\xi = x - ct$ . This transformation simplifies the PDEs and facilitates analytical and numerical exploration of how rapidly and in what manner the disease propagates spatially.

While these modifications significantly enhance the model's applicability to plant epidemiology, they also introduce new challenges. Chief among them is the requirement for accurate and high-resolution field data to estimate model parameters such as  $\beta$ ,  $\alpha$ , and  $D$ . Moreover, solving PDE-based models analytically is often intractable, necessitating the use of numerical simulations and computational tools. Environmental variability and heterogeneity in field conditions further complicate parameter estimation and model calibration [6].

Despite these challenges, the integration of spatial dynamics, time-varying transmission rates, and additional epidemiological compartments renders the SEIRD framework a powerful and adaptable tool for modeling plant disease outbreaks. By capturing the essential biological and environmental factors that drive disease dynamics in agricultural systems, the SEIRD model not only aids in understanding the underlying processes but also provides a quantitative foundation for designing and evaluating intervention strategies. This study aims to demonstrate the potential of this model in informing data-driven decision-making in plant disease management and contributing to the advancement of precision agriculture.

### **MODIFICATION OF SEIRD MODEL TO BE IMPLEMENTED TO PLANT DISEASE**

The authors in [7-11] developed a mathematical model to simulate epidemiological phenomena, particularly focusing on the dynamics of infectious diseases. A well-established framework was constructed and thoroughly analyzed to capture the essential components of disease transmission and control. The

SEIRD model is a well-established mathematical framework used to simulate and analyze the spread of infectious diseases in human populations. Given the biological similarities in the mechanisms of disease propagation, the SEIRD model can be adapted to understand the dynamics of plant disease epidemics. This report outlines how the SEIRD model can be applied to plant pathology with appropriate modifications.

In the context of plant diseases, the SEIRD model compartments can be reinterpreted as follows:

- S (Susceptible): Healthy plants that are vulnerable to infection.
- E (Exposed): Plants that have been infected but are not yet infectious (latent period).
- I (Infectious): Infected plants capable of transmitting the disease.
- R (Recovered): Plants that have recovered or are resistant to reinfection.
- D (Dead): Plants that have died or been removed due to the disease.

The differential equations governing the SEIRD model in plant pathology are similar to the human model but must consider plant-specific parameters:

$$\begin{aligned}\frac{dS}{dt} &= -\beta * S * I, \\ \frac{dE}{dt} &= \beta * S * I - \sigma * E, \\ \frac{dI}{dt} &= \sigma * E - \gamma * I - \mu * I, \\ \frac{dR}{dt} &= \gamma * I, \\ \frac{dD}{dt} &= \mu * I.\end{aligned}\tag{3}$$

Where:

$\beta$ : Transmission rate influenced by planting density, vector presence, and environmental conditions.

$\sigma$ : Rate of progression from exposed to infectious.

$\gamma$ : Recovery rate (e.g., due to resistance or treatment).

$\mu$ : Mortality or removal rate due to disease.

In adapting the SEIRD model to accurately represent plant disease epidemics, one of the most essential modifications involves the incorporation of environmental factors – chiefly temperature and humidity – into the core dynamics of disease transmission. Unlike human epidemiology, where contact patterns and behavior largely govern transmission, the progression and intensity of plant epidemics are heavily influenced by abiotic environmental conditions. Temperature and relative humidity, in particular, play a critical role in determining the rate at which plant pathogens infect hosts, survive in the environment, and reproduce. These variables fluctuate daily and seasonally, thereby making it necessary to define the transmission rate  $\beta(t)$  as a dynamic function that responds to environmental inputs.

To integrate this variability, the transmission rate can be modeled as:

$$\beta(t) = \beta_0 \times f(T(t), H(t)).\tag{4}$$

Here,  $\beta_0$  denotes the baseline transmission rate under optimal conditions, while  $T(t)$  and  $H(t)$  represent the temperature and relative humidity at time  $t$ , respectively. The function  $f(T(t), H(t))$  acts as an environmental response function, modifying the effective transmission rate based on how closely current environmental conditions align with the optimal thresholds for pathogen activity. A commonly used formulation for  $f(T(t), H(t))$  involves a product of two sigmoid functions:

$$f(T, H) = \frac{1}{1+e^{-k_1(T-T_{opt})}} \times \frac{1}{1+e^{-k_2(H-H_{opt})}}.\tag{5}$$

In this expression,  $T_{opt}$  and  $H_{opt}$  represent the optimal temperature and humidity conditions for disease transmission, while  $k_1$  and  $k_2$  determine the steepness of the response – how sensitive the pathogen is to deviations from its optimal climate. When temperature and humidity are near their respective optima,  $f(T,H)$  approaches 1, maximizing transmission. Conversely, under unfavorable conditions, the function declines, reducing  $\beta(t)$  and thus lowering the likelihood of infection. Beyond environmental responsiveness, the SEIRD model can also be improved by accounting for seasonal trends and plant growth stages. Susceptibility to infection often varies across the plant's development cycle – seedlings may be more vulnerable than mature plants, for instance. A time-dependent growth stage multiplier  $G(t)$  can be introduced to scale  $\beta(t)$  in accordance with the current phase of plant development. For example,  $G(t)$  might take values like 0.3 during the seedling stage, 0.6 during vegetative growth, and 1.0 during flowering, when susceptibility peaks. Another vital consideration is the effect of agricultural interventions – notably fungicide application, crop rotation, and physical removal of infected plants. These interventions can be incorporated into the SEIRD model by further modifying  $\beta(t)$  through a treatment decay function  $M(t)$ . Immediately after a fungicide is applied,  $M(t) = 1$ , indicating full effect; over time, its impact decays. Accordingly, the effective transmission rate becomes:

$$\beta(t) = \beta_0 \times f(T(t).H(t)) \times G(t) \times e^{-\lambda(t-t_0)} \quad (6)$$

Where  $\lambda$  is the rate of decay and  $t_0$  is the time of treatment. In some scenarios, interventions also enhance the recovery rate  $\gamma(t)$ , for instance, by accelerating the healing process or removing infected plants from the population. Altogether, these adaptations – environmental modulation, growth-stage sensitivity, and intervention effects – equip the SEIRD model with the structural flexibility needed to accurately simulate real-world plant epidemics. They enable the model to not only mirror the biological realities of plant-pathogen-environment interactions but also to evaluate the effectiveness of management strategies under changing conditions. These enhancements make the model an effective decision-support tool for agricultural disease control, particularly in the context of climate variability and precision agriculture

Plant susceptibility to disease is not static; it fluctuates in response to both seasonal environmental cycles and distinct physiological growth stages. For instance, plants are generally more vulnerable during early developmental phases such as the seedling stage, while resistance tends to increase as they mature. To mathematically capture the effect of seasonal patterns on disease transmission, a periodic adjustment function  $g(t)$  can be introduced:

$$g(t) = 1 + \alpha \cdot \sin\left(\frac{2\pi t}{365} + \phi\right) \quad (7)$$

In this equation,  $\alpha$  represents the amplitude of seasonal variation, indicating the intensity of fluctuation, while  $\phi$  denotes the phase shift used to align the seasonal peak with the critical infection period (e.g., spring or harvest time).

To further refine the model, a growth-stage sensitivity coefficient  $G(t)$  is applied to account for the plant's developmental phase. This coefficient can take discrete values based on observed susceptibility: for example,  $G(t) = 0.3$  for seedlings,  $G(t)=0.6$  during vegetative growth, and  $G(t)=1.0$  during flowering or fruiting stages.

Combining both seasonal and phenological influences, the time-dependent transmission rate becomes:

$$\beta(t) = \beta_0 \cdot g(t) \cdot G(t). \quad (8)$$

This composite function allows the model to reflect real-world variations in disease transmission potential driven by environmental seasonality and plant development.

Agricultural actions, such as chemical treatments or removal of infected plants, significantly affect epidemic dynamics and should be modeled explicitly.

The transmission rate can be reduced after treatment:

$$\beta(t) = \beta_0 \cdot e^{-\delta \cdot M(t)}. \quad (9)$$

Where:

$\delta$ : Reduction factor due to treatment

$M(t)$ : Treatment effectiveness function, such as:

- $M(t) = 1$  (immediately after application);
- $M(t) = e^{(-\lambda(t - t_0))}$  (decays over time since treatment at  $t_0$ ).

Similarly, the recovery rate  $\gamma(t)$  can be increased due to intervention:

$$\gamma(t) = \gamma_0 + \eta \cdot M(t) \quad (10)$$

### FINAL MODIFIED SEIRD EQUATIONS

$$\begin{aligned} dS/dt &= -\beta(t) \cdot S \cdot I/N \\ dE/dt &= \beta(t) \cdot S \cdot I/N - \sigma E \\ dI/dt &= \sigma E - \gamma(t)I - \mu I \\ dR/dt &= \gamma(t)I \\ dD/dt &= \mu I \end{aligned} \quad (11)$$

### APPLICATIONS IN PLANT DISEASE MANAGEMENT

The modified SEIRD model offers a powerful and versatile framework for managing plant disease outbreaks in agricultural systems. One of its key applications lies in its ability to simulate both the spatial and temporal dynamics of pathogen spread. By incorporating environmental conditions, plant development stages, and intervention effects, the model can realistically reflect how diseases propagate through a crop field over time and space. This capability allows researchers and agricultural decision-makers to visualize potential epidemic trajectories and understand how fast and in what direction an outbreak may unfold. Additionally, the model provides a structured approach for assessing the impact of various control strategies, such as fungicide applications, crop rotation, or the removal of infected plants. By comparing different intervention scenarios, the SEIRD framework enables users to identify the most effective and cost-efficient disease management practices.

Furthermore, the model can forecast the timing and intensity of epidemic peaks, which is crucial for optimizing the timing of interventions and minimizing crop loss. Finally, the model contributes to broader integrated pest and disease management (IPDM) efforts by offering a data-driven foundation for coordinating multiple control measures in a holistic and sustainable manner. Through its predictive capacity and flexibility, the SEIRD model serves as an essential tool in modern, precision-guided plant disease control.

### CONCLUSION

The SEIRD model, when appropriately modified, offers a robust and flexible framework for modeling the dynamics of plant disease epidemics. By incorporating critical agricultural factors – such as environmental conditions (e.g., temperature and humidity), seasonal variation, plant growth stages, and the decaying effects of control interventions – the model becomes highly relevant for real-world agricultural scenarios. These modifications enhance the model's ability to simulate spatial and temporal disease progression, estimate outbreak peaks, and assess the effectiveness of various management strategies. Furthermore, the inclusion of traveling wave analysis and time-varying transmission parameters allows for more accurate prediction of disease spread in heterogeneous crop environments. As a result, the adapted SEIRD model serves as a valuable decision-support tool in integrated pest and disease management

programs. It not only facilitates data-driven planning and timely intervention but also contributes to the development of precision agriculture by enabling more sustainable, efficient, and proactive responses to plant epidemics.

### References

1. Kermack W.O., McKendrick A.G. A contribution to the mathematical theory of epidemics // Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character. – 1927. – 115(772). – P. 700–721. <https://doi.org/10.1098/rspa.1927.0118>
2. Hethcote H.W. The mathematics of infectious diseases // SIAM Review. – 2000. – 42(4). – P. 599–653. <https://doi.org/10.1137/S0036144500371907>
3. Gilligan C.A., Gubbins S., Simons S.A. Analysis and fitting of an SIR model with host response to infection load for a plant disease // Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences. – 1997. – 352(1351) – P. 353–364. <https://doi.org/10.1098/rstb.1997.0026>
4. Al Basir F., Blyuss K.B., Ray S. Modelling the effects of awareness-based interventions to control the mosaic disease of *Jatropha curcas*. – 2018. – ArXiv preprint arXiv:1810.03158.
5. Brown D. H. Stochastic spatial models of plant diseases. – 2001. – ArXiv: math/0112094v1 [math.NA]
6. Konstantinov I.S., Taha A.T.T., Starchenko D.N. SIR Model Dynamics: Insights into Epidemics and Vaccination // Economics. Information technologies. – 2024. – 51(1). – P. 145–156. DOI 10.52575/2712-746X2024-51-1-145-156
7. Konstantinov I.S., Taha A.T.T. Mathematical analysis of SIR model with incubation period // Research result. Information technologies. – T.9, №3, 2024. – P. 3-9. DOI: 10.18413/2518-1092-2024-9-3-0-1
8. Agha H.R., Taha A.T. Methods of the epidemics spread mathematical modeling // Research result. Information technologies. – T.7, №4, 2022. – P. 25-33. DOI: 10.18413/2518-1092-2022-7-4-0-3
9. Taha A.T.T. Modeling the COVID-19 transmission protocol // Information Systems and Technologies. – 2025. – 2(148). – P. 48-59.
10. Taha A.T., Konstantinov I.S. Analyzing pandemic dynamics through traveling waves: a mathematical model. Proceedings of the Southwest State University. Series: IT Management, Computer Science, Computer Engineering. Medical Equipment Engineering. 2025. 15(1): 170-179. <https://doi.org/10.21869/2223-1536-2025-15-1-170-179>

### Список литературы

1. Kermack W.O., McKendrick A.G. A contribution to the mathematical theory of epidemics // Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character. – 1927. – 115(772). – P. 700–721. <https://doi.org/10.1098/rspa.1927.0118>
2. Hethcote H.W. The mathematics of infectious diseases // SIAM Review. – 2000. – 42(4). – P. 599–653. <https://doi.org/10.1137/S0036144500371907>
3. Gilligan C.A., Gubbins S., Simons S.A. Analysis and fitting of an SIR model with host response to infection load for a plant disease // Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences. – 1997. – 352(1351) – P. 353–364. <https://doi.org/10.1098/rstb.1997.0026>
4. Al Basir F., Blyuss K.B., Ray S. Modelling the effects of awareness-based interventions to control the mosaic disease of *Jatropha curcas*. – 2018. – ArXiv preprint arXiv:1810.03158.
5. Brown D. H. Stochastic spatial models of plant diseases. – 2001. – ArXiv: math/0112094v1 [math.NA]
6. Константинов И.С., Таха А.Т., Старченко Д.Н. Динамика модели SIR: взгляд на эпидемии и вакцинацию // Экономика. Информатика. – 2024. – 51(1). С. 145–156. DOI 10.52575/2712-746X-2024-51-1-145-156
7. Константинов И.С., Таха А.Т.Т. Математический анализ модели SIR с учетом периода инкубации // Научный результат. Информационные технологии. – Т.9, №3, 2024. – С. 3-9. DOI: 10.18413/2518-1092-2024-9-3-0-1
8. Агха Х.Р., Таха А.Т. Методы математического моделирования распространения эпидемий // Научный результат. Информационные технологии. – Т.7, №4, 2022. – С. 25-33. DOI: 10.18413/2518-1092-2022-7-4-0-3.
9. Таха А.Т.Т. Моделирование протокола передачи COVID-19 // Информационные системы и технологии. – 2025. – 2(148). – С. 48-59.
10. Таха А., Константинов И.С. Анализ динамики пандемий через волны передвижения: математическая модель. Известия Юго-Западного государственного университета. Серия: Управление,

вычислительная техника, информатика. Медицинское приборостроение. 2025. 15(1): 170-179. <https://doi.org/10.21869/2223-1536-2025-15-1-170-179>

**Konstantinov Igor Sergeevich**, Doctor of Technical Sciences, Professor, Director of the Institute of Information Technologies and Control Systems, Belgorod State Technological University named after V.G. Shukhov, Belgorod, Russia

**Taha Asraa Tariq Taha**, Postgraduate Student of the Department of Mathematical and Software of Information Systems, Belgorod State National Research University, Belgorod, Russia

**Goldobina Darya Mikhailovna**, Assistant, Department of Information and Telecommunication Systems and Technologies, Institute of Engineering and Digital Technologies, Belgorod State National Research University, Belgorod, Russia

**Константинов Игорь Сергеевич**, доктор технических наук, профессор, директор института информационных технологий и управляющих систем, Белгородский государственный технологический университет им. В.Г. Шухова, г. Белгород, Россия

**Таха Асраа Тарик Таха**, аспирант кафедры математического и программного обеспечения информационных систем, Белгородский государственный национальный исследовательский, г. Белгород, Россия

**Голдобина Дарья Михайловна**, ассистент кафедры информационно-телекоммуникационных систем и технологий, института инженерных и цифровых технологий, Белгородский государственный национальный исследовательский, г. Белгород, Россия

УДК 519.688

DOI: 10.18413/2518-1092-2025-10-3-0-7

Лихошерстный А.Ю.<sup>1</sup>  
Великая Я.Г.<sup>2</sup>

**СТРУННО-ВОЛНОВОЙ ПАРАЛЛЕЛЬНЫЙ АЛГОРИТМ  
РЕШЕНИЯ СЛАУ ДЛЯ РАЗРЕЖЕННЫХ МАТРИЦ**

<sup>1</sup> Научно-технический центр «АПМ»,  
Октябрьский б-р, 14, Московская область, г. Королев, 141070, Россия  
<sup>2</sup> Белгородский государственный национальный исследовательский университет,  
ул. Победы, 85, г. Белгород, 308015, Россия

*e-mail: ozzy.osbourne.man@gmail.com, velikaya@bsuedu.ru*

**Аннотация**

В статье рассматривается параллельный алгоритм решения систем линейных алгебраических уравнений для симметричных разреженных матриц, который позволяет разбивать большую задачу на множество мелких подзадач, тем самым как повышая быстродействие, так и уменьшая потребление памяти. В его основе лежит способ одновременного вычисления промежуточных значений при разложении матрицы с сохранением балансировки нагрузки на процессоры таким образом, чтобы при получении окончательного результата левых частей разложения от них не зависели правые части разложения. Такой подход позволяет исходную матрицу жесткости представить в виде произведения большого количества простых матриц и решать систему линейных алгебраических уравнений в виде последовательности решений методом подстановки. Для уменьшения заполнения разреженных матриц разложения использовался приближенный метод минимальной степени, который помимо того, что является одним из самых эффективных и быстродействующих из существующих на сегодняшний момент времени, позволяет для разработанного алгоритма более равномерно распределить нагрузку вычислений. Разработанный метод реализован в программных продуктах НТЦ АПМ для систем с общей памятью, но также может быть реализован и для систем с распределенной памятью.

**Ключевые слова:** алгоритм минимальной степени; струнно-волновой алгоритм; метод конечных элементов; система линейных алгебраических уравнений; матричная факторизация

**Для цитирования:** Лихошерстный А.Ю., Великая Я.Г. Струнно-волновой параллельный алгоритм решения СЛАУ для разреженных матриц // Научный результат. Информационные технологии. – Т.10, №3, 2025. – С. 72-79. DOI: 10.18413/2518-1092-2025-10-3-0-7

Likhosherstnyy A.Yu.<sup>1</sup>  
Velikaya Ya.G.<sup>2</sup>

**STRING-WAVE DIRECT PARALLEL SOLVER  
FOR SPARSE SYSTEM OF LINEAR EQUATIONS**

<sup>1</sup> APM Ltd.,  
14 Oktyabrsky Boulevard, Moscow region, Korolev, 141070, Russia  
<sup>2</sup> Belgorod State National Research University,  
85 Pobedy St., Belgorod, 308015, Russia

*e-mail: ozzy.osbourne.man@gmail.com, velikaya@bsuedu.ru*

**Abstract**

The article discusses a parallel algorithm for solving systems of linear algebraic equations for symmetric sparse matrices, which allows you to split a large task into many small subtasks, thereby both increasing performance and reducing memory consumption. It is based on a method of simultaneous calculation of intermediate values during matrix decomposition while maintaining load balancing on processors so that when the final result of the left parts of the decomposition is

obtained, the right parts of the decomposition do not depend on them. This approach allows the initial stiffness matrix to be represented as a product of a large number of simple matrices and solve a system of linear algebraic equations in the form of a sequence of solutions by substitution. To reduce the filling of sparse decomposition matrices, an approximate minimum degree method was used, which, in addition to being one of the most efficient and fastest existing at the moment, allows the developed algorithm to distribute the load of calculations more evenly. The developed method is implemented in NTC APM software products for systems with shared memory, but it can also be implemented for systems with distributed memory.

**Keywords:** approximate minimum degree; string-wave algorithm; finite element method; system of linear equations; matrix factorization

**For citation:** Likhosherstnyy A.Yu., Velikaya Ya.G. String-Wave Direct Parallel Solver for Sparse System of Linear Equations // Research result. Information technologies. – Т.10, №3, 2025. – P. 72-79. DOI: 10.18413/2518-1092-2025-10-3-0-7

## **ВВЕДЕНИЕ**

В настоящее время для инженерных расчетов в области строительства и машиностроения активно используется метод конечных элементов (МКЭ) [1]. Применение МКЭ предполагает решение систем линейных алгебраических уравнений (СЛАУ) большой размерности. Матрицы таких систем, называемые матрицами жёсткости, обладают рядом свойств, зависящих от структуры и количества узлов конечно-элементной сетки, а также характера задачи. Эти матрицы являются симметричными и в значительной степени разреженными.

С помощью СЛАУ решаются такие задачи как: статический расчет, динамический расчет, расчет собственных частот, расчет устойчивости, нелинейные расчеты и др. В общем случае СЛАУ [2] имеет следующий вид:

$$K_{n,n}X_{n,m} = F_{n,m}, \quad (1)$$

где

- $K$  – матрица жесткости;
- $X$  – векторы перемещений;
- $F$  – векторы нагрузки для системы конечных элементов;
- $n$  – размерность матрицы жесткости;
- $m$  – количество загрузений.

Ввиду того, что матрица жесткости является разреженной, она представлена в строчном разреженном формате CSR (Compressed Sparse Rows) [3, 4], где структура хранения использует три одномерных массива:

- 1) массив ненулевых элементов матрицы жесткости построчно;
- 2) массив номеров столбцов ненулевых элементов построчно;
- 3) массив местоположения первого ненулевого элемента в каждой строке.

Такой формат позволяет существенно сократить объем исходных данных и уменьшить количество выполняемых над ними операций, т.е. при расчетах сокращается расход памяти и увеличивается быстродействие.

## **ОСНОВНАЯ ЧАСТЬ**

На сегодня используют два различных подхода решения СЛАУ: методы прямого решения и итерационные методы. Каждый подход обладает определенными преимуществами, но чаще всего пользуются прямыми методами. Главной проблемой итерационных методов является частая плохая обусловленность матриц жесткости большой размерности, когда ее определитель равен малой величине или стремится к нулю. И даже процесс предобуславливания не дает абсолютной гарантии сходимости решения.

Для упрощения решения СЛАУ с симметричной матрицей жесткости применяют следующее ее разложение [5]:

$$K_{n,n} = L_{n,n}D_{n,n}L_{n,n}^T, \quad (2)$$

где

$L$  – нижняя треугольная матрица (хранится в разреженном формате), у которой все диагональные элементы равны единице;

$D$  – диагональная матрица (хранится в виде одномерного массива).

Решение СЛАУ  $LDL^T X = F$  в этом случае сводится к последовательному решению трех простейших задач методом подстановки:

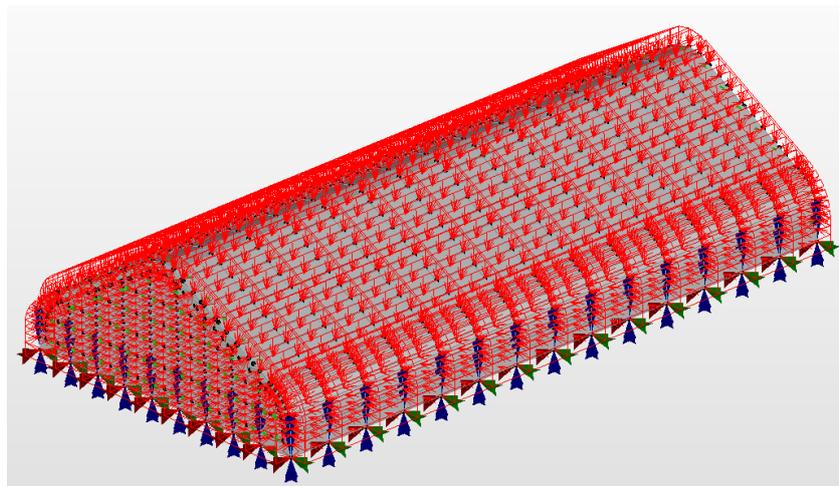
$$\begin{cases} L_{n,n}Y_{n,m} = F_{n,m}; \\ D_{n,n}Z_{n,m} = Y_{n,m}; \\ L_{n,n}^T X_{n,m} = Z_{n,m}. \end{cases} \quad (3)$$

Основной проблемой при таком разложении является заполнение разреженной треугольной матрицы  $L$  большим количеством ненулевых элементов, что влечет за собой значительный расход памяти и увеличение количества вычислительных операций над элементами этой матрицы. Поэтому перед процедурой разложения производят переупорядочивание строк и столбцов исходной матрицы жесткости. Для этого находят специальную матрицу переупорядочивания  $P$  такую, что:

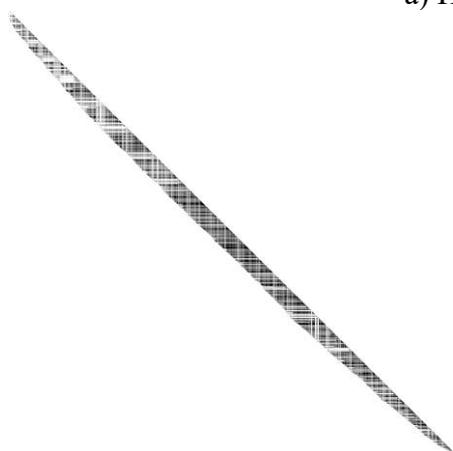
$$(P_{n,n}K_{n,n}P_{n,n}^T)(P_{n,n}X_{n,m}) = P_{n,n}F_{n,m}. \quad (4)$$

Матрица переупорядочивания  $P$  представляет собой матрицу, у которой в каждой строке и каждом столбце содержится только один элемент равный единице, а все остальные элементы равны нулю. Хранится эта матрица в виде массива номеров столбцов с единичными элементами.

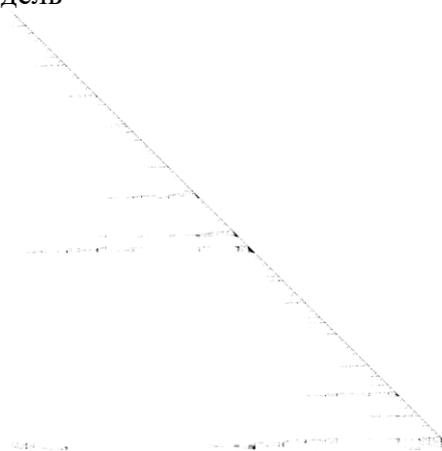
Для нахождения матрицы переупорядочивания часто применяют алгоритм Катхилла-Макки ввиду простоты его реализации, но такой алгоритм не является оптимальным с точки зрения уменьшения заполнения ненулевыми элементами матриц разложения. Более эффективными являются алгоритмы на основе применения метода минимальной степени. На практике данный метод в его исходном виде не используется в силу его значительной трудоемкости. Широко распространены две его модификации: множественный метод минимальной степени (Multiple Minimum Degree, MMD) и приближенный метод минимальной степени (Approximate Minimum Degree, AMD) [6]. Последний используется в программных продуктах компании НТЦ АПИМ. На рисунке 1 показан результат вычисления нижней треугольной матрицы  $L$  с переупорядочиванием с помощью алгоритма Катхилла-Макки и приближенным методом минимальной степени.



а) Исходная модель



б) Cuthill–McKee



в) AMD

Рис. 1. Результат разложения после переупорядочивания матрицы  
Fig. 1. The result of decomposition after matrix reordering

Здесь черными точками обозначены ненулевые элементы полученной матрицы разложения  $L$ . В таблице 1 показано количество полученных ненулевых элементов данной матрицы.

Таблица 1

Результаты работы алгоритмов переупорядочивания матрицы

Table 1

Results of matrix reordering algorithms

	Исходная матрица жесткости	Cuthill–McKee	AMD
Количество ненулевых элементов	279966	4510257	469879

После процедуры переупорядочивания производится непосредственно разложение матрицы жесткости. В классическом варианте разложение  $LDL^T$  вычисляется следующим образом:

$$\begin{cases} D_j = K_{jj} - \sum_{k=1}^{j-1} L_{jk} D_k L_{jk}^T; \\ L_{ij} = \frac{K_{ij} - \sum_{k=1}^{j-1} L_{ik} D_k L_{jk}^T}{D_j}; \\ i = 1, \dots, n; \quad j = 1, \dots, n. \end{cases} \quad (5)$$

Выражение (5) плохо поддается распараллеливанию ввиду большого количества информационных зависимостей. Кроме того, вычисления не сбалансированы, что значительно снижает их эффективность. Для устранения этих проблем был разработан алгоритм, адаптированный для параллельных вычислений [7, 8]. Треугольную матрицу разложения  $L$  можно визуально представить как музыкальный инструмент в виде арфы (рисунок 2). Каждая струна арфы соответствует столбцу матрицы  $L$ . Игра на струнах осуществляется поочередно. Игра на  $j$ -ой струне означает нахождение окончательных значений элементов  $j$ -ого столбца  $L$ . При этом, каждая струна распространяет незатухающую звуковую волну в сторону струн, расположенных от нее правее. Т.е., каждый  $j$ -й столбец матрицы  $L$  частично изменяет значения элементов всех столбцов с номером выше  $j$ . Таким образом, можно параллельно вычислять влияние текущего столбца на все столбцы, расположенные правее.



$$L = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ l_{21} & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ l_{31} & l_{32} & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ l_{41} & l_{42} & l_{43} & 1 & 0 & 0 & 0 & 0 & 0 \\ l_{51} & l_{52} & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ l_{61} & l_{62} & 0 & l_{64} & 0 & 1 & 0 & 0 & 0 \\ l_{71} & l_{72} & l_{73} & 0 & 0 & 0 & 1 & 0 & 0 \\ l_{81} & l_{82} & 0 & l_{84} & 0 & l_{86} & 0 & 1 & 0 \\ l_{91} & l_{92} & 0 & 0 & l_{95} & l_{96} & l_{97} & l_{98} & 1 \end{pmatrix}$$

Рис. 2. Визуальное представление матрицы разложения в виде арфы  
Fig. 2. Visual representation of the decomposition matrix in the form of a harp

Математически струнно-волновой алгоритм можно описать следующим образом. Сначала матрицам  $D$  и  $L$  присваиваются соответствующие значения матрицы жесткости:

$$\begin{cases} D_{ii} = K_{ii}; \\ L_{ij} = K_{ij}; \\ i = 1, \dots, n; \quad j = 1, \dots, i - 1. \end{cases} \quad (6)$$

Далее элементы этих матриц изменяются согласно вышеописанному алгоритму:

$$\begin{cases} L_{ij} = \frac{L_{ij}}{D_{jj}}; \\ L_{ik} = L_{ik} - L_{kj} D_{jj} L_{ij}; \\ D_{ii} = D_{ii} - L_{ij} D_{jj} L_{ij}; \\ j = 1, \dots, n; \quad i = j + 1, \dots, n; \quad k = 2, \dots, i - 1. \end{cases} \quad (7)$$

Т.к. матрица разложения  $L$  является разреженной, то не на все правые столбцы оказывает влияние текущий столбец. В ходе проведенных исследований было выявлено, что изменяются только те столбцы, номер которых соответствует номерам строк ненулевых элементов текущего столбца. Данное обстоятельство позволяет устранить лишние итерации цикла параллельного алгоритма.

При решении СЛАУ больших размерностей прямым методом часто не хватает памяти. Струнно-волновой алгоритм позволяет уменьшить размер задачи, например, применив следующее разложение:

$$K = L_1 B L_1^T, \quad (8)$$

где

$L_1$  – первые столбцы треугольной матрицы разложения  $L$ ;

$B$  – полученная влиянием  $L_1$  подматрица меньшей размерности (рисунок 3).

$$B = \begin{pmatrix} d_{11} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & d_{22} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & d_{33} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \vdots & \vdots \\ 0 & 0 & 0 & \boxed{S} & 0 & 0 & 0 & 0 & 0 & 0 \\ \vdots & \vdots \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}, \quad L_1 = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ & & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ & & & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ & & & & 1 & 0 & 0 & 0 & 0 & 0 \\ & & & & & 1 & 0 & 0 & 0 & 0 \\ & & & & & & 0 & 1 & 0 & 0 \\ & & & & & & & 0 & 0 & 1 & 0 \\ & & & & & & & & 0 & 0 & 0 & 1 & 0 \\ & & & & & & & & & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

Рис. 3. Уменьшение размерности задачи

Fig. 3. Reducing the dimensionality of the problem

В свою очередь, для матрицы  $B$  можно применить аналогичное разложение. В итоге, струнно-волновой алгоритм позволяет получить следующее выражение:

$$K = L_1 L_2 \dots L_{n-1} L_n D L_n^T L_{n-1}^T \dots L_2^T L_1^T \quad (9)$$

Данное разложение матрицы жесткости позволяет разбить большую задачу на множество мелких подзадач, где решение СЛАУ также находится методом подстановки:

$$\begin{cases} L_1 X_n = F; \\ L_2 X_{n-1} = X_n; \\ \dots \\ L_n X_{i-1} = X_{i-2}; \\ D X_i = X_{i-1}; \\ L_n^T X_{i+1} = X_i; \\ \dots \\ L_2^T X_2 = X_3; \\ L_1^T X_1 = X_2. \end{cases} \quad (10)$$

Для оценки эффективности параллельных вычислений применялись параметры ускорения и эффективности [9]. Под ускорением понимается отношение времени выполнения последовательного алгоритма ко времени выполнения параллельного алгоритма на  $p$  процессорах:

$$S_p = \frac{T_1}{T_p} \quad (11)$$

Эффективность использования параллельным алгоритмом процессоров при решении задачи определяется как отношение ускорения к количеству процессоров:

$$E_p = \frac{S_p}{p} \quad (12)$$

Результаты вычислительных экспериментов оценки эффективности параллельного алгоритма представлены в таблице 2. Алгоритм реализован с помощью технологии OpenMP [10]. Вычисления производились на шестиядерном процессоре Intel Core i7-8700K с 32 ГБ оперативной памяти DDR4. Для проведения вычислительных экспериментов использовалась программная система ARMStructure3D, в которой загружались модели различных размерностей и при замере времени выполнения параллельного алгоритма учитывалось время на формирование матрицы жесткости.

Таблица 2

Оценка эффективности параллельного алгоритма

Table 2

Efficiency evaluation of the parallel algorithm

Размерность матрицы жесткости, $n$	Ускорение, $S_p$	Эффективность, $E_p$
94692	3,54	0,59
174267	3,75	0,63
244620	4,86	0,81
527919	4,72	0,79
755181	4,78	0,80

Таким образом, струнно-волновой алгоритм эффективно работает в связке с алгоритмом переупорядочивания AMD, разбивая большую задачу на множество мелких подзадач с наименьшими информационными зависимостями, что позволяет при его параллельной реализации достичь оптимальных значений параметров ускорения и эффективности. Разработанный алгоритм успешно применяется в программных продуктах компании НТЦ АПМ и постоянно модернизируется с целью повышения своего быстродействия.

### ЗАКЛЮЧЕНИЕ

В статье представлен новый алгоритм решения систем линейных алгебраических уравнений для симметричных разреженных матриц. Данный алгоритм включает в себя метод переупорядочивания разреженных матриц AMD и струнно-волновой метод решения СЛАУ собственной разработки. Он успешно применяется в программных продуктах компании НТЦ АПМ. Разработанный решатель используется в таких программных модулях как статический и динамический расчет, расчет собственных частот и устойчивости при решении обобщенной задачи на собственные значения, расчет гармонических колебаний и многих других.

Разработанный алгоритм является параллельным и реализован с помощью технологии OpenMP. Вычислительные эксперименты показывают, что с увеличением количества процессоров и объема входных данных ускорение увеличивается без падения эффективности. Т.е. разработанный параллельный алгоритм является масштабируемым.

Дальнейшие исследования будут сосредоточены на повышении его быстродействия и реализации с помощью технологии CUDA для графических ускорителей NVIDIA и технологии MPI для систем с распределенной памятью.

### Список литературы

1. Бате К., Вилсон Е. Численные методы анализа и метод конечных элементов. – Москва: Стройиздат, 1982. – 448 с.
2. Джеммель Дж. Вычислительная линейная алгебра. Теория и приложения. – Москва: Мир, 2001. – 430 с.
3. Писсанецки С. Технология разреженных матриц. – Лондон: Academic Press Inc, 1984. – 312 с.

4. Ахо А., Хопкрофт Дж., Ульман Дж. Построение и анализ вычислительных алгоритмов. – Москва: Мир, 1979. – 542 с.
5. Андреев Б. Численные методы. – Москва: издательство ВМК МГУ им Ломоносова, 2013. – 336 с.
6. Heggenes P., Eisenstat S.C., Kumfert G. The Computational Complexity of the Minimum Degree Algorithm // Четырнадцатая норвежская конференция компьютерных наук, Осло, Норвегия, 26-28 ноября, 2001. URL: <https://digital.library.unt.edu/ark:/67531/metadc1410671/> (дата обращения: 09.07.2025).
7. Воеводин В., Воеводин Вл. Параллельные вычисления. – СПб: BHV, 2015. – 608 с.
8. Воеводин В.В. Математические основы параллельных вычислений. – Москва: МГУ, 1991. – 345 с.
9. Гергель В.П. Высокопроизводительные вычисления для многопроцессорных многоядерных систем. – Москва: Издательство Московского Университета, 2010. – 544 с.
10. Гергель В.П. Современные языки и технологии параллельного программирования. – Москва: Издательство Московского Университета, 2012. – 408 с.

### References

1. Bathe K.J., Wilson E.L. Numerical methods in finite element analysis. Moscow: Stroyizdat, 1982. 448 p.
2. Demmel J. Computational linear algebra. Theory and applications. Moscow: Word, 2001. 430 p.
3. Pissanetsky S. Sparse Matrix Technology. London: Academic Press Inc, 1984. 312 p.
4. Aho A.V., Hopcroft J.E., Ulman J.D.: The design and Analysis of Computer Algorithms. Addison – Wesley: Reading, MA. Third printing, 1976. 542 p.
5. Andreev B. Numerical methods. Moscow: Publishing department of the Faculty of Computational Mathematics and Mathematics of Lomonosov Moscow State University, 2013. 336 p.
6. Heggenes P., Eisenstat S.C., Kumfert G. The Computational Complexity of the Minimum Degree Algorithm // 14-th Norwegian Computer Science Conference, Oslo, Norway, November 26-28, 2001. <https://doi.org/10.2172/15002765>
7. Voevodin V., Voevodin V. Parallel Computing. Saint Petersburg: BHV, 2015. 608 p.
8. Voevodin V.V. Mathematical foundations of parallel computing. World Scientific Publishing Co., Series in computer science. 1992. V.33. 343 p.
9. Gergel V.P. High-performance computing for multiprocessor, multi-core systems. Moscow: Moscow University Publishing House, 2010. 544 p.
10. Gergel V.P. Modern languages and parallel programming technologies. Moscow: Moscow University Publishing House, 2012. 408 p.

**Лихошерстный Алексей Юрьевич**, кандидат технических наук, математик-программист, научно-технический центр «АПМ», г. Королёв, Россия

**Великая Яна Геннадьевна**, кандидат технических наук, доцент кафедры математического и программного обеспечения информационных систем, Белгородский государственный национальный исследовательский университет, г. Белгород, Россия

**Likhosherstnyy Alexey Yuryevich**, Candidate of Technical Sciences, Mathematician and Programmer, APM Ltd, Korolev, Russia

**Velikaya Yana Gennadievna**, Candidate of Technical Sciences, Associate Professor of the Department of Mathematical and Software Support for Information Systems, Belgorod State National Research University, Belgorod, Russia

УДК 004.942

DOI: 10.18413/2518-1092-2025-10-3-0-8

Ганичева А.В.<sup>1</sup>  
Ганичев А.В.<sup>2</sup>

**МАТЕМАТИЧЕСКАЯ МОДЕЛЬ УСВОЕНИЯ  
УЧЕБНОГО МАТЕРИАЛА**

<sup>1</sup>) Тверская государственная сельскохозяйственная академия,  
ул. Маршала Василевского (Сахарово), 7, г. Тверь, 7170904, Россия

<sup>2</sup>) Тверской государственный технический университет,  
наб. Аф. Никитина, 22, г. Тверь, 170026, Россия

*e-mail: tgan55@yandex.ru, alexej.ganichev@yandex.ru*

**Аннотация**

Проблема формирования знаний и профессиональных навыков будущих специалистов является важнейшей в учебном процессе. Особую актуальность данная проблема приобретает с внедрением в учебный процесс цифровых технологий. Объектом исследования является образовательная система получения, усвоения и оценивания знаний учащихся. Цель данной статьи – разработка новой математической модели усвоения обучаемыми учебного материала. Для осуществления данной цели математически описаны (формализованы в виде модели) процессы усвоения лекционного учебного материала с учетом коэффициентов запоминания и усвоения. Разработана модель борьбы процесса усвоения знаний с процессом не усвоения знаний. Рассмотрены ситуации обучения одного учащегося и группы обучаемых. В основу модели положен математический аппарат разностных схем. Введено новое понятие «ступора» в учебном процессе, получены формульные соотношения, определяющие его появление. Разработанные в статье модели позволяют исследовать учебный процесс и выработать рекомендации по его совершенствованию.

**Ключевые слова:** моделирование; знания; лекция; учащиеся; коэффициент; процесс усвоения материала; запоминание

**Для цитирования:** Ганичева А.В., Ганичев А.В. Математическая модель усвоения учебного материала // Научный результат. Информационные технологии. – Т.10, №3, 2025. – С. 80-88.  
DOI: 10.18413/2518-1092-2025-10-3-0-8

Ganicheva A.V.<sup>1</sup>  
Ganichev A.V.<sup>2</sup>

**MATHEMATICAL MODEL OF LEARNING MATERIAL  
ASSIMILATION**

<sup>1</sup>) Tver State Agricultural Academy,  
7 Marshal Vasilevsky St. (Sakharovo), Tver, 7170904, Russia

<sup>2</sup>) Tver State Technical University,  
22 nab. Af. Nikitin, Tver, 170026, Russia

*e-mail: tgan55@yandex.ru, alexej.ganichev@yandex.ru*

**Abstract**

The problem of formation of knowledge and professional skills of future specialists is the most important in the educational process. This problem becomes particularly relevant with the introduction of digital technologies into the educational process. The object of the research is the educational system of obtaining, assimilating and evaluating students' knowledge. The purpose of this article is to develop a new mathematical model of learners' learning of educational material. To achieve this goal, the processes of assimilation of lecture educational material are mathematically described (formalized in the form of a model), taking into account the coefficients of memorization and assimilation. The models developed in the article allow us to explore the learning process and develop recommendations for its improvement.

**Keywords:** modeling; knowledge; lecture; students; coefficient; process of assimilation of material; memorization

**For citation:** Ganicheva A.V., Ganichev A.V. Mathematical Model of Learning Material Assimilation // Research result. Information technologies. – Т.10, №3, 2025. – P. 80-88. DOI: 10.18413/2518-1092-2025-10-3-0-8

## **ВВЕДЕНИЕ**

Внедрение в учебный процесс цифровых технологий делает проблему усвоения учебного материала особенно важной и актуальной.

Современное общество – общество, основанное на знаниях. Знания – это зафиксированная и проверенная практикой обработанная информация, которая использовалась и может многократно использоваться для принятия решений. Знания обобщают опыт, полученный человеком в ходе выполнения какой-либо практической деятельности, и представляют собой закономерности предметной области. Основным путь получения знаний – это запоминание и усвоение учебного материала в образовательных организациях. Получение и запоминание знаний в учебных заведениях высшего и среднего специального образования осуществляется преимущественно на лекционных занятиях. Несмотря на многообразие форм лекционных занятий, они имеют общую основу: во время их проведения осуществляется процесс борьбы процесса усвоения знаний с противоположным процессом не усвоения знаний. Имеется промежуточное состояние между этими процессами, которое назовем «ступор». Своевременное обнаружение данного состояния имеет особое значение при управлении учебным процессом, т.к. позволяет своевременно принять меры для вывода из него обучаемого или группы обучаемых.

Целью данной статьи является разработка математической модели процесса усвоения учебного материала, а также модели борьбы за качество данного процесса.

### **1. ОБЗОР НАУЧНЫХ ПУБЛИКАЦИЙ ПО МОДЕЛИРОВАНИЮ УЧЕБНОГО МАТЕРИАЛА**

Сложность и многогранность проблемы качественного усвоения учебного материала вызывает необходимость применения для ее решения метод педагогико-математического моделирования. Кратко рассмотрим основные направления исследований в данной области.

Концепция применения метода математического моделирования в образовательных системах сформулирована в статьях [9, 19]. В работе [22] отмечается, что математическое моделирование является мощным инструментом, который может быть использован для улучшения образования. Типология, особенности и классификация педагогических моделей рассмотрены в статьях [10, 11, 17]. В работе [21] рассмотрены различные подходы к интегрированию математического моделирования в процессы преподавания и обучения. Автор исследования [16] отмечает взаимосвязи между моделями обучения, математическим материалом и характером учащихся. Обзор литературы, отражающей различные точки зрения на рассматриваемую проблему в мировом сообществе, дан в статье [20]. В статье [5] математическая модель представления элементов образовательного процесса и построения траектории обучения рассматривается как основа для разработки образовательной информационной системы. С помощью математического моделирования, как показано в работе [4], можно исследовать различные компоненты учебного процесса. В статье [13] предложена классификация моделей, приведена доказательная база их применения для учебного процесса. Отдельные характеристики учебной информации как средства совершенствования учебного процесса рассмотрены в [4]. В статье [24] разработаны четыре математические модели эффективности обучения: 1) проверки знаний на экзаменах, 2) развития талантов, 3) карьерного роста, 4) модель повышения качества образования. Специальная панель инструментов оценки знаний разработана в [15].

Следует отметить, что многие разработанные в настоящее время модели учебного процесса обладают большой математической и вычислительной сложностью. Так, в работе [12] предлагается описывать процесс усвоения материала системой обыкновенных дифференциальных уравнений двенадцатого порядка.

Математические модели получения знаний, организации контрольно-проверочных мероприятий и оценки эффективности учебного процесса разработаны в статьях [1-3].

Наиболее близкими по решаемым задачам и используемому математическому аппарату к настоящей статье исследованиями являются работы [6-8]. В этих статьях рассматриваются разностные компьютерные модели дидактических систем «учитель-ученик». Следует отметить, что в этих исследованиях коэффициент понимания рассматривается зависящим только от одного фактора – сложности восприятия материала, а коэффициент усвоения – в диапазоне от 0,1 до 0,2. В предлагаемой работе рассматриваются коэффициенты запоминания и усвоения, которые могут определяться разными аспектами (сложностью предложенного материала, усвоением предыдущего материала, самочувствием обучаемого и т.д.).

## 2. МОДЕЛИРОВАНИЕ ПРОЦЕССА УСВОЕНИЯ ЛЕКЦИОННОГО УЧЕБНОГО МАТЕРИАЛА

Рассмотрим усвоение лекционного учебного материала. Сначала рассмотрим одного учащегося. Пусть  $r_i$  – знания, полученные на  $i$ -ой лекции,  $R_i$  – суммарные значения, полученные им в результате прослушивания  $i$  лекций,  $\alpha_i$  и  $\beta_i$  соответственно коэффициенты запоминания и усвоения, характеризующие долю предыдущего суммарного знания, которую получил и усвоил учащийся к началу лекции с номером  $i$ , т.е.  $\alpha_i, \beta_i \in (0, 1 - \delta)$ ,  $\delta = \text{const}$ ,  $0 < \delta < 1$ . При этом считаем, что имеет место равенство  $R_1 = r_1$ . Тогда имеет место формула:

$$R_i = r_i + \alpha_{i-1}\beta_{i-1}R_{i-1}, \quad i = \overline{2, m}. \quad (1)$$

Значения  $r_i$  оцениваются в баллах в результате либо летучек по основным понятиям лекции, либо в результате решения заданий на соответствующем практическом занятии. Коэффициенты  $\alpha_i$  и  $\beta_i$  проверяются либо по анкетам психологической службы, либо с дублированием материала по предыдущим летучкам и заданиям. Может использоваться свертка баллов по решению летучек, задач и заданий, данных психологической службы. Коэффициенты дробные, каждое слагаемое делится на их общую сумму.

Пусть рассматривается учебная группа, состоящая из  $n$  учащихся, и  $j$  – порядковый номер учащегося в этой группе ( $j = \overline{1, n}$ ),  $\alpha_{j,i-1}$  и  $\beta_{j,i-1}$  соответственно коэффициенты запоминания и усвоения  $j$ -ым учащимся материала, который он получит и усвоит к началу  $i$ -ой лекции, причем  $\alpha_{ji}, \beta_{ji} \in (0, 1 - \delta_j)$ ,  $0 < \delta_j < 1$ ,  $\delta = \text{const}$ ,  $r_{ji}$  – знания  $j$ -го учащегося, полученные на  $i$ -ой лекции,  $R_{ji}$  – суммарные значения, полученные им после  $i$  лекций. Тогда получим:

$$R_{ji} = r_{ji} + \alpha_{j,i-1}\beta_{j,i-1}R_{j,i-1}, \quad (2)$$

и суммарные знания данной группы можно определить по формуле:

$$\sum_{j=1}^n R_{ji} = \sum_{j=1}^n r_{ji} + \sum_{j=1}^n \alpha_{j,i-1}\beta_{j,i-1}R_{j,i-1}. \quad (3)$$

Если поделить равенство (3) на  $n$ , то получим средние значения по группе, соответственно:

$$\bar{R}_i = \frac{1}{n} \sum_{j=1}^n R_{ji}, \quad \bar{r}_i = \frac{1}{n} \sum_{j=1}^n r_{ji}. \quad (4)$$

Тогда из выражения (3) получим:

$$\bar{R}_i = \bar{r}_i + \frac{1}{n} \sum_{j=1}^n \alpha_{j,i-1}\beta_{j,i-1}R_{j,i-1}. \quad (5)$$

Таким образом, среднее значение суммарного знания группы из  $n$  обучаемых, полученное в результате прослушивания  $i$  лекций, равно среднему значению суммарного знания на  $i$ -ой лекции с учетом коэффициентов запоминания и усвоения.

Однако,

$$\frac{1}{n} \sum_{j=1}^n \alpha_{j,i-1} \beta_{j,i-1} R_{j,i-1} = \frac{1}{n} \left( \sum_{j=1}^n \alpha_{j,i-1} \beta_{j,i-1} \right) \cdot \left( \sum_{k=1}^n R_{k,i-1} \right) -$$

$$-\frac{1}{n} \sum_{\substack{k,j=1 \\ k \neq j}}^n \alpha_{j,i-1} \beta_{j,i-1} R_{k,i-1} = \overline{\alpha_i \beta_i} \cdot \sum_{j=1}^n R_{j,i-1},$$

т.к. второе слагаемое правой части суммы равно нулю, а  $\overline{\alpha_i \beta_i}$  – среднее значение произведений  $\alpha_{j,i} \cdot \beta_{j,i}$  ( $j = \overline{1, n}$ ). Отсюда с учетом соотношения (5) получаем:

$$\overline{R}_i = \overline{r}_i + \overline{\alpha_i \beta_i} \cdot \sum_{j=1}^n R_{j,i-1}, \quad (6)$$

т.е. формула (6) отличается от формулы (5) вторым слагаемым, которое равно среднему значению произведения коэффициентов запоминания и усвоения, умноженному на суммарное знание, полученное до  $i$ -ой лекции.

Из выражения (6) получаем это значение:

$$\overline{\alpha_i \beta_i} = (\overline{R}_i - \overline{r}_i) / \sum_{j=1}^n R_{j,i-1}. \quad (7)$$

Таким образом, коэффициент  $\overline{\alpha_i \beta_i}$  равен частному от деления разности между средним значением суммарного знания группы при чтении  $i$  лекций и средним значением суммарного знания группы на  $i$ -ой лекции, деленной на суммарное значение знания группы до  $i$ -ой лекции.

Коэффициент  $\overline{\alpha_i \beta_i}$  назовем обобщенным коэффициентом группового запоминания и усвоения.

Пусть  $\alpha_{ji} = \alpha_j$ ,  $\beta_{ji} = \beta_j$ ,  $r_{ji} = \tilde{r}_j$ , т.е. у учащегося с номером  $j$  коэффициенты усвоения и запоминания не изменяются. Тогда формула (1) преобразуется к виду:

$$R_i = \tilde{r}_j \frac{1 - (\alpha_j \beta_j)^i}{1 - (\alpha_j \beta_j)}. \quad (8)$$

Не нарушая общности, рассмотрим доказательство этой формулы для случая  $j=1$ ,  $i=5$ . Имеем следующее соотношение:

$$R_5 = \tilde{r}_1 + \alpha_1 \beta_1 R_4 = \tilde{r}_1 + \alpha_1 \beta_1 (\tilde{r}_1 + \alpha_1 \beta_1 (\tilde{r}_1 + \alpha_1 \beta_1 (\tilde{r}_1 + \alpha_1 \beta_1 (\tilde{r}_1 + \alpha_1 \beta_1 \tilde{r}_1))) =$$

$$= \tilde{r}_1 + \alpha_1 \beta_1 \tilde{r}_1 + \alpha_1^2 \beta_1^2 \tilde{r}_1 + \alpha_1^3 \beta_1^3 \tilde{r}_1 + \alpha_1^4 \beta_1^4 \tilde{r}_1 = \tilde{r}_1 \frac{1 - (\alpha_1 \beta_1)^5}{1 - (\alpha_1 \beta_1)}.$$

### 3. МОДЕЛЬ БОРЬБЫ ПРОЦЕССА УСВОЕНИЯ ЗНАНИЙ С ПРОЦЕССОМ НЕ УСВОЕНИЯ ЗНАНИЙ

Рассмотрим модель борьбы процесса усвоения знаний с процессом не усвоения знаний. Сначала рассмотрим эту модель для одного учащегося.

Тогда усвоение происходит по модели (1), а не усвоение пусть описывается аналогичной формулой (но при этом  $Q_i < 0$ ):

$$Q_i = q_i + \gamma_{i-1} \delta_{i-1} Q_{i-1}, \quad i = \overline{2, m}$$

и  $Q_1 = q_1$ .

Здесь  $\gamma_{i-1}$  и  $\delta_{i-1}$  – соответственно коэффициенты не усвоения и не запоминания. Эти коэффициенты определяются либо по баллам решения задач и заданий, либо по результатам анкетирования психологической службы. Причем чем хуже результат, тем отрицательное число, соответствующее данному показателю, больше по модулю. При этом рассматриваются приведенные коэффициенты, т.е. с учетом деления на их общую сумму. Отрицательные

коэффициенты в этой сумме берутся по модулю, т.е. общая сумма приведенных коэффициентов равна единице.

Рассмотрим лекцию с номером  $i$ .

Предположим, что  $r_i q_i < 0$ , т.е.  $q_i < 0$ . Кроме того,  $|q_i| > \gamma_{i-1} \delta_{i-1} Q_{i-1}$ . Тогда усвоение побеждает не усвоение, если  $|Q_i| < R_i$ , не усвоение побеждает усвоение, если  $|Q_i| > R_i$ , если  $|Q_i| = R_i$ , то имеет место, так называемое в психологии состояние «ступора». В учебном процессе это ситуация ожидания и отсутствия реакции на поступающую на лекции информацию. Данная ситуация может измениться в ту или иную сторону под воздействием внутренних и внешних факторов учебного процесса. Например, обучаемый отвлекся и перестал понимать излагаемый преподавателем материал, но своевременное замечание или пояснение помогло ему понять и освоить пропущенный фрагмент изучаемого материала. Возможно и противоположное развитие ситуации: обучаемый продолжил отвлекаться и перестал понимать суть данного фрагмента, а возможных, дидактически связанных с ним.

В первом случае принимается решение о том, что процесс усвоения происходит нормально, во втором случае необходимы соответствующие меры (беседы, дополнительные задания и консультации). Рассмотрим подробно третий случай. Для него имеем:

$$r_i + \alpha_{i-1} \beta_{i-1} R_{i-1} + q_i + \gamma_{i-1} \delta_{i-1} Q_{i-1} = 0. \quad (9)$$

Предположим, что во время «ступора»  $r_i = -q_i$ . Тогда выражение (9) принимает следующий вид:

$$\alpha_{i-1} \beta_{i-1} R_{i-1} = \gamma_{i-1} \delta_{i-1} |Q_{i-1}|,$$

при этом получается равенство  $|Q_i| = R_i$ .

В этом случае получаем:

$$\alpha_{i-1} \beta_{i-1} = -\gamma_{i-1} \delta_{i-1}, \quad (10)$$

т.е. произведение коэффициентов запоминания и усвоения материала, предшествующего  $i$ -ой лекции, равно соответствующему произведению коэффициентов не запоминания и не усвоения.

Из выражения (10) можно найти номер лекции, на которой произойдет «ступор» у данного учащегося, а именно:

$$\alpha_{i-1} \beta_{i-1} = (1 - \alpha_{i-1})(1 - \beta_{i-1}), \text{ т.е. } \alpha_{i-1} \beta_{i-1} = 1. \quad (11)$$

Коэффициенты известны, находим номер лекции  $i - 1$  для выполнения условия (11).

Рассмотрим данный вопрос для группы из  $n$  обучаемых. В этом случае имеет место формула не усвоения, аналогичная формуле (8), но с данными по не усвоению, т.е. получаем выражение:

$$Q_i = \tilde{q}_j \frac{1 - (\gamma_j \delta_j)^i}{1 - (\gamma_j \delta_j)}. \quad (12)$$

В случае «ступора» группы  $|Q_i| = R_i$  и предполагаем, что  $r_j = q_j$ . Тогда имеет место соотношение

$$\frac{1 - (\alpha_j \beta_j)^i}{1 - (\alpha_j \beta_j)} = \frac{1 - (\gamma_j \delta_j)^i}{1 - (\gamma_j \delta_j)}. \quad (13)$$

Обозначим для краткости, но не нарушая общности,  $\gamma_j \delta_j$  через  $a$ ,  $\alpha_j \beta_j$  – через  $b$ . Воспользуемся свойством пропорции и преобразуем (13) к следующему виду:

$$1 - a - b^i + ab^i = 1 - b - a^i + ba^i.$$

Отсюда получаем формулу:

$$(a - 1)(b^i - 1) = (b - 1)(a^i - 1),$$

или имеет место равенство:

$$\frac{1-a}{1-b} = \frac{1-a^i}{1-b^i}. \quad (14)$$

Представим выражения  $1-a^i$  и  $1-b^i$  в виде:

$$1-a^i = (1-a)(a^{i-1} + a^{i-2} + \dots + a + 1);$$

$$1-b^i = (1-b)(b^{i-1} + b^{i-2} + \dots + b + 1).$$

С учетом этих двух соотношений выражение (14) преобразуется к следующему виду:

$$b^{i-1} + b^{i-2} + \dots + b + 1 - (a^{i-1} + a^{i-2} + \dots + a + 1) = 0. \quad (15)$$

Пусть  $b = ka$ , тогда из выражения (15) получаем:

$$a^{i-1}(k^{i-1} - 1) + a^{i-2}(k^{i-2} - 1) + \dots + a(k - 1) = 0.$$

Заметим, что если  $k > 1$ , то левая часть последнего равенства больше, чем нуль, если  $k < 1$ , то левая часть меньше, чем нуль. Равенство достигается в том и только в том случае, когда  $k = 1$ , т.е. имеет место равенство  $b = a$ . В этом случае выполняется условие  $\gamma_j \delta_j = \alpha_j \beta_j$ .

Таким образом, соотношение  $\gamma_j \delta_j = \alpha_j \beta_j$  является необходимым и достаточным условием для «ступора» группы при условии, что коэффициенты усвоения и запоминания, а также коэффициенты не усвоения и не запоминания для каждого учащегося не изменяются от лекции к лекции.

Данное обстоятельство подтверждает известный результат педагогических исследований о том, что процесс изложения учебного материала должен происходить с увеличением количества знаний в течение учебного процесса (так называемый процесс непрерывного обучения). Непрерывное обучение достигается качественным составлением учебного плана подготовки специалистов данного профиля и планов изучаемых дисциплин в течение всего периода обучения.

### **ЗАКЛЮЧЕНИЕ**

Основными результатами данной статьи являются:

- 1) разработана модель усвоения лекционного учебного материала одним учащимся и группой обучающихся;
- 2) получена формула обобщенного коэффициента группового запоминания и усвоения;
- 3) доказано утверждение о том, что знания, полученные  $j$ -ым учащимся в результате прослушивания  $i$ -ой лекции при условии, что коэффициенты запоминания и усвоения от лекции к лекции не изменяются, выражаются через сумму геометрической прогрессии.
- 4) разработана модель борьбы процесса усвоения знаний с процессом не усвоения знаний для одного учащегося и группы обучающихся;
- 5) получены условия, когда усвоение побеждает не усвоение и наоборот, когда не усвоение доминирует над усвоением;
- 6) введено понятие «ступора» в учебном процессе, когда отсутствует реакция на поступающую на лекции информацию;
- 7) получены формульные соотношения, позволяющие распознать его появление.

Направлением дальнейших исследований является адаптация разработанных моделей к условиям наличия нечеткой, неполной и плохо структурированной информации об учебном процессе, а также использование подхода, объединяющего количественные и качественные методы сбора данных [18]. Многие направления перспективных исследований в области математического моделирования в образовании сформулированы в монографии [23].

### **Список литературы**

1. Ганичева А.В. Моделирование показателей учебного процесса // В мире научных открытий. – 2011. – № 10-2 (22). – С. 1016.

2. Ганичева А.В. Оптимальное решение и оценка полезности организационных вопросов // Ярославский педагогический вестник. – 2011. – Т. 3. – № 2. С. 53.
3. Ганичева А.В. Оценка эффективности процесса обучения // Интеллект. Инновации. Инвестиции. – 2011. – № 2. – С. 134-137.
4. Гефан Г.Д. Подходы к математическому моделированию учебного процесса // Тенденции развития науки и образования. – 2024. – № 111-1. – С. 60-63.
5. Киселева О.М., Быков А.А., Скуратова Н.А. Математическая модель как основа для разработки образовательной информационной системы // Естественные и технические науки. – 2023. – № 10 (185). – С. 12-16.
6. Майер Р.В. Закономерности усвоения, забывания и имитационное моделирование обучения // Инновации в образовании. – 2017. – № 5. – С. 145-152.
7. Майер Р.В. Методология компьютерного моделирования дидактических систем «учитель-ученик» // Вестник педагогического опыта. – 2024. – № 4 (62). – С. 14-21.
8. Майер Р.В. Модель дидактического процесса, учитывающая распределение элементов учебного материала по сложности // Ученые записки Крымского федерального университета имени В.И. Вернадского. Социология. Педагогика. Психология. – 2021. – Т. 7 (73). – № 4. – С. 94-104.
9. Нахман А.Д. Компетенция математического моделирования в контексте современной образовательной парадигмы // Научное обозрение. Педагогические науки. – 2017. – № 3. – С. 71-79.
10. Писаренко В.И. Педагогические модели: типология и особенности // Проблемы современного образования. – 2024. – № 1. – С. 58-76.
11. Сапронов В.А. Классификация моделей в педагогике // Педагогика и просвещение. – 2024. – № 1. – С. 84-100.
12. Титов Б.А., Рябинова Е.Н. Математическая модель усвоения учебного материала // Проблемы и перспективы развития двигателестроения: материалы межд. науч.-техн. конф. Ч. 1. – Самара: СГАУ, 2011. – С. 117-118.
13. Тишина Е.М. Методологические основы использования принципов математического моделирования в учебном процессе // Образование и наука в современном мире. – Инновации. – 2018. – № 5 (18). – С. 28-33.
14. Чуйко Л.В., Капацина Н.П. Моделирование учебной информации как условие оптимального управления процессом обучения // Вестник Приднестровского университета. Серия: Гуманитарные науки. – 2017. – № 1 (55). – С. 92-98.
15. An Analytical Dashboard of Collaborative Activities for the Knowledge Building / C. Gutiérrez-Braojos [et al.] // Technology, Knowledge and Learning. – 2023. – P. 1-27.
16. Analysis of Learning Model and Mathematical Material to Build Students' Character through / S. Fadillah [et al.] // Mathematics Learning Journal Of Education, Teaching and Learning. – 2019. – 4 (2). – P. 345-349.
17. Bakwai B., Yisa H.M. Mathematical modelling: a scientific tool in educational management // Al-hikma Journal of Education. – 2014. – 1 (2). – P. 38-46.
18. Bulut N., Ferri R.B. Mathematical Modelling and Education for Sustainable Development in Pre-Service Primary Teacher Education // Educ. Sci. – 2025. 15 (2), 248 p.
19. Montejo-Gámez J., Fernández-Ahumada E., Adamuz-Povedano N. A Tool for the Analysis and Characterization of School Mathematical Models / C. Gutiérrez-Braojos [et al.] // Mathematics. – 2021. – 9 (13). – P. 1569.
20. Five different perspectives on mathematical modeling in mathematics education / A. Abassian A. [et al.] // Investigations in Mathematics Learning. – 2020. – 12 (1). – P. 53-65.
21. Kaiser G. Mathematical Modelling and Applications in Education // Encyclopedia of Mathematics Education. Springer, Cham. – 2020. P. 553–561.
22. Kumar M., Behera B. Mathematical Modelling: A Pedagogical Tool for Competency-Based Learning // Asian Journal for Mathematics Education. – 1 (3). – P. 359-380.
23. Stillman G.A., Brown J.P. Lines of Inquiry in Mathematical Modelling Research in Education: Monograph. Springer ICME-13, 2019. – 259 p.
24. Sun B. Mathematical Models of Learning Efficiency // EURASIA. Journal of Mathematics Science and Technology Educatio. – 2017. – 13 (7). P. 4261-4270.

## References

1. Ganicheva A.V. Modeling of educational process indicators // In the World of Scientific Discoveries. – 2011. No. 10-2 (22). – P. 1016.
2. Ganicheva A.V. Optimal solution and assessment of the usefulness of organizational issues // Yaroslavl Pedagogical Bulletin. – 2011. – Vol. 3. – No. 2. – P. 53.
3. Ganicheva A.V. Assessment of the effectiveness of the learning process // Intelligence. Innovation. Investments. – 2011. – No. 2. – P. 134-137.
4. Gefan G.D. Approaches to mathematical modeling of the educational process // Trends in the development of science and education. – 2024. – No. 111-1. – P. 60-63.
5. Kiseleva O.M., Bykov A.A., Skuratova N.A. Mathematical Model as a Basis for Developing an Educational Information System // Natural and technical sciences. – 2023. – No. 10 (185). – P. 12-16.
6. Majer R.V. Patterns of Assimilation, Forgetting, and Simulation Modeling of Learning // Innovations in Education. – 2017. No. 5. – P. 145-152.
7. Majer R.V. Methodology of Computer Modeling of Teacher-Student Didactic Systems // Bulletin of Pedagogical Experience. – 2024. – No. 4 (62). – P. 14-21.
8. Majer R.V. Model of the Didactic Process Taking into Account the Distribution of Elements of Educational Material by Complexity // Scientific Notes of V.I. Vernadsky Crimean Federal University. Sociology. Pedagogy. Psychology. – 2021. – Vol. 7 (73). – No. 4 – P. 94-104.
9. Nahman A.D. Competence of mathematical modeling in the context of the modern educational paradigm // Scientific Review. Pedagogical Sciences. – 2017. – No. – P. 71-79.
10. Pisarenko V.I. Pedagogical models: typology and features // Problems of modern education. – 2024. – No. 1 – P. 58-76.
11. Saponov V.A. Classification of models in pedagogy // Pedagogy and education. – 2024. – No. 1. – P. 84-100.
12. Titov B.A., Ryabinova E.N. Mathematical model of assimilation of educational material // Problems and prospects of development of engine building: Proc. of the Int. Scientific and Technical Conf. Part 1. – Samara: SGAU, 2011. – P. 117-118.
13. Tishina E.M. Methodological foundations for using the principles of mathematical modeling in the educational process // Education and Science in the Modern World. Innovations. – 2018. – No. 5 (18). – P. 28-33.
14. Chuyko L.V., Kapatsina N.P. Modeling educational information as a condition for optimal management of the learning process // Bulletin of the Pridnestrovian University. Series: Humanities. – 2017. – No. 1 (55). – P. 92-98.
15. An Analytical Dashboard of Collaborative Activities for the Knowledge Building / C. Gutiérrez-Braojos [et al.] // Technology, Knowledge and Learning. – 2023. – P. 1-27.
16. Analysis of Learning Model and Mathematical Material to Build Students' Character through / S. Fadillah [et al.] // Mathematics Learning Journal Of Education, Teaching and Learning. – 2019. – 4 (2). – P. 345-349.
17. Bakwai B., Yisa H.M. Mathematical modelling: a scientific tool in educational management // Al-hikma Journal of Education. – 2014. -1 (2). – P. 38-46.
18. Bulut N., Ferri R.B. Mathematical Modelling and Education for Sustainable Development in Pre-Service Primary Teacher Education // Educ. Sci. – 2025. 15 (2), 248 p.
19. Montejo-Gámez J., Fernández-Ahumada E., Adamuz-Povedano N. A Tool for the Analysis and Characterization of School Mathematical Models / C. Gutiérrez-Braojos [et al.] // Mathematics. – 2021. – 9 (13). – P. 1569.
20. Five different perspectives on mathematical modeling in mathematics education / A. Abassian A. [et al.] // Investigations in Mathematics Learning. – 2020. – 12 (1). – P. 53-65.
21. Kaiser G. Mathematical Modelling and Applications in Education // Encyclopedia of Mathematics Education. Springer, Cham. – 2020. P. 553–561.
22. Kumar M., Behera B. Mathematical Modelling: A Pedagogical Tool for Competency-Based Learning // Asian Journal for Mathematics Education. – 1 (3). – P. 359-380.
23. Stillman G.A., Brown J.P. Lines of Inquiry in Mathematical Modelling Research in Education: Monograph. Springer ICME-13, 2019. – 259 p.
24. Sun B. Mathematical Models of Learning Efficiency // EURASIA. Journal of Mathematics Science and Technology Educatio. – 2017. – 13 (7). P. 4261-4270.

**Ганичева Антонина Валериановна**, кандидат физико-математических наук, доцент, профессор кафедры физико-математических дисциплин и информационных технологий, Тверская государственная сельскохозяйственная академия, г. Тверь, Россия

**Ганичев Алексей Валерианович**, старший преподаватель кафедры информатики и прикладной математики, Тверской государственный технический университет, г. Тверь, Россия

**Ganicheva Antonina Valerianovna**, Candidate of Physical and Mathematical Sciences, Associate Professor, Professor of the Department Physical and Mathematical Disciplines and Information Technology, Tver State Agricultural Academy, Tver, Russia

**Ganichev Alexey Valerianovich**, Senior lecturer of the Department of Informatics and Applied Mathematics, Tver State Technical University, Tver, Russia