

# ПРИКАСПИЙСКИЙ ЖУРНАЛ



УПРАВЛЕНИЕ И ВЫСОКИЕ  
ТЕХНОЛОГИИ

2024  
№3 (67)



ISSN 2074-1707

АСТРАХАНСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ  
ИМЕНИ В. Н. ТАТИЩЕВА

## **ПРИКАСПИЙСКИЙ ЖУРНАЛ: управление и высокие технологии**

НАУЧНО-ТЕХНИЧЕСКИЙ ЖУРНАЛ

**2024  
№ 3 (67)**

Журнал включен в перечень рецензируемых научных изданий, рекомендованных ВАК России для публикации основных научных результатов диссертаций на соискание ученой степени кандидата наук, на соискание ученой степени доктора наук по следующим научным специальностям.

**Группа специальностей 1.2 «Компьютерные науки и информатика»:**

1.2.2 – Математическое моделирование, численные методы и комплексы программ (технические науки).

**Группа специальностей 2.2 «Электроника, фотоника, приборостроение и связь»:**

2.2.4 – Приборы и методы измерения (по видам измерений) (технические науки);

2.2.11 – Информационно-измерительные и управляющие системы (технические науки);

2.2.12 – Приборы, системы и изделия медицинского назначения (технические науки).

**Группа специальностей 2.3 «Информационные технологии и телекоммуникации»:**

2.3.1 – Системный анализ, управление и обработка информации (технические науки);

2.3.4 – Управление в организационных системах (технические науки);

2.3.5 – Математическое и программное обеспечение вычислительных систем, комплексов и компьютерных сетей (технические науки);

2.3.6 – Методы и системы защиты информации, информационная безопасность (технические науки).

Журнал входит в базу данных Ulrich's Periodicals Directory.

Астрахань  
Астраханский государственный университет имени В. Н. Татищева  
2024

Рекомендовано к печати редакционно-издательским советом  
Астраханского государственного университета имени В. Н. Татищева

**ПРИКАСПИЙСКИЙ ЖУРНАЛ:  
управление и высокие технологии**

НАУЧНО-ТЕХНИЧЕСКИЙ ЖУРНАЛ

**2024  
№ 3 (67)**

**Редакционная коллегия**

*И.М. Азмухамедов*, доктор технических наук, профессор, декан факультета цифровых технологий и кибербезопасности, профессор кафедры «Информационная безопасность» Астраханского государственного университета им. В. Н. Татищева (главный редактор)

*И.В. Аникин*, доктор технических наук, профессор, заведующий кафедрой «Системы информационной безопасности» Казанского национального исследовательского технического университета им. А.Н. Туполева – КАИ

*А.А. Большаков*, доктор технических наук, профессор, профессор кафедры «Системы автоматизированного проектирования и управления» Санкт-Петербургского государственного технологического института (технического университета)

*Л.А. Демидова*, доктор технических наук, профессор, профессор кафедры «Вычислительной и прикладной математики» Рязанского государственного радиотехнического университета (г. Рязань)

*А.С. Катасёв*, доктор технических наук, профессор, профессор кафедры систем информационной безопасности Казанского национального исследовательского технического университета им. А.Н. Туполева – КАИ»

*И.Ю. Квятковская*, доктор технических наук, профессор, директор Института информационных технологий и коммуникаций Астраханского государственного технического университета

*А.Г. Кравец*, доктор технических наук, профессор, профессор кафедры «Системы автоматизированного проектирования и поискового конструирования» Волгоградского государственного технического университета

*В.Ю. Кузнецова*, кандидат технических наук, старший преподаватель кафедры информационной безопасности Астраханского государственного университета им. В. Н. Татищева

*Ю.В. Литовка*, доктор технических наук, профессор, профессор кафедры «Системы автоматизированной поддержки принятия решений» Тамбовского государственного технического университета

*А.М. Лихтер*, доктор технических наук, профессор, заведующий кафедрой «Общая физика» Астраханского государственного университета им. В. Н. Татищева

*А.А. Лобатый*, доктор технических наук, профессор, заведующий кафедрой «Информационные системы и технологии» Белорусского национального технического университета (Республика Беларусь, г. Минск)

*Е.В. Никульчев*, доктор технических наук, профессор, профессор кафедры «Управление и моделирование систем» Московского технологического университета (МИРЭА) (г. Москва)

*В.О. Осипян*, доктор физико-математических наук, доцент, профессор кафедры «Информационные технологии» Кубанского государственного университета (г. Краснодар)

*И.Ю. Петрова*, доктор технических наук, профессор, первый проректор Астраханского государственного архитектурно-строительного университета, заведующая кафедрой САПР Астраханского государственного архитектурно-строительного университета

*А.В. Рыбиков*, кандидат физико-математических наук, директор «Физико-математического института» Астраханского государственного университета им. В.Н. Татищева; доцент кафедры электротехники, электроники и автоматики Астраханского государственного университета им. В.Н. Татищева

*А.В. Скрипаль*, доктор физико-математических наук, профессор, заведующий кафедрой «Медицинская физика» Саратовского национального исследовательского государственного университета им. Н.Г. Чернышевского

*И.Б. Старченко*, доктор технических наук, профессор, ООО «Параметрика», научный руководитель (г. Таганрог Ростовской области)

*Т.Л. Тен*, доктор физико-математических наук, профессор кафедры «Информационно-вычислительные системы» Карагандинского экономического университета (Республика Казань, г. Караганда)

*Е.Н. Тищенко*, доктор экономических наук, профессор, заведующий кафедрой «Информационные технологии и защита информации» Ростовского государственного экономического университета (РИНХ) – г. Ростов-на-Дону

*С.А. Филит*, доктор технических наук, профессор, профессор кафедры «Биомедицинская инженерия» Юго-Западного государственного университета (г. Курск)

*Л.Р. Фионова*, доктор технических наук, профессор, декан факультета вычислительной техники, заведующая кафедрой «Информационное обеспечение управления и производства» Пензенского государственного университета

*В.А. Цимбал*, заслуженный деятель науки РФ, доктор технических наук, профессор, профессор кафедры «Автоматизированные системы управления» (Филиал Военной академии РВСН им. Петра Великого МО в г. Серпухов Московской области)

*Н.К. Юрков*, заслуженный деятель науки РФ, доктор технических наук, профессор, заведующий кафедрой «Конструирование и производство радиоаппаратуры» Пензенского государственного университета

*N.A. Kolesova*, PhD, Check Point Software Technologies LTD, Tel-Aviv, Israel

*Serg Miranda*, PhD (Toulouse University, France), – Master thesis at UCLA (University of California, Los Angeles with an INRIA Scholarship), Professor of Computer Science, University of Nice – Sophia Antipolis (Nice, France), Director of the CS dept. and MBDS innovation lab ([www.mbds-fr.org](http://www.mbds-fr.org))

Журнал выходит 4 раза в год  
Все материалы, поступающие в редколлегию журнала,  
проходят независимое рецензирование

© Астраханский государственный университет  
имени В. Н. Татищева, 2024  
© Гайфитдинова С. Ю., дизайн обложки, 2024

ASTRAKHAN TATISHCHEV STATE UNIVERSITY

**PRIKASPIYSKIY ZHURNAL:  
Upravlenie i Vysokie Tekhnologii**

**CASPIAN JOURNAL:  
Control and High Technologies**

A SCIENTIFIC AND TECHNICAL JOURNAL

**2024  
No. 3 (67)**

The journal is included in the list of the reviewed scientific journals recommended by VAK of Russia for the publication of the main scientific results of theses for the candidate of science degree, for the doctor of science degree on the following scientific specialties.

**Group of specialties 1.2 “Computer science and informatics”:**

1.2.2 – Mathematical modelling, numerical methods and complexes of programmes (technical sciences).

**Group of specialties 2.2 “Electronics, photonics, instrument engineering and communication”:**

2.2.4 – Instruments and methods of measurement (by type of measurement) (technical sciences);

2.2.11 – Information-measuring and control systems (technical sciences);

2.2.12 – Medical devices, systems and products (technical sciences).

**Group of specialties 2.3 “Information technologies and telecommunications”:**

2.3.1 – System analysis, information control and processing (technical sciences);

2.3.4 – Management in organizational systems (technical sciences);

2.3.5 – Mathematical software and software for computing systems, complexes and computer networks (technical sciences);

2.3.6 – Information security methods and systems, information security (technical sciences).

The journal is included into the database Ulrich’s Periodicals Directory.

Astrakhan  
Astrakhan Tatishchev State University  
2024

Recommended by the Editorial and Publishing Board  
of Astrakhan Tatishchev State University

**CASPIAN JOURNAL:  
Control and High Technologies**

A SCIENTIFIC AND TECHNICAL JOURNAL

**2024**

**No. 3 (67)**

**Editorial Board**

*I.M. Azhmukhamedov*, Doct. Sci. (Engineering), Professor, Dean of the Faculty of Digital Technologies and Cybersecurity, Professor of Information Security Department, Astrakhan Tatishchev State University (**Editor-in-Chief**)

*I.V. Anikin*, Doct. Sci. (Engineering), Professor, Head of Information Security System Department, Kazan National Research Technical University named after A.N. Tupolev – KAI

*A.A. Bolshakov*, Doct. Sci. (Engineering), Professor of «Systems of Automated Design Engineering and Control» department, St. Petersburg State Technological Institute (Technical University)

*L.A. Demidova*, Doct. Sci. (Engineering), Professor, Professor of the Computational and Applied Mathematics Department, Ryazan State Radio Engineering University (Ryazan)

*A.S. Katasev*, Doct. Sci. (Engineering), Associate Professor, Professor of the Department of Information Security Systems, Kazan National Research Technical University. A.N. Tupolev – KAI

*I.Yu. Kvyatkovskaya*, Doct. Sci. (Engineering), Professor, Head of “Information Technologies and Communications” Institute of the Astrakhan State Technical University

*A.G. Kravets*, Doct. Sci. (Engineering), Professor, Professor of the Automated Design Engineering Systems and Search Constructing Department, Volgograd State Technical University

*V.Yu. Kuznetsova*, Cand. Sci. (Engineering), Senior Lecturer of Information Security Department, Astrakhan Tatishchev State University

*Yu.V. Litovka*, Doct. Sci. (Engineering), Professor, Professor of the Department of Automated Support System for Decision-Making, Tambov State Technical University

*A.M. Likhter*, Doct. Sci. (Engineering), Professor, Head of the Department of General Physics, Astrakhan Tatishchev State University

*A.A. Lobaty*, Doct. Sci. (Engineering), Professor, Head of Information Systems and Technologies Department, Belarusian National Technical University (Belarus, Minsk)

*E.V. Nikulchev*, Doct. Sci. (Engineering), Professor, Professor of the System Management and Modeling Department, Moscow Technological University (Moscow)

*V.O. Osipyan*, Doct. Sci. (Physics and Mathematics), Professor of the Kuban State University (Krasnodar)

*I.Yu. Petrova*, Doct. Sci. (Engineering), Professor, First Vice-Rector of the Astrakhan State Architectural and Construction University, Head of the CAD department of Astrakhan State Architectural and Construction University

*A.V. Rybakov*, Cand. Sci. (Physics and Mathematics), Director of the Institute of Physics and Mathematics, Astrakhan Tatishchev State University

*A.V. Skripal*, Doct. Sci. (Physics and Mathematics), Professor, Head of Medical Physics Department of the Saratov National Research State University named after N.G. Chernyshevsky

*I.B. Starchenko*, Doct. Sci. (Engineering), Professor, OOO «Parametrica» (Taganrog, Rostov Oblast), Research Supervisor

*T.L. Ten*, Doct. Sci. (Engineering), Professor, Karaganda Economic University (Republic of Kazakhstan, Karaganda)

*E.N. Tishchenko*, Doct. Sci. (Economics), Professor, Head of the Information Technologies & Information Security Department, Rostov State University of Economics, Rostov-on-Don

*S.A. Filist*, Doct. Sci. (Engineering), Professor, Professor of Biomedical Engineering Department, Southwest State University (Kursk)

*L.R. Fionova*, Doct. Sci. (Engineering), Professor, Dean of the Computer Technology Faculty, Head of the Department «Information Support of Management and Production, Penza State University

*V.A. Tsimbal*, Doct. Sci. (Engineering), Honored Worker of Science of the Russian Federation, Professor, Professor of the Automated Control Systems Department (Branch of the Military Academy of the Russian Strategic Missile Forces named after Peter the Great of the Moscow Oblast, Serpukhov, Moscow Oblast)

*N.K. Yurkov*, Honored worker of science of the Russian Federation, Doct. Sci. (Engineering), Professor, Head of the department «Designing and production of the radio equipment», Penza State University

*N.A. Kolesova*, PhD, Check Point Software Technologies LTD, Tel-Aviv, Israel

*Serg Miranda*, PhD (Toulouse University, France), – Master thesis at UCLA (University of California, Los Angeles with an INRIA Scholarship), Professor of Computer Science dept., University of Nice – Sophia Antipolis (Nice, France), Director of the CS department and MBDS innovation lab ([www.mbd.fr.org](http://www.mbd.fr.org))

The journal is published four times a year  
All materials that come to the Editorial Board of the journal  
are subject to independent peer-review

© Tatishchev Astrakhan State University, 2024  
© S. Yu. Gayfitdinova, cover design, 2024

# СОДЕРЖАНИЕ

## ИНФОРМАТИКА, ВЫЧИСЛИТЕЛЬНАЯ ТЕХНИКА И УПРАВЛЕНИЕ

### **МАТЕМАТИЧЕСКОЕ И ПРОГРАММНОЕ ОБЕСПЕЧЕНИЕ ВЫЧИСЛИТЕЛЬНЫХ МАШИН, КОМПЛЕКСОВ И КОМПЬЮТЕРНЫХ СЕТЕЙ**

#### **Д. С. Пономарёв**

Применение корреляционного анализа с примерами  
на языках программирования «Python» и «R» для исследования  
статистических данных из форм отчетности пенитенциарной системы..... 9–15

#### **В. Ю. Кузнецова, А. Р. Зоева**

Информационная система обработки данных о пациентах  
для диагностики аномальных маточных кровотечений ..... 16–23

### **СИСТЕМНЫЙ АНАЛИЗ, УПРАВЛЕНИЕ И ОБРАБОТКА ИНФОРМАЦИИ**

#### **С. Ю. Петрова**

Структурный поиск скрытых связей в онтологической базе знаний  
дефектных состояний электрооборудования..... 24–35

#### **Т. А. Глазунов, А. А. Соколов, В. П. Маликов**

Классификация методов распознавания конструктивных элементов  
на чертежах технических устройств..... 36–43

#### **О. И. Захарова, С. Г. Бедняк, Д. А. Шамрин**

Сравнительный анализ методов машинного обучения  
с применением методов снижения размерности  
для классификации жестов рук..... 44–48

#### **А. Г. Бондаренко, А. Г. Кравец**

Инструменты прогнозирования технологического развития  
на основе данных из открытых источников:  
систематическое исследование русскоязычных документов..... 49–62

### **МЕТОДЫ И СИСТЕМЫ ЗАЩИТЫ ИНФОРМАЦИИ, ИНФОРМАЦИОННАЯ БЕЗОПАСНОСТЬ**

#### **Е. А. Кушко, И. И. Трофимычев**

Метод защиты от исследования локальной вычислительной сети  
на основе реконфигурации топологии сетевого уровня..... 63–72

#### **Н. О. Демин, Р. Ю. Демина, И. М. Ажмухамедов, А. М. Меркулова**

Методика оценки эффективности средств защиты от DoS и DDoS-атак ..... 73–80

#### **В. А. Частикова, Е. В. Шоничев**

Разработка модели поведенческой аналитики пользователей  
и сущностей на основе частного профилирования ..... 81–88

#### **И. А. Проворнов, И. А. Калмыков**

Разработка математической модели  
отказоустойчивого SPN-шифратора AES ..... 89–101

**МАТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ,  
ЧИСЛЕННЫЕ МЕТОДЫ И КОМПЛЕКСЫ ПРОГРАММ**

**П. Ф. Чернавин, Н. П. Чернавин,  
Ф. П. Чернавин, И. А. Черняев**

Минимизация числа входных признаков  
в задачах медицинской диагностики..... 102–109

**Ф. К. Ильясов, А. К. Аюпова**

Морфологический анализ структур высушенных капель  
водно-солевого раствора сывроточного альбумина,  
обработанного холодной плазмой..... 110–116

**ПРИБОРОСТРОЕНИЕ, МЕТРОЛОГИЯ  
И ИНФОРМАЦИОННО-ИЗМЕРИТЕЛЬНЫЕ  
ПРИБОРЫ И СИСТЕМЫ**

**ПРИБОРЫ И МЕТОДЫ ИЗМЕРЕНИЯ  
(ПО ВИДАМ ИЗМЕРЕНИЙ)**

**А. Н. Брысин, Д. Р. Арсентьев, А. С. Новикова,  
Ю. А. Журавлева, О. Е. Железникова, О. Ю. Коваленко**

Разработка и программно-аппаратная реализация  
сдвоенного 14-разрядного АЦП ..... 117–127

**ПРАВИЛА ДЛЯ АВТОРОВ ..... 128**

# CONTENTS

## INFORMATICS, COMPUTER TECHNIQUE AND CONTROL

### **MATHEMATICAL SOFTWARE AND SOFTWARE FOR COMPUTING MACHINES, COMPLEXES AND COMPUTER NETWORKS**

#### **D. S. Ponomarev**

Application of correlation analysis with examples  
in Python and R to study statistical data  
from reporting forms of the federal penalty service..... 9–15

#### **V. Yu. Kuznetsova, A. R. Zueva**

Information system for processing patient data  
for diagnosis of abnormal uterine bleeding..... 16–23

### **SYSTEM ANALYSIS, CONTROL AND INFORMATION PROCESSING**

#### **S. Yu. Petrova**

Structural search of hidden relations  
in ontological knowledge base  
of electrical equipment faulty states..... 24–35

#### **T. A. Glazunov, A. A. Sokolov, V. P. Malikov**

Classification of methods for recognizing structural elements  
in drawings of technical devices ..... 36–43

#### **O. I. Zaharova, S. G. Bednyak, D. A. Shamrin**

Comparative analysis of machine learning methods  
with application of size reduction methods  
for hand gesture classification..... 44–48

#### **A. G. Bondarenko, A. G. Kravets**

Open data-based tools for technological development forecasting:  
systematic mapping study of Russian-language sources..... 49–62

### **METHODS AND SYSTEMS OF INFORMATION PROTECTION, INFORMATION SECURITY**

#### **E. A. Kushko, I. I. Trofimychev**

Local network exploration countering method  
based on network layer topology reconfiguration ..... 63–72

#### **N. O. Demin, R. Yu. Demina, I. M. Azhmuamedov, A. M. Merkulova**

Methodology for evaluating the effectiveness  
of protection against DoS and DDoS-attacks..... 73–80

#### **V. A. Chastikova, E. V. Shonichev**

Development of a behavioral analytics model  
for users and entities based on personal profiling ..... 81–88

#### **I. A. Provornov, I. A. Kalmykov**

Development of a mathematical model  
of a fault-tolerant implementation of the SPN encoder..... 89–101

**MATHEMATICAL MODELLING, NUMERICAL METHODS  
AND PROGRAM SYSTEMS**

**P. F. Chernavin, N. P. Chernavin, F. P. Chernavin, I. A. Cherniaev**  
Minimizing the number of input features in medical diagnostics ..... 102–109

**F. K. Ilyasov, A. K. Ayupova**  
Morphological analysis of the structures of dried droplets  
of a water-salt solution of serum albumin treated with cold plasma ..... 110–116

**INSTRUMENT ENGINEERING, MEASUREMENT SCIENCE,  
INFORMATION AND MEASURING DEVICES AND SYSTEMS**

**INSTRUMENTS AND METHODS OF MEASUREMENT  
(BY TYPE OF MEASUREMENT)**

**A. N. Brysin, D. R. Arsentiev, A. S. Novikova,  
Yu. A. Zhuravleva, O. E. Zheleznikova, O. Yu. Kovalenko**  
Development and software-hardware realization  
of a double 14-bit ADC ..... 117–127

**RULES FOR THE AUTHORS** ..... 128

# **ИНФОРМАТИКА, ВЫЧИСЛИТЕЛЬНАЯ ТЕХНИКА И УПРАВЛЕНИЕ**

## **МАТЕМАТИЧЕСКОЕ И ПРОГРАММНОЕ ОБЕСПЕЧЕНИЕ ВЫЧИСЛИТЕЛЬНЫХ МАШИН, КОМПЛЕКСОВ И КОМПЬЮТЕРНЫХ СЕТЕЙ**

УДК 303.732.4; 303.71

### **ПРИМЕНЕНИЕ КОРРЕЛЯЦИОННОГО АНАЛИЗА С ПРИМЕРАМИ НА ЯЗЫКАХ ПРОГРАММИРОВАНИЯ «PYTHON» И «R» ДЛЯ ИССЛЕДОВАНИЯ СТАТИСТИЧЕСКИХ ДАННЫХ ИЗ ФОРМ ОТЧЕТНОСТИ ПЕНИТЕНЦИАРНОЙ СИСТЕМЫ**

*Пономарёв Дмитрий Сергеевич*, Научно-исследовательский институт Федеральной службы исполнения наказаний, Российская Федерация, 125130, г. Москва, ул. Нарвская, 15а, строение 1, кандидат технических наук, ORCID: 0000-0003-1562-2956, e-mail: [ponomarev.dmitry1990@mail.ru](mailto:ponomarev.dmitry1990@mail.ru)

В приведенном научном исследовании освещается значимость практического применения корреляционного анализа данных в исследованиях и формировании статистической отчетности Федеральной службы исполнения наказаний. В работе рассмотрены правила и особенности использования коэффициентов корреляции Спирмена, Кендалла и Пирсона. Приводятся примеры применения тестов Шапиро – Уилка и Холмогорова – Смирнова. Представлены общепринятые подходы для предварительной подготовки данных перед проведением корреляционного анализа. Реализация представленных методов и подходов проводилась при помощи языков программирования «Python» и «R»: в работе приведены примеры расчетов коэффициентов корреляции, создания матриц коэффициентов корреляций и взаимных распределений, а также разработка тепловых карт. Рассмотрено практическое применение таких популярных библиотек, как `seaborn` и `psych` (для построения матриц взаимных распределений и тепловых карт), `corplot` (для расчетов непосредственно самих коэффициентов корреляций), `ggplot2` (для построения графиков нормального распределения), `scipy` (для проведения тестов Шапиро – Уилка и Холмогорова – Смирнова). Приведен обзор существующих интегрированных сред разработки и web-сервисов для проведения расчетов при помощи упомянутых языков программирования. В качестве практического примера в исследовании были проанализированы статистические данные трудового сектора пенитенциарной системы. В ходе проведенного корреляционного анализа были выявлены взаимосвязи между исследуемыми параметрами, также была проведена интерпретация полученных результатов. Полученные в ходе исследования выводы могут быть применимы для определения перспектив и преимуществ использования корреляционного анализа в служебной практике. Результаты исследования предоставляют основу для обсуждения возможных перспектив и преимуществ практического применения данных методов в работе.

**Ключевые слова:** корреляционный анализ, Python, R, системный анализ, пенитенциарная система

### **APPLICATION OF CORRELATION ANALYSIS WITH EXAMPLES IN PYTHON AND R TO STUDY STATISTICAL DATA FROM REPORTING FORMS OF THE FEDERAL PENALTY SERVICE**

*Ponomarev Dmitriy S.*, Federal State Institution Research Institute of the Federal Penitentiary Service of Russia, building 1, 15a Narvskaya St., Moscow, 125130, Russian Federation, Cand. Sci. (Engineering), ORCID: 0000-0003-1562-2956, e-mail: [ponomarev.dmitry.1990@mail.ru](mailto:ponomarev.dmitry.1990@mail.ru)

The above scientific research highlights the importance of the practical application of correlation data analysis in research and the formation of statistical reporting of the Federal Penitentiary Service. The paper discusses the rules and features of using Spearman, Kendall and Pearson correlation coefficients. Examples of application of the Shapiro-Wilk and Kholmogorov-Smirnov tests are given. Common approaches for preliminary data preparation before conducting correlation analysis are presented. The implementation of the presented methods and approaches was carried out using the Python and R programming languages: the work provides examples of calculations of correlation coefficients, creation of matrices of correlation coefficients and mutual distributions, as well as the development of heat maps. The practical application of such popular libraries as `seaborn` and `psych` (for constructing matrices of mutual distributions and heat maps), `corplot` (for calculating the correlation coefficients themselves), `ggplot2` (for constructing normal distribution graphs), `scipy` (for conducting Shapiro-Wilk and Kholmogorov tests) is considered (Smirnova). An overview of existing integrated development environments and web services for carrying out calculations using the mentioned programming languages is provided. As a practical example, the study analyzed statistical data from the labor sector of the penitentiary system. During the correlation analysis, relationships between the studied parameters were identified, and the results were also interpreted. The conclusions obtained during the study can be applied to determine the prospects and advantages of using correlation analysis in professional practice. The results of the study provide a basis for discussing possible prospects and advantages of the practical application of these methods in work.

**Keywords:** correlation analysis, Python, R, statistics, penitentiary system

## ВВЕДЕНИЕ

На сегодняшний день в уголовно-исполнительной системе активно проводятся исследования в области науки и инженерных решений, особенно актуальными являются разработки методов и моделей управления и поддержки принятия решений [1–5]. Осуществление сбора и учета информации, создание базы данных являются важной составляющей не только практически любого из перечисленных процессов, но и полноценного функционирования системы исполнения наказаний в целом. В соответствии с Приказом Федеральной службы исполнения наказаний Российской Федерации от 6 августа 2009 г. № 357, с 1 октября 2009 г. введена в эксплуатацию автоматизированная информационная система (далее – АИС). Данная АИС предоставляет возможность поддерживать общую базу данных, осуществлять передачу, прием и обработку статистической отчетности, а также создавать сводные, оперативные и регламентные аналитические отчеты. В результате использования данной системы в Федеральной службе исполнения наказаний была накоплена обширная база данных. Стоит отметить при этом, что не менее важными составляющими для накопленных данных являются и разработки методических рекомендаций для анализа статистической информации, прогнозирования, оценки их достоверности и поиска скрытых взаимосвязей между параметрами [6]. В этом контексте одним из ключевых элементов может быть использование методов системного и статистического анализов данных. В работе рассмотрены возможные преимущества и особенности применения корреляционного анализа для данных, сформированных на основе форм статистической отчетности Федеральной службы исполнения наказаний.

Таким образом, **целью** данной работы стало исследование преимуществ и возможностей применения корреляционного анализа для статистических данных из форм отчетности Федеральной службы исполнения наказаний.

**Методы исследования.** В работе использованы методы статистической обработки данных и системного анализа, такие как: нормализация значений, построение матриц распределений и гистограмм, проведение тестов Шапиро – Уилка и Холмогорова – Смирнова, построение «тепловых карт» основанных на матрицах коэффициентов корреляции Спирмена, а также были рассмотрены условия применения корреляционного анализа, основанного на расчете коэффициентов Пирсона и Кендалла. Реализация методов проводилась при помощи библиотек языка «R», также были рассмотрены и возможности применения библиотеки на языке «Python».

**Научными результатами** в ходе проведенной работы явились: разработанные правила для проведения корреляционного анализа для данных, сформированных на основе форм статистической отчетности Федеральной службы исполнения наказаний; особенности нормализации значений исследуемых параметров; преимущества проведения тестов Шапиро – Уилка и Холмогорова – Смирнова для исследуемых данных; а также непосредственно сами результаты проведенного корреляционного анализа данных.

## МЕТОДЫ КОРРЕЛЯЦИОННОГО АНАЛИЗА

Коэффициенты корреляции Пирсона, Кендалла и Спирмена – это три различных метода измерения степени связи между двумя переменными. Они различаются в том, как они учитывают и измеряют эту связь и в каких ситуациях они наиболее подходят [7].

Для проведения корреляционного анализа были рассмотрены следующие условия для каждого метода:

1. Для применения корреляции Пирсона были рассмотрены условия: измерение линейной связи между двумя непрерывными переменными; наличие нормального распределения данных; наличие измерения, насколько сильно две переменные меняются вместе в одном направлении; диапазон значений от  $-1$  до  $1$ , где  $-1$  обозначает сильную отрицательную линейную связь,  $1$  – сильную положительную связь и  $0$  – отсутствие линейной связи. Для двух параметров ( $X$ ,  $Y$ ) коэффициент корреляции ( $r_{xy}$ ) рассчитывался по формуле (1) [7]:

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{(n-1)s_x s_y}, \quad (1)$$

где  $n$  – объем выборки;  $x_i, y_i$  – значения параметров  $X, Y$ ;  $\bar{x}, \bar{y}$  – средние значения параметров  $X, Y$ ;  $s_x, s_y$  – среднеквадратичное отклонение (2):

$$s_x = \sqrt{\sum_{i=1}^n (x_i - \bar{x})^2}, \quad (2)$$

при этом (3):

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i. \quad (3)$$

2. Применение корреляции Кендалла можно охарактеризовать следующим образом: измеряет степень согласованности между двумя переменными; основана на сравнении пар значений переменных; является непараметрическим методом, не требует нормального распределения данных и устойчив к выбросам; измеряет, насколько сильно значения переменных согласованы при соблюдении порядка (ранжирования) данных; подходит для измерения степени согласованности между двумя

переменными, не предполагая линейности данных, и когда данные содержат ранги или категории. Для расчетов коэффициентов корреляции Кендалла применяется формула (4):

$$r = \frac{4P(p)}{n(n-1)} - 1, \quad (4)$$

где  $P$  – число совпадений.

3. Применение корреляции Спирмена: измеряет степень монотонной связи между двумя переменными; основана на ранжировании данных, т. е. замене данных их порядковыми номерами; непараметрический метод, не требует нормального распределения данных и устойчив к выбросам; измеряет, насколько сильно значения переменных согласованы при соблюдении порядка данных; подходит для измерения монотонных связей между переменными, когда данные содержат ранги или категории. Расчет корреляции Спирмена, как правило, проводится по формуле (5) [7]:

$$r = 1 - \frac{6 \sum_{i=1}^n d^2}{n(n^2-1)}, \quad (5)$$

где  $\sum d^2$  – сумма квадратов разностей рангов.

Далее следует рассмотреть методы реализации приведенных методов корреляционного анализа.

#### ИСПОЛЬЗОВАНИЕ БИБЛИОТЕК ЯЗЫКОВ «R» И «PYTHON»

На языке «R» проведение корреляционного анализа возможно при помощи встроенных библиотек, а также таких библиотек, как «ggplot2», «corrplot» и др. – для более удобной визуализации [8]. Среди наиболее часто применимых IDE («Integrated development environment» – интегрированная среда разработки) для работы на «R» достаточно популярна RStudio.

Пример построения матрицы корреляций на «R» представлен далее (где  $df$  – исследуемый массив данных):

```
corrplot(corr(df, method = "pearson"), number.cex = 1, number.digits = 1, diag = F, bg = "grey",
outline = "black", addgrid.col = "white")
```

При этом изменение коэффициентов корреляции Спирмена, Пирсона и Кендалла будет соответственно изменяться благодаря соответствующим командам “pearson”, “kendall”, “spearman” (например: `method = “pearson”`).

На языке «Python» проведение корреляционного анализа возможно в таких средах, как «Google Colaboratory», «Jupyter Notebook», «PyCharm» [9] и др. составить матрицу корреляций и представить в графическом виде (например, при помощи «тепловой карты»), можно используя библиотеки «pandas», «matplotlib», «seaborn» и др. Далее представлена разработка тепловой карты для матрицы коэффициентов корреляции при помощи библиотеки «seaborn»:

```
sns.heatmap(df.corr(method = "pearson"), xticklabels = df.corr().columns, yticklabels =
df.corr().columns, cmap = "RdYlGn", center = 0, annot = T)
```

Как и в предыдущем примере, изменение коэффициентов корреляции Пирсона, Кендалла и Спирмена будет соответствовать командам “pearson”, “kendall”, “spearman” [9].

#### ПРАКТИЧЕСКОЕ ПРИМЕНЕНИЕ

Рассмотрим применение корреляционного анализа с использованием библиотек языков «Python» и «R» для исследования статистических данных из форм отчетности Федеральной службы исполнения наказаний. Для этого были проанализированы такие сферы, как трудовая адаптация осужденных, получение профессионального образования, распространение социально значимых заболеваний. В результате было выделено несколько параметров, которые не только являются актуальными для современного общества, но и представляют особый интерес с научной точки зрения.

Расчет проводился на примере следующих параметров: «Среднесписочная численность осужденных, получающих профессиональное образование и профессиональное обучение в профессиональных образовательных организациях» (присвоена переменная X1); «Среднесписочная численность трудоустроенных осужденных» (присвоена переменная X2); «Среднесписочная численность лиц, больных наркоманией» (присвоена переменная X3); «Среднесписочная численность лиц, больных алкоголизмом» (присвоена переменная X4).

Учет данных параметров ведется согласно Приказу ФСИН России № 754 от 30.08.2019 «Об утверждении формы статистической отчетности ОТАО «Отчет о трудовой адаптации осужденных» и инструкции по ее заполнению и представлению» (далее Приказ ФСИН России № 754 от 30.08.2019) и Приказу ФСИН России № 1033 от 12.11.2019 «Об утверждении формы ведомственной статистической отчетности ФСИН-6 «Сведения о социально значимых заболеваниях у лиц, содержащихся в учреждениях уголовно-исполнительной системы Российской Федерации, и отдельных показателях деятельности медицинской службы» и инструкции по ее заполнению и представлению» (далее Приказ ФСИН России № 1033 от 12.11.2019). Выборка была сформирована на основе данных восьми федеральных округов Российской Федерации в период с 2019 по 2022 г., объем выборки составил  $n = 128$ .

**Предобработка данных.** В первую очередь следует рассмотреть предварительную обработку данных для проведения корреляционного анализа. Как правило, нормализация значений проводится по формуле [10] (6):

$$\dot{x}_n = \frac{(x-x_{min})}{(x_{max}-x_{min})} \quad (6)$$

Однако, согласно проведенным ранее исследованиям [11], для параметров X1–X4, предобработку данных более предпочтительно проводить по формуле (7):

$$\dot{x}_n = \frac{x_n}{Z} \quad (7)$$

где  $Z$  – параметр, который отражает среднесписочную численность осужденных. В результате значения исследуемых параметров будут нормализованы в пределах от 0 до 1.

Проверка на нормальность проводилась при помощи тестов Шапиро – Уилка (8) [12]:

$$W = \frac{1}{s^2} \sum_{i=1}^n a_{n-i+1} (x_{n-i+1} - x_i) s^2, \quad (8)$$

где значение коэффициента  $a_{n-i+1}$  является табличным, если  $W < W(\alpha)$ , то нулевая гипотеза о нормальности распределения данных отклоняется ( $W(\alpha)$  – табличное значение,  $\alpha$  – заданный уровень значимости).

Проведение теста Шапиро – Уилка и построение диаграммы распределения на примере X1 на языке «R» представлено далее (где  $df$  – сформированный набор данных, уровень значимости принят 0,05):

```
shapiro.test(df$X1)
hist(df$X4, col="coral2", xlab="distance", main="X1")
```

Проведение теста Шапиро – Уилка на языке «Python» также возможно: для этого можно использовать библиотеку «scipy.stats». Далее представлено проведение теста Шапиро – Уилка с использованием условных операторов «if», «else»:

```
statistic, p_value = stats.shapiro(X1)
print("Statistic:", statistic)
print("p-value:" p_value)
alpha = 0.05
if p_value > alpha:
    print("Гипотеза о нормальности не отвергается")
else:
    print("Гипотеза о нормальности отвергается")
```

В результате было получено, что практически для всех параметров (за исключением X1)  $p\text{-value} < 0,05$ , что говорит о том, что нулевая гипотеза о нормальности распределения данных отклоняется. Был также проведен тест Колмогорова – Смирнова, были получены аналогичные результаты (только теперь уже для всех исследуемых параметров  $p\text{-value} < 0,05$ , включая X1). Более наглядно распределение данных можно рассмотреть на примере построения гистограмм. Гистограммы распределений представлены на рисунке 1.

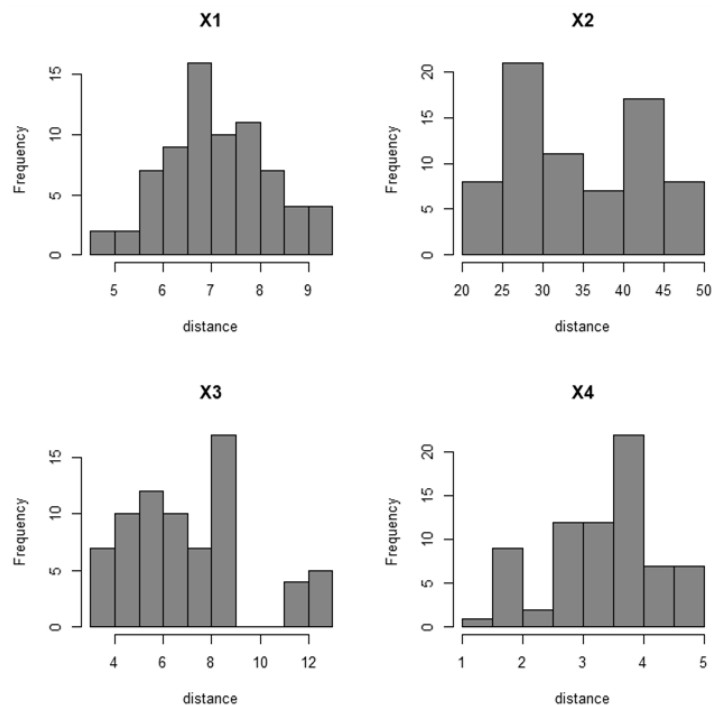


Рисунок 1 – Гистограммы распределений для нормализованных параметров X1–X4

Таким образом, применение корреляции Спирмена является наиболее подходящим решением для проведения корреляционного анализа между параметрами.

Стоит отметить при этом, что если привести распределение значений параметров выборки к нормальному распределению, то также представляется возможным и применение корреляции Пирсона. Однако данные вопросы требуют отдельного масштабного исследования, что выходит за рамки представленной статьи.

**Результаты корреляционного анализа.** Рассмотрим более подробно применение корреляции Спирмена для вышеприведенных данных. На рисунке 2 представлена матрица коэффициентов корреляции для исследуемых параметров в виде тепловой карты.

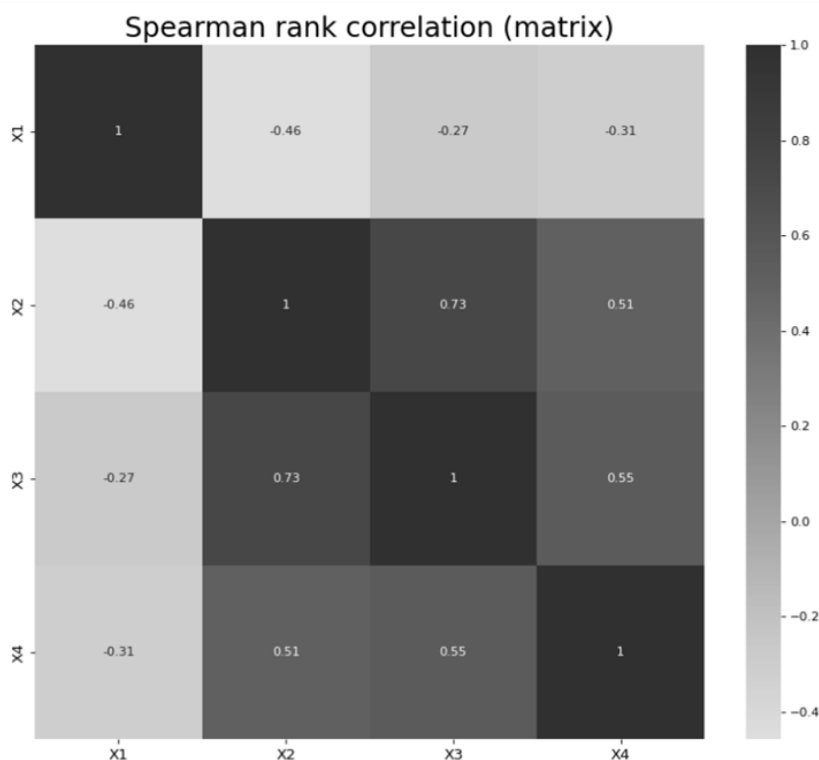


Рисунок 2 – Пример матрицы коэффициентов корреляции Спирмена в виде «тепловой карты» с использованием библиотеки «seaborn» на языке «Python»

При помощи метода «pairs.panels» из библиотеки «psych» на языке R можно совместить матрицы взаимных распределений (в графическом виде) с матрицей коэффициентов корреляции, при этом на диагонали будут размещены гистограммы распределений значений исследуемых параметров. Кроме того, в данном методе по умолчанию установлено построение центров распределения данных в графическом виде. Результаты применения данного метода представлены на рисунке 3 (для расчетов коэффициентов корреляции было выбрано «method = 'spearman'»), т. е. расчет коэффициентов корреляции Спирмена).

Как видно из данной матрицы, сильную прямую корреляцию ( $0,7 < r < 1,0$ ) имеет лишь одна пара параметров: X2 и X3 ( $r = 0,73$ ). Среднюю прямую корреляцию ( $0,5 < r < 0,7$ ) имеют пары параметров: X2 и X4 ( $r = 0,51$ ); X3 и X4 ( $r = 0,55$ ). Отсутствие корреляции можно наблюдать между параметрами: X1 и X2 ( $r = -0,46$ ); X1 и X3 ( $r = -0,27$ ); X1 и X4 ( $r = -0,31$ ).

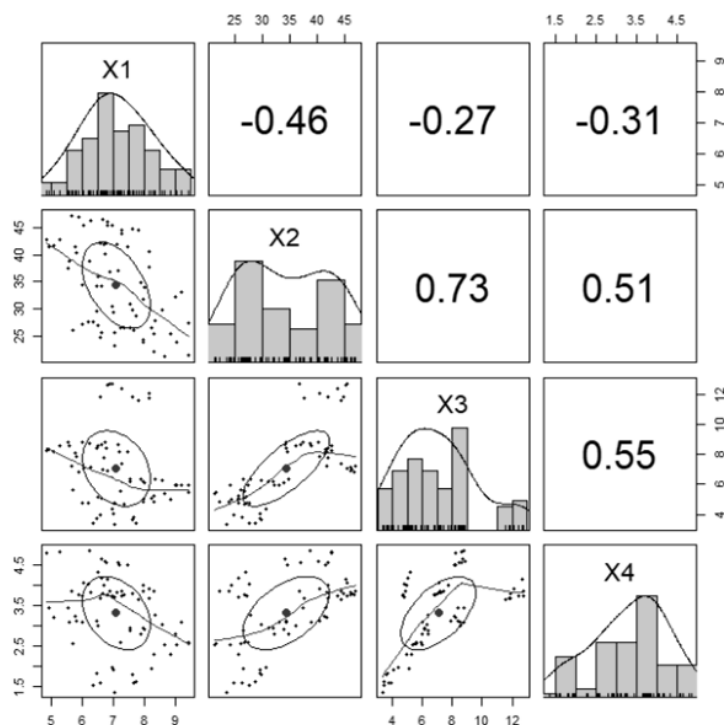


Рисунок 3 – Совмещенная матрица взаимных распределений и коэффициентов корреляции Спирмена для параметров X1–X4 с использованием метода «pairs.panels» из библиотеки «psych» на языке «R»

### ИНТЕРПРЕТАЦИЯ ПОЛУЧЕННЫХ РЕЗУЛЬТАТОВ

Из полученных результатов можно наблюдать корреляцию между параметрами «Среднесписочная численность трудоустроенных осужденных» (X2) и «Среднесписочная численность лиц, больных наркоманией» (X3) ( $r = 0,73$ ). Данный результат можно объяснить тем, что подразделениями ФСИН России уделяется должное внимание трудоустройству осужденных, которые имеют наркотическую зависимость. Аналогичная ситуация наблюдается между параметрами «Среднесписочная численность трудоустроенных осужденных» (X2) и «Среднесписочная численность лиц, больных алкоголизмом» (X4) (но в меньшей степени здесь  $r = 0,51$ ).

Взаимосвязь между параметрами «Среднесписочная численность лиц, больных наркоманией» (X3) и «Среднесписочная численность лиц, больных алкоголизмом» (X4) ( $r = 0,55$ ) свидетельствует о том, что социально значимые заболевания и вредные привычки, как правило, могут порождать одна другую.

Достаточно интересный результат получился для параметра «Среднесписочная численность осужденных, получающих профессиональное образование и профессиональное обучение в профессиональных образовательных организациях» (X1): здесь хоть и не наблюдается ярко выраженной обратной корреляции (т. е. случаи, когда  $-1,00 < r < -0,50$ ), но тем не менее все коэффициенты в матрице для данного параметра являются отрицательными. Другими словами, можно выдвинуть предположение о том, что параметр, который отражает стремление получить профессиональное образование имеет тенденцию к противоположному тренду относительно наличия социально значимых заболеваний и вредных привычек (таких как алкоголизм и наркозависимость). Наличие обратной корреляции с трудоустройством для данного параметра объясняется тем, что получение профессионального образования происходит с отрывом от производства и трудовой деятельности.

### ЗАКЛЮЧЕНИЕ

Грамотное использование корреляционного анализа для параметров может позволить как выявить эффективность работы подразделений ФСИН России в рассматриваемых вопросах, так и определить возможные недостатки в работе. Стоит отметить при этом, что, как и для любого метода статистической обработки, применение данного метода должно сопровождаться правильной интерпретацией результатов, а также рассмотрением возможных вероятностей получения ошибочных результатов (например, из-за неправильного учета данных, неправильной их обработки, наличия неучтенных факторов и дополнительных скрытых зависимостей между параметрами).

В качестве продолжения данного исследования можно рассмотреть применение корреляции Пирсона (для этого нужно привести значения параметров к нормальным распределениям), применение кластерного анализа и анализ «выбросов» в исследуемых данных. Также в качестве продолжения полученных результатов исследования можно рассматривать разработку регрессионных моделей на основе проведенного корреляционного анализа.

**Список источников**

1. Меньших, В. В. Методика расчета устойчивости функционирования модели управления при возникновении чрезвычайных обстоятельств на объекте уголовно-исполнительной системы / В. В. Меньших, А. В. Папонов // Вестник Воронежского института МВД России. – 2013. – № 3. – С. 71–77.
2. Кочедыков, С. С. Разработка методики комплексной оценки деятельности в уголовно-исполнительной системе / С. С. Кочедыков, Л. В. Россихина // Вестник Воронежского института МВД России. – 2013. – № 1. – С. 209–214.
3. Благодатский, Г. А. Системный анализ организационной структуры медицинской службы уголовно-исполнительной системы и управление ее реформированием / Г. А. Благодатский, М. М. Горохов, С. Б. Пономарев. – Ижевск : Изд-во ИжГТУ, 2017. – 104 с.
4. Пономарев, С. Б. К вопросу о применении информационных систем для оптимизации тактики ведения больных в местах лишения свободы / С. Б. Пономарев, М. М. Горохов, А. В. Серебrenников, С. Г. Логинова // Интеллектуальные системы в производстве. – 2007. – № 2. – С. 100–103.
5. Россихина, Л. В. Задача разработки календарных планов в уголовно-исполнительной системе по критерию упущенной выгоды / Л. В. Россихина // Вестник Воронежского института МВД России. – 2014. – № 1. – С. 286–291.
6. Пономарев, Д. С. Обработка форм статистической отчетности Федеральной службы исполнения наказаний на основе методов разведочного анализа данных библиотек языков «Python» и «R» / Д. С. Пономарев, М. М. Горохов, С. Б. Пономарев // Вестник Воронежского института ФСИН России. – 2023. – № 2. – С. 106–112.
7. Bruce, P. Practical statistics for Data Scientists / P. Bruce, A. Bruce, P. Gedeck. – O'Reilly, 2020. – P. 1–86.
8. Robert, I. Kabacoff. R in Action: Data analysis and graphics with R / Robert I. Kabacoff. – Manning Publications Co, 2013. – P. 27–98.
9. Samir, Madhavan. Mastering Python for Data Science / Samir Madhavan. – Packet Publishing, 2018. – 276 p.
10. Tukey, John W. The Future of Data Analysis / Tukey John W. // The Annals of Mathematical Statistics. – 1962. – № 1. – P. 1–67.
11. Пономарев, Д. С. Производственная и трудовая деятельность как фактор медико-социальной адаптации осужденных : отчет о НИР / Д. С. Пономарев, М. М. Горохов, С. Б. Пономарев, И. В. Разумова ; Федеральное казенное учреждение «Научно-исследовательский институт Федеральной службы исполнения наказаний». – 2022. – С. 5–40.
12. Shapiro, S. S. An analysis of variance test for normality (complete samples) / S. S. Shapiro, M. B. Wilk // Biometrika. – 1965. – № 52. – P. 591–611.

**References**

1. Menshikh, V. V., Paponov, A. V. Methodology for calculating the stability of the functioning of the management model in the event of emergency circumstances at the facility of the penal system. *Bulletin of the Voronezh Institute of the Ministry of Internal Affairs of Russia*, 2013, no. 3, pp. 71–77 (In Russ.).
2. Kochedykov, S. S., Rossikhina, L. V. Development of a methodology for comprehensive assessment of activities in the penal system. *Bulletin of the Voronezh Institute of the Ministry of Internal Affairs of Russia*, 2013, no. 1, pp. 209–214 (In Russ.).
3. Blagodatskiy, G. A., Gorohov, M. M., Ponomarev, S. B. *System analysis of the organizational structure of the medical service of the penal system and management of its reform*. Izhevsk: Izhevsk State Technical University Publ., 2017. 104 p. (In Russ.).
4. Ponomarev, S. B., Gorohov, M. M., Serebrennikov, A. V., Loginova S. G. On the issue of using information systems to optimize the tactics of managing patients in places of deprivation of liberty. *Intellectual Systems in Production*, 2007, no. 2, pp. 100–103 (In Russ.).
5. Rossikhina, L. V. The task of developing calendar plans in the penal system according to the criterion of lost profits. *Bulletin of the Voronezh Institute of the Ministry of Internal Affairs of Russia*, 2014, no. 1, pp. 286–291 (In Russ.).
6. Ponomarev, D. S., Gorokhov, M. M., Ponomarev, S. B. Processing of statistical reporting forms of the Federal Penitentiary Service based on methods of exploratory analysis of data from Python and R language libraries. *Bulletin of the Voronezh Institute of the Federal Penitentiary Service of Russia*, 2023, no. 2, pp. 106–112 (In Russ.).
7. Bruce, P., Bruce, A., Gedeck P. Practical statistics for Data Scientists. O'Reilly, 2020, pp. 1–86.
8. Robert, I. Kabacoff. *R in Action: Data analysis and graphics with R*. Manning Publications Co, 2013, pp. 27–98.
9. Samir, Madhavan. *Mastering Python for Data Science*. Packet Publishing, 2018. 276 p.
10. Tukey, John W. The Future of Data Analysis. *The Annals of Mathematical Statistics*, 1962, no. 1, pp. 1–67.
11. Ponomarev, D. S., Gorokhov, M. M., Ponomarev, S. B., Razumova, I. V. *Production and labor activity as a factor in the medical and social adaptation of convicts: research report*. Federal government institution "Research Institute of the Federal Penitentiary Service", 2022, pp. 5–40 (In Russ.).
12. Shapiro, S. S., Wilk, M. B. An analysis of variance test for normality (complete samples). *Biometrika*, 1965, no. 52, pp. 591–611.

Статья поступила в редакцию 03.05.2024; одобрена после рецензирования 27.05.2024; принята к публикации 27.05.2024.

The article was submitted 03.05.2024; approved after reviewing 27.05.2024; accepted for publication 27.05.2024.

УДК 004.001

### ИНФОРМАЦИОННАЯ СИСТЕМА ОБРАБОТКИ ДАННЫХ О ПАЦИЕНТАХ ДЛЯ ДИАГНОСТИКИ АНОМАЛЬНЫХ МАТОЧНЫХ КРОВОТЕЧЕНИЙ

**Кузнецова Валентина Юрьевна**, Астраханский государственный университет им. В. Н. Татищева, 414056, Российская Федерация, г. Астрахань, ул. Татищева, 20а,

кандидат технических наук, доцент кафедры информационной безопасности, ORCID 0000-0002-6954-5020, e-mail: arhelia@bk.ru

**Зоева Аделя Ренатовна**, Астраханский государственный медицинский университет, 414000, Российская Федерация, г. Астрахань, ул. Бакинская, 121,

кандидат медицинских наук, доцент кафедры акушерства и гинекологии педиатрического факультета с курсом последипломного образования, ORCID: 0000-0001-8596-6514, e-mail: abzhalilova@mail.ru

В статье рассматривается проблема диагностики аномальных маточных кровотечений в отечественной и зарубежной практике. Отмечено, что в рамках классического осмотра и опроса пациенток в условиях ограниченного времени медицинские работники могут получить неполную картину их состояния, тем самым допуская ложные выводы об отсутствии хронических кровотечений, которые кажутся нормальными для многих пациенток. Авторами предлагается авторский опросник, положенный в основу информационной системы обработки данных для диагностики аномальных маточных кровотечений. Информационная система представлена в виде веб-приложения, которое может быть использовано врачами акушерами-гинекологами для первичного сбора информации о пациентках, самими пациентками для выявления факторов риска для дальнейшего обследования, а также для сбора статистической информации по теме аномальных маточных кровотечений.

**Ключевые слова:** цифровые технологии в медицине, аномальные маточные кровотечения, обработка данных пациентов, система поддержки принятия решений

### INFORMATION SYSTEM FOR PROCESSING PATIENT DATA FOR DIAGNOSIS OF ABNORMAL UTERINE BLEEDING

**Kuznetsova Valentina Yu.**, Astrakhan Tatishchev State University, 20a Tatishchev St., Astrakhan, 414056, Russian Federation,

Cand. Sci. (Engineering), Associate Professor, Department of Digital Technologies, and Cybersecurity, ORCID: <https://orcid.org/0000-0002-6954-5020>, e-mail: arhelia@bk.ru

**Zoeva Adelya R.**, Astrakhan State Medical University, 121 Bakinskaya St., Astrakhan, 414000, Russian Federation,

Cand. Sci. (Medicine), Associate Professor, Department of Obstetrics and Gynecology, Faculty of Pediatrics with a postgraduate education course, ORCID: <https://orcid.org/0000-0001-8596-6514>, e-mail: abzhalilova@mail.ru

The article discusses the problem of diagnosing abnormal uterine bleeding in domestic and foreign practice. It is noted that during a classic examination and interview of patients under limited time conditions, medical workers may not obtain a complete picture of their condition, thereby allowing false conclusions about the absence of chronic bleeding, which seem normal for many patients. The authors propose an original questionnaire that serves as the basis for an information processing system for diagnosing abnormal uterine bleeding. The information system is presented in the form of a web application that can be used by obstetricians-gynecologists for the initial collection of information about patients, by the patients themselves to identify risk factors for further examination, as well as to collect statistical information on the topic of abnormal uterine bleeding.

**Keywords:** digital technologies in medicine, abnormal uterine bleeding, patient data processing, decision support system

#### Graphical annotation (Графическая аннотация)



## **ВВЕДЕНИЕ**

На сегодняшний день проблема нарушений менструального цикла в практике врачей акушеров-гинекологов как отечественных, так и зарубежных остается весьма актуальной [1–3]. Аномальные маточные кровотечения (АМК) занимают ведущую позицию в их структуре [4]. В соответствии с определением АМК у женщин детородного возраста – это кровотечение из матки, которое не соответствует нормальной картине менструальных циклов. А именно оно возникает слишком часто или нерегулярно либо длится дольше или обильнее, чем нормальные менструации.

Согласно положениям экспертной группы Международной классификации акушерства и гинекологии [3, 5, 6], АМК могут быть проявлениями различных анатомических (полипы эндометрия и цервикального канала, аденомиоз, лейомиома матки, гиперплазия и рак эндометрия) и функциональных нарушений (коагулопатии, овуляторные нарушения [6], эндометриальная дисфункция, ятрогенные и неклассифицируемые причины) женской репродуктивной системы.

Казалось бы, что диагностика АМК не должна вызывать затруднений, ведь описаны четкие критерии для постановки этого диагноза. Но зачастую пациентки неадекватно оценивают объем теряемой крови при МЦ, не обращая внимания на наличие сгустков в менструальных выделениях, их длительность и нерегулярность, особенно если эти симптомы начали проявляться с пубертатного возраста.

В связи с ограниченным временем, отведенным на прием пациента, опрос ограничивается оценкой информации о возрасте менархе, длительности менструального цикла и менструального кровотечения. В обязательном порядке врач задает вопрос об обильности менструальной кровопотери, но восприятие объема теряемой менструальной крови у всех пациенток субъективно. На прием может прийти пациентка с патологией гемостаза [7], у которой с пубертатного возраста менструации длятся по 7 дней, они регулярны, но в них присутствуют сгустки, и больно приходится использовать двойную защиту от «протеканий» (гигиеническая прокладка и тампон одновременно) [8]. Но, учитывая давность процесса, женщина скажет, что менструации необильные, только потому что они у нее были такими всегда, и подобная ситуация для нее является нормой. А если доктор не задаст ей дополнительных вопросов, например, о способах защиты от «протеканий», объемом прокладок, которыми пользуется пациентка, то может сложиться ложное впечатление об отсутствии хронических АМК.

В связи с этим возникает необходимость разработки автоматизированного опросника для выявления возможного наличия аномальных маточных кровотечений. Опросник может быть использован врачами акушерами-гинекологами для первичного сбора информации о пациентках еще до приема (пациент самостоятельно проходит тест и приходит на приём уже с результатами для обсуждения с врачом), самими пациентками для выявления факторов риска, что сподвигнет их на прохождение тщательного обследования, а также для сбора обезличенной информации по теме аномальных маточных кровотечений в научных и статистических целях.

## **ОБЗОР ЛИТЕРАТУРНЫХ ИСТОЧНИКОВ И АНАЛОГОВ**

Согласно множеству литературных источников, у женщин с АМК чаще встречаются нарушения репродуктивной функции: гиперпластические процессы эндометрия, бесплодие, невынашивание беременности и преэклампсия [8]. Поэтому так важно на раннем этапе возникновения диагностировать АМК, выявить их причину и провести коррекцию заболевания.

Критерии постановки диагноза АМК четко прописаны в клиническом протоколе «Аномальные маточные кровотечения» [9], но задача акушера-гинеколога – грамотно задать вопросы пациентке, которые укажут на наличие патологии. В последние годы появляются исследования, направленные на выделение групп риска по развитию АМК. Например, в работе А. К. Красильниковой на основании изучения клинико-anamnestических данных у женщин репродуктивного возраста с АМК, обусловленными гиперплазией эндометрия, предлагается установить факторы риска развития данной патологии [10]. Кроме того, АМК рассматриваются с точки зрения психоэмоционального стресса, и Ю. А. Петровым предлагается оценить уровень эмоционального напряжения у женщин с кровотечениями [11]. Авторы одной из работ предпринимали попытки установить взаимосвязь между АМК, ожирением и полиморфизмом генов [14].

Часто аномальные маточные кровотечения являются одним из симптомов заболеваний или результатом побочного действия препаратов. Например, применение оральных контрацептивов может вызывать АМК [9].

Зачастую обильные менструации в пубертатном периоде являются манифестацией патологии гемостаза. Стоит отметить, что своевременно поставленный диагноз даст возможность не только корректировать объем теряемой крови, но и предотвратить развитие любых кровотечений в дальнейшем. Основными этиологическими звеньями в формировании патологии системы свертывания являются повреждение сосудистой стенки, тромбоцитов, дефицит факторов свертывания крови. Характерными признаками данной патологии является появление обильных менструаций с менархе при сохранении регулярности и цикличности менструального цикла. Для таких пациенток характерны проявления геморрагического синдрома [7].

Не стоит забывать и о ятрогенных причинных развития АМК. Поэтому так важно уточнить у пациенток с обильными менструациями препараты, которые они принимают. Так эстрогены, прогестины, андрогены, нестероидные противовоспалительные препараты, антикоагулянты, лекарственные средства, влияющие на метаболизм дофамина, включая фенотиазины и трициклические антидепрессанты, способны влиять на формирование физиологического менструального цикла и, как следствие, приводить к маточным кровотечениям [1, 7].

Существуют и работы, направленные на оптимизацию опроса пациенток с помощью IT-технологий, где пациентке предлагается пройти опрос с целью оценки характера ее менструаций [12, 13], позволяющий проводить оценку объема менструальной кровопотери, прогнозировать риски развития железодефицитного состояния. Но на сегодняшний день не существует удобного электронного анкетирования пациенток, которое бы было направлено не только на оценку менструальных кровотечений и, возможно, наличия симптомов анемии, но способного по результатам тестирования определить возможную причину АМК, с формированием перечня обследований, рекомендованного для женщины.

В связи с этим видится целесообразным разработка программного продукта для дистанционной оценки симптомов и признаков у пациентки, которые могли бы указывать на наличие у нее АМК. Тем самым, врач на приеме, без затраты времени, сможет выделить группу риска по развитию данной патологии, после чего назначить дообследование для уточнения причины их возникновения.

#### РАЗРАБОТКА ОПРОСНИКА И ЕГО ПРОГРАММНАЯ РЕАЛИЗАЦИЯ

Тестирование на выявление аномальных маточных кровотечений состоит из трех информационных блоков. Первый блок – основной. В нём собирается первичная информация, которая позволяет однозначно установить наличие аномального маточного кровотечения в соответствии с протоколом «Аномальные маточные кровотечения». В первом блоке вопросов пациентка указывает свои данные о возрасте, росте и весе. Рост и вес необходимы в дальнейшем для расчета индекса массы тела, который сигнализирует о наличии ожирения или избыточной массы тела, которые, в свою очередь, также могут влиять на развитие аномальных маточных кровотечений (рис. 1).

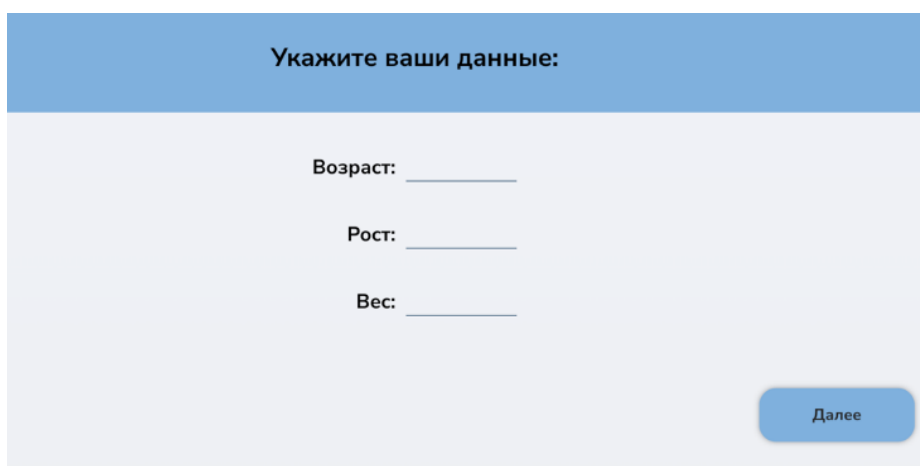


Рисунок 1 – Первый экран интерфейса опросника

Далее первый блок включает в себя вопросы о частоте, регулярности, длительности менструального цикла, а также о наличии кровянистых выделений в середине цикла. В каждом вопросе есть вариант ответа, за который начисляется 1 балл. Данный ответ является симптомом АМК. Если в рамках первого блока вопросов пациентка набрала хотя бы 1 балл, то наличие АМК для нее имеет высокую вероятность.

Второй блок вопросов включает в себя дополнительную информацию, которая косвенно может свидетельствовать о наличии аномальных маточных кровотечений. Этот блок вопросов особенно важен в том случае, если в первом блоке пациентка набрала 0 баллов, так как не имеет очевидных симптомов АМК. Данные вопросы связаны с качеством жизни, а именно – частота смены средств гигиены, в том числе в ночной период времени, их влагоемкость, необходимость подбора темной одежды в период менструаций. При этом вес ответов на вопросы различается: в вопросе о необходимости смены гигиенической прокладки в ночное время суток начисляется 0,5 баллов, если пациентка совершает данную процедуру 1 раз за ночь, а если 2 и более раз, то начисляется 1 балл. Аналогично в вопросах касательно степени защиты гигиенических средств и скорости наполнения прокладок. Если по итогам второго блока вопросов пациентка суммарно набрала 1 балл и более, вероятность АМК достаточно высокая.

4. Есть ли межменструальные кровянистые выделения?

Нет

Мажущие кровянистые выделения перед началом менструального цикла 2 и более дней

Беспорядочные менструации

Межменструальные кровянистые выделения

Рисунок 2 – Один из вопросов первого блока опросника

6. Пользуетесь ли Вы двойной защитой от «протеканий» - тампон+прокладка?

Да

Нет

Рисунок 3 – Один из вопросов второго блока опросника

На втором этапе тестирование завершается для тех пациенток, которые набрали менее 1 балла по результатам обоих блоков тестирования (рис. 4).

По отмеченным вами симптомам вероятность наличия аномальных маточных кровотечений минимальна. Если у вас всё-таки есть подозрения на данное нарушение, рекомендуем вам обратиться к акушер-гинекологу.

Рисунок 4 – Результат тестирования для пациенток с низким уровнем риска

Третий блок тестирования предназначен для формирования плана обследования пациентки с высоким риском АМК. Данные вопросы включают в себя уточнения касательно приема комбинированных оральных контрацептивов, заместительной гормонотерапии, факторов риска в анамнезе и нюансы, связанные с наследственностью. Помимо стандартных вопросов, задаваемых на приеме у врача акушера гинеколога (возраст менархе, характер менструального цикла), тестирование было дополнено пунктами, которые направлены на оценку объема кровопотери, путем детального анализа наполняемости, частоты замены и уровня защиты использования гигиенических средств во время менструаций. В соответствии с классификацией PALM-COEIN, в тестирование вошли вопросы, нацеленные на выявление неорганических причин возникновения АМК – о признаках, соответствующих геморрагическому синдрому (рис. 5).

**14. Есть ли у Вас нижеперечисленные факторы?**

Обильные менструации с момента, как они пришли впервые

Послеродовые кровотечения

Кровотечение во время операций

Кровотечения во время лечения у стоматолога

Появление синяков на теле 1-2 и более раз в месяц

Носовое кровотечения 1-2 и более раз в месяц

Кровоточивость десен при чистке зубов

Заболевания свертывающей системы крови в семье

**12. Принимаете ли Вы нижеуказанные препараты?**

Препараты, содержащие эстрогены и гестагены (комбинированные оральные контрацептивы, заместительная гормонотерапия, ВМИ с левоноргестрелом- Кайлина, Мирена)

Нестероидный противовоспалительные препараты

Антидепрессанты

Антикоагулянты

Тамоксифен

Нет

Рисунок 5 – Блок дополнительных вопросов для выявления возможных причин высокого риска кровотечений

На основании третьего блока вопросов формируется тот перечень анализов, которые, возможно, назначит лечащий акушер-гинеколог, так как на основании опроса не назначаются обследования, опрос носит исключительно информативный характер. К возможным анализам относятся: ХГЧ крови, общий анализ крови, коагулограмма, ферритин, микроскопическое исследование отделяемого из цервикального канала и заднего свода влагалища и УЗИ органов малого таза. При наличии дополнительных факторов формируется дополнительный список обследований: при высоком ИМТ и наследственности предлагается биопсия эндометрия под контролем гистероскопии, при частных кровотечениях – консультация гематолога, в случае, если пациентка принимает препараты, которые могут влиять на свертываемость крови, – указание на смену препарата при строгой консультации с лечащим врачом.

Рекомендуем вам обратиться за консультацией акушер-гинеколога в связи с подозрением на аномальные маточные кровотечения.

Возможно, акушер-гинеколог назначит вам следующий перечень анализов:

- ХГЧ крови
- Общий анализ крови
- Коагулограмма
- Ферритин
- Микроскопическое исследование отделяемого из цервикального канала и заднего свода влагалища
- УЗИ органов малого таза
- Биопсия эндометрия под контролем гистероскопии

Также вам рекомендована консультация гематолога

Возможно, что аномальные маточные кровотечения вызваны приемом Тамоксифен. Рекомендована консультация специалиста, назначившего лекарственное средство, для его замены или отмены.

Рисунок 6 – Пример рекомендаций по итогам тестирования пациентки

По завершении тестирования пациентка может выгрузить результаты своего тестирования в формате PDF, чтобы представить в распечатанном или электронном виде лечащему врачу для дальнейшего выбора тактики лечения. Блок-схема алгоритма работы программного продукта приведена на рисунке 7.

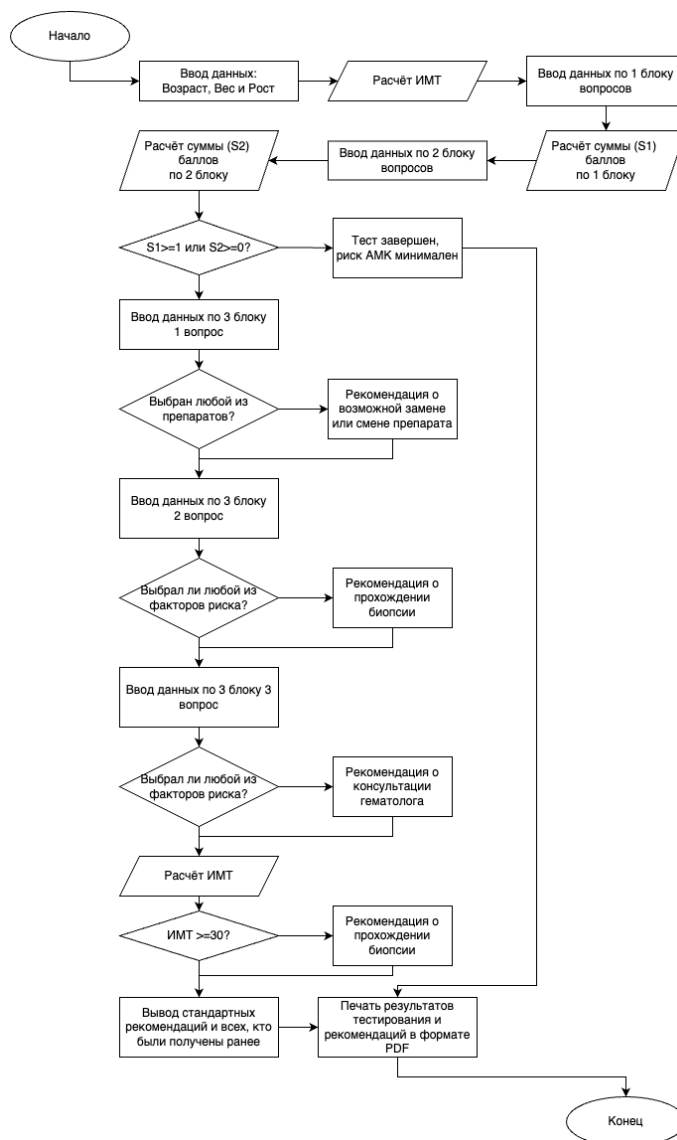


Рисунок 7 – Алгоритм работы программного продукта

Также в базу данных сервиса сохраняется результат тестирования в обезличенном формате для проведения дальнейших статистических исследований врачами-ординаторами.

### ЗАКЛЮЧЕНИЕ

В статье рассматривается проблема диагностики аномальных маточных кровотечений в отечественной и зарубежной практике. Отмечено, что в рамках классического осмотра и опроса пациенток в условиях ограниченного времени медицинские работники могут получить неполную картину их состояния, тем самым допуская ложные выводы об отсутствии хронического кровотечения, которые кажутся нормальными для многих пациенток. Авторами предлагается авторский опросник, положенный в основу информационной системы обработки данных для диагностики аномальных маточных кровотечений. Информационная система представлена в виде веб-приложения, которое может быть использовано врачами акушерами-гинекологами для первичного сбора информации о пациентках, самими пациентками для выявления факторов риска для дальнейшего обследования, а также для сбора статистической информации по теме аномальных маточных кровотечений.

### Список источников

1. Aldeewan, M. A. Histopathological Evaluation of Biopsy Material of the Endometrium of Premenopausal and Postmenopausal Women with Abnormal Uterine Bleeding / M. A. Aldeewan, O.F. Hameed, Z. A. Maktoof // *Laboratory Diagnostics. Eastern Europe*. – 2023. – Vol. 12, № 3. – P. 408–413. – DOI: 10.34883/PI.2023.12.3.009.
2. Мишина, А. И. Аномальные маточные кровотечения у девочек пубертатного периода, тактика ведения / А. И. Мишина, А. В. Казакова, О. И. Линева [и др.] // *American Scientific Journal*. – 2021. – № 49–1. – С. 28–31. – DOI 10.31618/asj.2707-9864.2021.1.49.104.
3. Енькова, Е. В. Аномальные маточные кровотечения: этиология и патогенез (описательный обзор) / Е. В. Енькова, Е. В. Киселева, О. В. Хоперская [и др.] // *Научные результаты биомедицинских исследований*. – 2022. – Т. 8, № 3. – С. 365–381. – DOI: 10.18413/2658-6533-2022-8-3-0-8.
4. Киселева, Н. И. Аномальные маточные кровотечения вследствие овulatoryной дисфункции / Н. И. Киселева, Н. П. Жукова, И. М. Арестова [и др.] // *Охрана материнства и детства*. – 2022. – № 1 (39). – С. 70–78.
5. Доброхотова, Ю. Э. Аномальные маточные кровотечения: алгоритмы ведения, методы терапии / Ю. Э. Доброхотова, Л. В. Сапрыкина, Л. А. Филатова, М. Р. Нариманова // *РМЖ. Мать и дитя*. – 2020. – Т. 3, № 1. – С. 55–60. – DOI: 10.32364/2618-8430-2020-3-1-55-60.
6. Горбачева, М. И. Аналитический обзор систем поддержки принятия решения в составе медицинских информационных систем акушерства и гинекологии / М. И. Горбачева // *Молодежь и системная модернизация страны*. – 2022. – С. 18–20.
7. Клинические рекомендации – Аномальные маточные кровотечения – 2021–2022–2023 (01.07.2021). – Утверждены Минздравом РФ.
8. Крылов, К. Ю. Использование тромбоэластографии в диагностике нарушений гемостаза при аномальных маточных кровотечениях / К. Ю. Крылов, В. С. Афончиков, А. Ю. Каськов // *Global Reproduction*. – 2021. – № S1. – С. 68–72.
9. Мынко, О. И. Комбинированные оральные контрацептивы: частота применения и возникновения побочных эффектов у женщин репродуктивного возраста / О. И. Мынко, Д. Р. Ашрафзянова, В. В. Лобанова // *Энигма*. – 2020. – № 26. – С. 105–114. – EDN LILSOE.
10. Николаева, А. Е. Аномальные маточные кровотечения как манифестация патологии системы гемостаза / А. Е. Николаева, В. Ф. Беженарь, Ф. Р. Кутуева // *Вопросы гинекологии, акушерства и перинатологии*. – 2022. – Т. 21, № 4. – С. 159–162. – DOI: 10.20953/1726-1678-2022-4-159-162.
11. Обернихин, К. И. Оптимизация тактики ведения пациенток с аномальными маточными кровотечениями в условиях цифровой трансформации здравоохранения / К. И. Обернихин, В. С. Иванов // *Молодежный инновационный вестник*. – 2023. – Т. 12, № S2. – С. 36–37.
12. Петров, Ю. А. Оценка психоэмоционального состояния женщин с хроническими аномальными маточными кровотечениями / Ю. А. Петров, Н. В. Палиева, А. Ю. Тарасова, С. С. Старков // *Главный врач Юга России*. – 2023. – № 4 (90). – С. 34–37.
13. Кулигин, А. В. Патолофизиологические основы аномальных маточных кровотечений / А. В. Кулигин, Е. В. Астафьева, А. И. Самсонова, Т. Д. Кочадаева // *Клиническая патофизиология*. – 2021. – Т. 27, № 4. – С. 27–34. – EDN UVMDMO.
14. Сибирская, Е. В. Современная классификация и методы коррекции аномальных маточных кровотечений у женщин позднего репродуктивного возраста / Е. В. Сибирская, Л. Г. Пивазян, И. Е. Колтунов, Ю. В. Выхристюк // *Трудный пациент*. – 2021. – Т. 19, № 4. – С. 18–24. – DOI: 10.224412/2074-1005-2021-4-18-24.
15. Свидетельство о государственной регистрации программы для ЭВМ № 2022619335 Российская Федерация. Оценка и прогнозирование течения аномальных маточных кровотечений : № 2022618598 : заявл. 13.05.2022 : опубл. 20.05.2022 / Е. В. Енькова, К. И. Обернихин, В. В. Енькова [и др.] ; заявитель Федеральное государственное бюджетное образовательное учреждение высшего образования «Воронежский государственный медицинский университет имени Н.Н. Бурденко» Министерства здравоохранения Российской Федерации. – EDN UCUMDL.
16. Факторы риска развития гиперплазии эндометрия с аномальными маточными кровотечениями у женщин репродуктивного возраста / А. К. Красильникова, Д.А.Мальшикина, С. М. Гасанова [и др.] // *Российский вестник акушера-гинеколога*. – 2023. – Т. 23, № 6. – С. 48–54. – DOI: 10.17116/rosakush 20232306148.

17. Хамошина, М. Б. Аномальные маточные кровотечения у пациенток с ожирением и полиморфизм генов eNOS, VEGF-A и HIF-1A / М. Б. Хамошина, Ю. С. Артеменко, М. М. Азова, Л. В. Цховребова, А. В. Агаджанян, С. А. Петровская // Акушерство и гинекология: Новости. Мнения. Обучения. – 2023. – № 3 (41). – URL: <https://cyberleninka.ru/article/n/anomalnye-matochnye-krovotecheniya-u-patsientok-s-ozhireniem-i-polimorfizm-genov-enos-vegf-a-i-hif-1a> (дата обращения: 20.03.2024).

#### References

1. Aldeewan, M. A., Hameed, O. F., Maktoof, Z. A. Histopathological Evaluation of Biopsy Material of the Endometrium of Premenopausal and Postmenopausal Women with Abnormal Uterine Bleeding. *Laboratory Diagnostics. Eastern Europe*, 2023, vol. 12, no. 3, pp. 408–413. – DOI: 10.34883/PI.2023.12.3.009 (In Russ.).
2. Mishina, A. I., Kazakova, A. V., Lineva, O. I. et al. Abnormal uterine bleeding in girls during puberty, management tactics. *American Scientific Journal*, 2021, no. 49-1. – P. 28-31. – DOI: 10.31618/asj.2707-9864.2021.1.49.104 (In Russ.).
3. Enkova, E. V., Kiseleva, E. V., Khoperskaya, O. V. et al. Abnormal uterine bleeding: etiology and pathogenesis (descriptive review). *Scientific results of biomedical research*, 2022, vol. 8, no. 3, pp. 365–381. DOI: 10.18413/2658-6533-2022-8-3-0-8 (In Russ.).
4. Kiseleva, N. I., Zhukova, N. P., Arestova, I. M. et al. Abnormal uterine bleeding due to ovulatory dysfunction. *Protection of Motherhood and Childhood*, 2022, no. 1 (39), pp. 70–78 (In Russ.).
5. Dobrokhotova, Yu. E., Saprykina, L. V., Filatova, L. A., Narimanov, M. R. Abnormal uterine bleeding: management algorithms, methods of therapy. *Breast Cancer. Mother and Child*, 2020, vol. 3, no. 1, pp. 55–60. DOI: 10.32364/2618-8430-2020-3-1-55-60 (In Russ.).
6. Gorbachev, M. I. Analytical review of decision support systems as composition of medical information systems of obstetrics and gynecology. *Youth and Systemic Modernization of the Country*, 2022, pp. 18–20 (In Russ.).
7. *Clinical guidelines – Abnormal uterine bleeding – 2021–2022–2023 (07.01.2021)*. Approved by the Ministry of Health of the Russian Federation (In Russ.).
8. Krylov, K. Yu., Afonchikov, V. S., Kaskov, A. Yu. The use of thromboelastography in the diagnosis of hemostasis disorders in abnormal uterine bleeding. *Global Reproduction*, 2021, no. S1, pp. 68–72 (In Russ.).
9. Mynko, O. I., Ashrafzyanova, D. R., Lobanova, V. V. Combined oral contraceptives: frequency of use and occurrence of side effects in women of reproductive age. *Enigma*, 2020, no. 26, pp. 105–114. EDN LILSOE (In Russ.).
10. Nikolaeva, A. E., Bezhenar, V. F., Kutueva, F. R. Abnormal uterine bleeding as a manifestation of the pathology of the hemostatic system. *Issues of Gynecology, Obstetrics and Perinatology*, 2022, vol. 21, no. 4, pp. 159–162. DOI: 10.20953/1726-1678-2022-4-159-162 (In Russ.).
11. Obernikhin, K. I., Ivanov, V. S. Optimization of tactics for managing patients with abnormal uterine bleeding in the context of digital transformation of healthcare. *Youth Innovative Bulletin*, 2023, vol. 12, no. S2, pp. 36–37 (In Russ.).
12. Petrov, Yu. A., Palieva, N. V., Tarasova, A. Yu. Starkov, S. S. Assessment of the psycho-emotional state of women with chronic abnormal uterine bleeding. *Chief Physician of the South of Russia*, 2023, no. 4 (90), pp. 34–37.
13. Kuligin, A. V. E. V. Astafieva, A. I. Samsonova, T. D. Kochedaeva Pathophysiological basis of abnormal uterine bleeding. *Clinical Pathophysiology*, 2021, vol. 27, no. 4, pp. 27–34. EDN UVMDMO (In Russ.).
14. Sibirskaia, E. V., Pivazyan, L. G., Koltunov, I. E., Vykhristyuk, Yu. V. Modern classification and methods of correction of abnormal uterine bleeding in women of late reproductive age. *Difficult Patient*, 2021, vol. 19, no. 4, pp. 18–24. DOI: 10.224412/2074-1005-2021-4-18-24 (In Russ.).
15. *Certificate of state registration of a computer program No. 2022619335 Russian Federation. Assessment and prediction of the course of abnormal uterine bleeding: no. 2022618598 : application. 05.13.2022 : publ. 05.20.2022*. E. V. Enkova, K. I. Obernikhin, V. V. Enkova et al. ; applicant Federal state budgetary educational institution of higher education “Voronezh State Medical University named after N.N. Burdenko of the Ministry of Health of the Russian Federation. EDN UCUMDL (In Russ.).
16. Krasilnikova, A. K., Malyshkina, D. A., Gasanova, S. M. Risk factors for the development of endometrial hyperplasia with abnormal uterine bleeding in women of reproductive age. *Russian Bulletin of Obstetrician-Gynecologist*, 2023, vol. 23, no. 6, pp. 48–54. DOI: 10.17116/rosakush20232306148 (In Russ.).
17. Khamoshina, M. B., Artemenko, Yu. S., Azova, M. M., Tskhovrebova, L. V., Agadzhanian, A. V., Petrovskaya, S. A. Abnormal uterine bleeding in obese patients and polymorphism of the eNOS, VEGF-A and HIF-1A genes. *Obstetrics and Gynecology: News. Opinions. Training*, 2023, no. 3 (41). URL: <https://cyberleninka.ru/article/n/anomalnye-matochnye-krovotecheniya-u-patsientok-s-ozhireniem-i-polimorfizm-genov-enos-vegf-a-i-hif-1a> (accessed 03.20.2024) (In Russ.).

Статья поступила в редакцию 15.06.2024; одобрена после рецензирования 01.07.2024; принята к публикации 04.09.2024.

The article was submitted 15.06.2024; approved after reviewing 01.07.2024; accepted for publication 04.09.2024.

## **СИСТЕМНЫЙ АНАЛИЗ, УПРАВЛЕНИЕ И ОБРАБОТКА ИНФОРМАЦИИ**

УДК 004.82, 007.5

### **СТРУКТУРНЫЙ ПОИСК СКРЫТЫХ СВЯЗЕЙ В ОНТОЛОГИЧЕСКОЙ БАЗЕ ЗНАНИЙ ДЕФЕКТНЫХ СОСТОЯНИЙ ЭЛЕКТРООБОРУДОВАНИЯ**

*Петрова Светлана Юрьевна*, Севастопольский государственный университет, 299053, Российская Федерация, г. Севастополь, ул. Университетская, 33, кандидат технических наук, доцент, ORCID: 0000-0002-1081-2833, e-mail: master-cim@yandex.ru

Старое электросетевое оборудование часто становится источником аварий и сбоев, что приводит к простоям, потерям электроэнергии и дополнительным расходам на ремонт. При этом замена всего устаревшего оборудования требует значительных финансовых и временных затрат. Одним из развивающихся направлений в неразрушающей диагностике становится использование искусственного интеллекта для выявления, верификации и ранжирования дефектов на электросетевых объектах. Общее научное исследование посвящено разработке моделей кратковременного машинного обучения для прогнозирования дефектов и предаварийного состояния электросетевого оборудования при ограниченном наборе исходных данных. Источником информации для обучения и прогнозирования будет служить онтологическая база знаний дефектных состояний электрооборудования. В работе дается формальное представление онтологической базы знаний, представлен обзор научных исследований и предлагаемых решений в области обработки связанных между собой данных (графы). Статья представляет собой постановку фундаментальной задачи – восстановления (или прогнозирования) отсутствующих связей в графе для идентификации дефектов и прогнозирования аварийного состояния на электросетевых объектах. Дано формальное описание задачи структурного поиска скрытых связей в онтологической базе знаний. Выдвинуто предположение о возможности представления базовой структуры граф с помощью многослойного перцептрона, представленного в виде направленного ациклического графа, в котором узлы соответствуют нейронам в нейронной сети, а грани – связям между нейронами.

**Ключевые слова:** инженерия знаний, системный анализ, онтологическая база знаний, графовые нейронные сети, неразрушающий контроль оборудования

**Финансирование:** исследование выполнено при финансовой поддержке РФФИ в рамках научного проекта № 24-21-20021.

### **STRUCTURAL SEARCH OF HIDDEN RELATIONS IN ONTOLOGICAL KNOWLEDGE BASE OF ELECTRICAL EQUIPMENT FAULTY STATES**

*Petrova Svetlana Yu.*, Sevastopol State University, 33 Universitetskaya St., Sevastopol, 299053, Russian Federation,

Cand. Sci. (Engineering), Associate Professor, ORCID: 0000-0002-1081-2833, e-mail: master-cim@yandex.ru

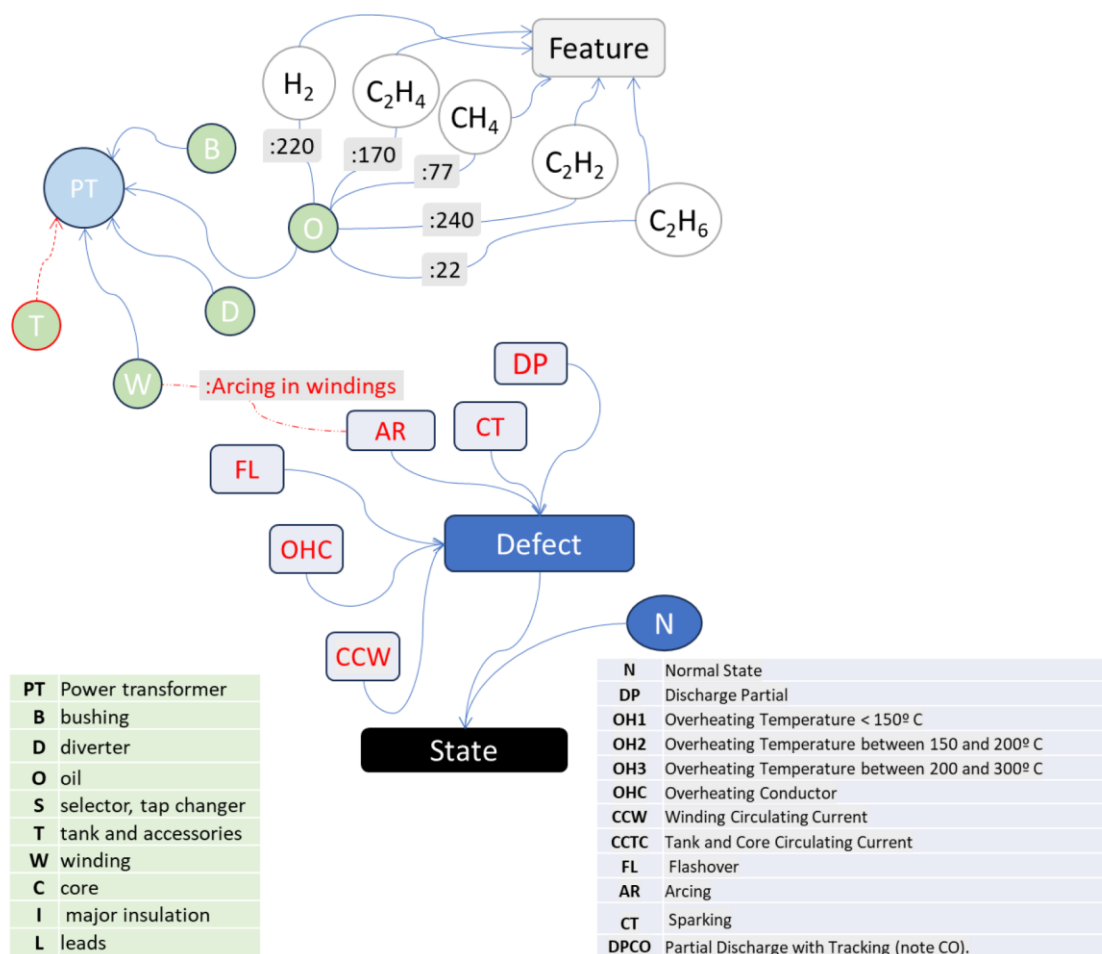
Old power grid equipment often becomes a source of accidents and failures, which leads to downtime, power losses and additional repair costs. At the same time, replacing all obsolete equipment requires significant financial and time expenditures. One of the developing trends in non-destructive diagnostics is the use of artificial intelligence to detect, verify and rank defects on power grid facilities. The general scientific research is devoted to the development of one-shot machine learning models for predicting defects and the pre-emergency state of electric grid equipment with a limited set of initial data. The source of information for training and prediction will be an ontological knowledge base of defective states of electrical equipment. The paper gives a formal representation of the ontological knowledge base, presents an overview of scientific research and proposed solutions in the field of processing linked data (graphs). The paper presents a formulation of the fundamental problem - restoration (or prediction) of missing links in a graph for identification of defects and prediction of emergency condition on electric grid facilities. A formal description of the problem of structural search for hidden links in the ontological knowledge base is given. An assumption has been made about the possibility of representing the basic structure of a graph using a multilayer perceptron, represented as a directed acyclic graph in which nodes correspond to neurons in a neural network, and faces correspond to connections between neurons.

**Keywords:** knowledge engineering, system analysis, ontological knowledge base, graph neural networks, non-destructive testing of equipment

**Financial Support:** the work was supported by the Russian Science Foundation, Project № 24-21-20021.

Graphical annotation (Графическая аннотация)

Структурный поиск скрытых связей между узлами графа



**ВВЕДЕНИЕ**

Электроэнергетическая отрасль сталкивается с проблемой наличия большого количества старого электрооборудования, что негативно сказывается на ее общем состоянии.

Возрастающая потребность в электроэнергии, растущая нагрузка на распределительные сети, а также требования к эффективности и надежности электрооборудования создают необходимость в модернизации и замене устаревшего оборудования. Старое сетевое оборудование может иметь низкую эффективность, большой расход энергии, высокие затраты на обслуживание и ремонт, а также устаревшие технологии, которые могут быть несовместимы с современными требованиями и стандартами. Это может приводить к снижению надежности сети, частым сбоям и перебоям в энергоснабжении. Для решения этой проблемы необходимо провести комплексную программу модернизации и замены старого сетевого оборудования, которая может включать в себя замену устаревших трансформаторов, высоковольтных и низковольтных линий, а также внедрение современных систем управления и мониторинга энергосетей. Однако такая программа модернизации может потребовать значительных инвестиций и времени для ее реализации.

В официальных отчетных документах Министерства энергетики говорится, что более 70 % оборудования отработало свой срок службы. То есть возраст оборудования подстанций линий превышает 40, а зачастую 50 лет. В период с 2022 по 2023 г. аварийность в распределительном комплексе составила 58 %. Такие большие цифры говорят о том, что вся техническая система находится на правой части кривой надёжности, когда оборудование по мере старения начинает чаще отказывать и выходить из строя. И соответственно, угроза, которая, в связи с этим возникает, – это экспоненциальное снижение качества, надежности и доступности электроэнергии для населения и бизнеса.

По оценкам экспертов, если проводить полную модернизацию электросетевого хозяйства традиционным способом, как это делали последние 50–60 лет, когда старый актив менялся на новый, без изменения топологии и без изменения технологий, то для того, чтобы решить эти проблемы, в России нужно выделить примерно 10 триллионов рублей – это очень большие суммы, которые могут привести к значительному росту тарифов.

Альтернативой может служить своевременная диагностика неисправностей и оценка состояния электросетевого оборудования. Однако диагностика старого оборудования может быть сложной и вызывать определенные трудности. Так старое оборудование зачастую лишено документации. Отсутствие информации о технических характеристиках, истории эксплуатации и предыдущих ремонтах затрудняет процесс диагностики и усложняет определение причины неисправности. Старое оборудование может не иметь современных систем диагностики или сенсоров, что затрудняет проведение точных и надежных тестов. Кроме того, возможность применения традиционных методов диагностики, таких как визуальный осмотр и испытания, может быть ограничена из-за сложности доступа к внутренним компонентам и недостаточной информации о состоянии оборудования.

Несмотря на эти проблемы, разработка и использование специализированных методов диагностики для оценки состояния старого оборудования может помочь в обнаружении и предотвращении неисправностей, а также в планировании ремонтных работ и замены оборудования. Современная диагностика может включать в себя использование неразрушающих методов, таких как ультразвуковая дефектоскопия, инфракрасная термография, акустическая эмиссия и другие, которые позволяют оценить состояние оборудования без его разборки или отключения от сети. Также важно разработать алгоритмы и программное обеспечение для анализа полученных данных и принятия решений о необходимости ремонта или замены оборудования.

Наше исследование посвящено разработке моделей машинного обучения прогнозирования дефектов и предаварийного состояния электросетевого оборудования при ограниченном наборе исходных данных. Это современный метод неразрушающего контроля выявления, верификации и ранжирования дефектов на электросетевых объектах. За последние несколько лет в области машинного обучения был достигнут ряд важнейших достижений в решении таких задач, как распознавания изображений, речи, текста и перевода языка. Решения чаще всего построены на глубоких нейронных сетях как аппроксиматоров функций и их способности обучаться сложным представлениям на основе большого количества входных данных. К сожалению, практический эффект от этих инноваций в области энергетики ограничен, поскольку большинство вышеупомянутых моделей для своего обучения требуют большие объемы данных, что контрастирует с состоянием современных систем диагностики электрооборудования, которые зачастую не в состоянии охарактеризовать даже несколько десятков дефектов.

То есть одним из ключевых препятствий на пути успешного применения алгоритмов искусственного интеллекта является проблема нехватки доступных данных в задаче диагностики электрооборудования. Решение этой проблемы требует тщательного планирования и исследований, чтобы обеспечить наличие достаточного количества данных для обучения модели и ее оценки.

Собрать большое количество необходимых данных для решения этой задачи очень сложно. Важно отметить, что маркированные данные мало доступны в этой области. Одна из причин этой проблемы может быть связана с тем, что в реальных условиях неисправности случаются редко и для сбора достаточного количества данных о неисправных состояниях может потребоваться много времени. Кроме того, не всегда возможно получить данные сразу от большого количества оборудования, что также может ограничить доступ к достаточному количеству данных. Так, согласно российским нормам, обмотку трансформатора следует проверять не реже одного раза в год. Другие конструктивные элементы (например, ленты, шпильки и т. д.) следует проверять не реже одного раза в 4 года. Расчеты передаточного числа трансформатора и сравнение фактических данных с показателями, указанными в технической документации, следует выполнять не реже одного раза в шесть лет. Для трансформаторов сухого типа испытания проводятся каждые шесть лет. Для определения эксплуатационных характеристик трансформатора раз в год отбираются пробы масла. Редкие испытания порождают малое количество данных, которого, очевидно, недостаточно для качественного обучения модели. Важно отметить, что количество классов дефектов для каждого типа электрооборудования огромно и динамически меняется, что также усложняет решение.

Последние достижения в области машинного обучения показали, что в некоторых случаях на основе небольшого объема данных можно делать нетривиальные прогнозы. Можно применить обычные методы кластеризации или классификации, основанные на плоской структуре данных, где каждый экземпляр оборудования связан с вектором параметрических атрибутов. Однако в задаче диагностирования состояния оборудования необходимо дополнительно исследовать отношения между параметрами оборудования и их состоянием. Онтологическое обучение позволяет решить эту и подобные ситуации, когда преобразование данных в плоскую структуру не является естественным.

Так, например, графовые нейронные сети (GNN), впервые представленные в работах [1, 2], дали многообещающие результаты для моделирования скрытых представлений графов. Алгоритм GNN расширяет возможности рекурсивных нейронных сетей и может быть применен к большинству типов графов, включая направленные, ненаправленные, маркированные и циклические графы. GNN уже достигли высочайшего уровня производительности в таких задачах, как прогнозирование связей, классификацию узлов и генерацию графов.

В нашем исследовании мы адаптируем метод кратковременного обучения графовых нейронных сетей к задаче поиска дефектов в электрооборудовании. Источником информации для обучения и прогнозирования будет служить онтологическая база знаний дефектных состояний электрооборудования.

### ОНТОЛОГИЧЕСКАЯ БАЗА ЗНАНИЙ ДЕФЕКТНЫХ СОСТОЯНИЙ ЭЛЕКТРООБОРУДОВАНИЯ

Формирование онтологической базы знаний о дефектных состояниях электрооборудования играет важную роль в развитии новой научной тематики. Это позволяет исследователям и инженерам более глубоко понять причины возникновения дефектов, разрабатывать эффективные методы и технологии их обнаружения и предотвращения. Одной из основных проблем, с которыми сталкиваются специалисты в области электрооборудования, является недостаток информации о типичных дефектах и способах их устранения. База знаний о дефектных состояниях электрооборудования позволяет собрать и систематизировать такую информацию, делая ее доступной для всех заинтересованных сторон.

Создание базы знаний о дефектных состояниях электрооборудования требует сбора и анализа данных о типичных дефектах, их симптомах и последствиях. Это может включать в себя информацию о различных видах электрооборудования, их характеристиках и особенностях эксплуатации, а также о методах диагностики и ремонта. База знаний о дефектных состояниях электрооборудования может быть использована в различных областях, включая научные исследования, инженерную практику и образование. Она позволит улучшить процессы обнаружения и устранения дефектов, повысит надежность и безопасность работы электрооборудования, а также позволит снизить затраты на его обслуживание. Анализ знаний, накопленных в базе, позволит выявлять скрытые паттерны и предсказывать возможные дефектные состояния оборудования, а также использовать для разработки новых неразрушающих методов диагностики электросетевого оборудования.

Для описания абстрактной структуры данных мы используем термин «онтологический граф», т. е. данные, структурированные в виде графов. Реальное воплощение этой структуры представляет собой сеть в виде графовой базы данных, описывающую связь реальных объектов, их свойств и состояний (от нормального до дефектного).

Онтологическая база знаний хранит коллекции кортежей вида (субъект, предикат, объект), которые удобно представлять в виде направленных маркированных мультиграфов, в которых сущности соответствуют узлам, а связи кодируются маркированными ребрами.

Формально онтологический граф  $G = (\mathcal{S}, \mathcal{R}, \mathcal{O})$  задается множеством двух типов узлов: субъектных  $\mathcal{S}$  и объектных  $\mathcal{O}$ , а также множеством ребер  $\mathcal{R}$  между этими узлами. Это означает, что мы можем разбить множество всех узлов на непересекающиеся множества  $\mathcal{S} \cap \mathcal{O} = \emptyset$ . Мы обозначаем ребро, идущее от узла  $s \in \mathcal{S}$  к узлу  $\sigma \in \mathcal{O}$ , как  $(s, t, \sigma) \in \mathcal{R}$ , где  $t$  определяет тип связи между узлами. Ребра обычно удовлетворяют ограничению, согласно которому определенного типа ребра соединяют только определенные типы узлов, т. е.  $(s, t_i, \sigma) \in \mathcal{R} \rightarrow s \in \mathcal{S}, \sigma \in \mathcal{O}$ .

Обычно онтологический граф представляют в виде трехмерной матрицы смежности для каждого типа связи  $A_t \in \mathbb{R}^{|\mathcal{S}| \times |\mathcal{R}| \times |\mathcal{O}|}$ , для этого субъекты, предикаты и объекты графа упорядочиваются таким образом, чтобы каждый субъект был проиндексирован по  $x$ -измерению, объект по  $y$ -измерению, а предикат по  $z$ -измерению.

Онтологическая база знаний дефектных состояний электрооборудования, прежде всего, должна содержать наиболее критичное и важное электрооборудование для энергетического сектора, например, подстанцию, трансформатор, генератор, а также параметры и неисправности, которые необходимо отслеживать и анализировать. В работе был проведен анализ конструкции наиболее критичного для отрасли электрооборудования, включающий изучение структуры и компонентов оборудования, анализ иерархии и связей между ними, а также изучение физических и механических характеристик каждого компонента. Анализ условий эксплуатации, включающий изучение рабочих параметров и режимов работы оборудования, таких как температура, влажность, напряжение, ток и другие факторы, которые могут влиять на его состояние: анализ и классификация дефектов, повреждений и отказов, их характеристики, причины возникновения, а также их влияние на работоспособность оборудования и связанных с ними аварийных состояний. Важно отметить, что в общих исследованиях практически не рассматривается вопрос о наличии нескольких, одновременно развивающихся дефектов, которые могут существенно исказить картину результатов диагностики оборудования. Онтологическая база знаний может представить причинно-следственную связь дефектов и повреждений и отобразить комплексность аварийного состояния.

Результатом работы стала онтология (рис. 1) и графическая база данных Neo4J (рис. 2).

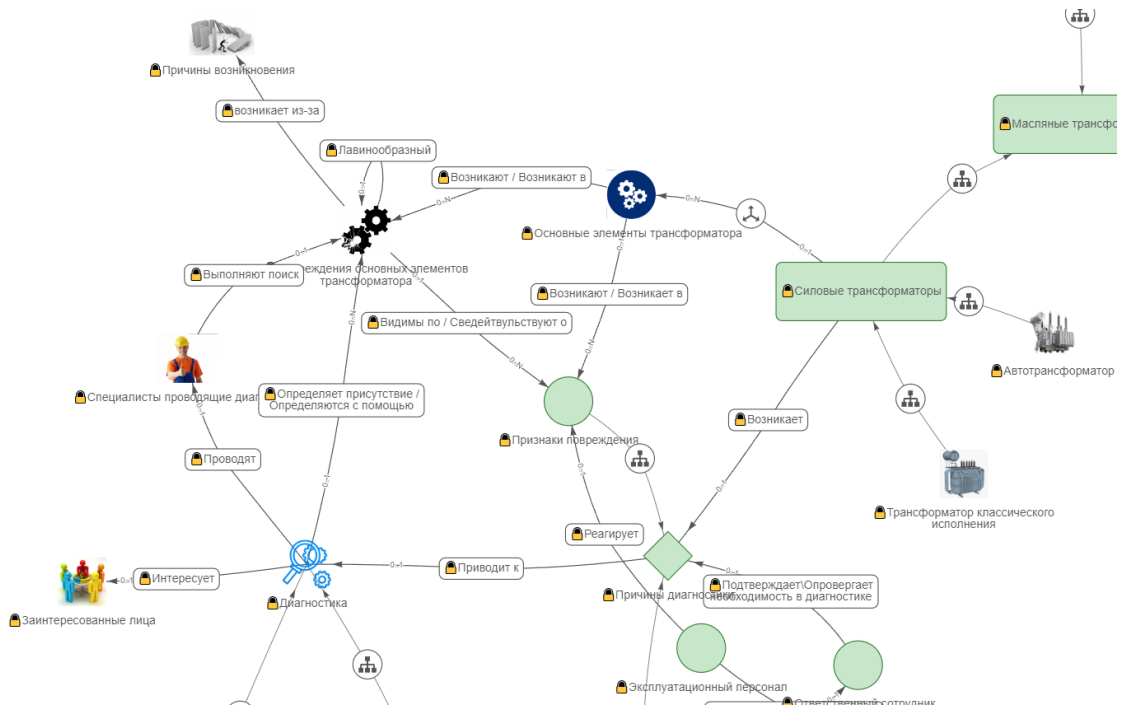


Рисунок 1 – Онтологическая база знаний дефектных состояний электрооборудования

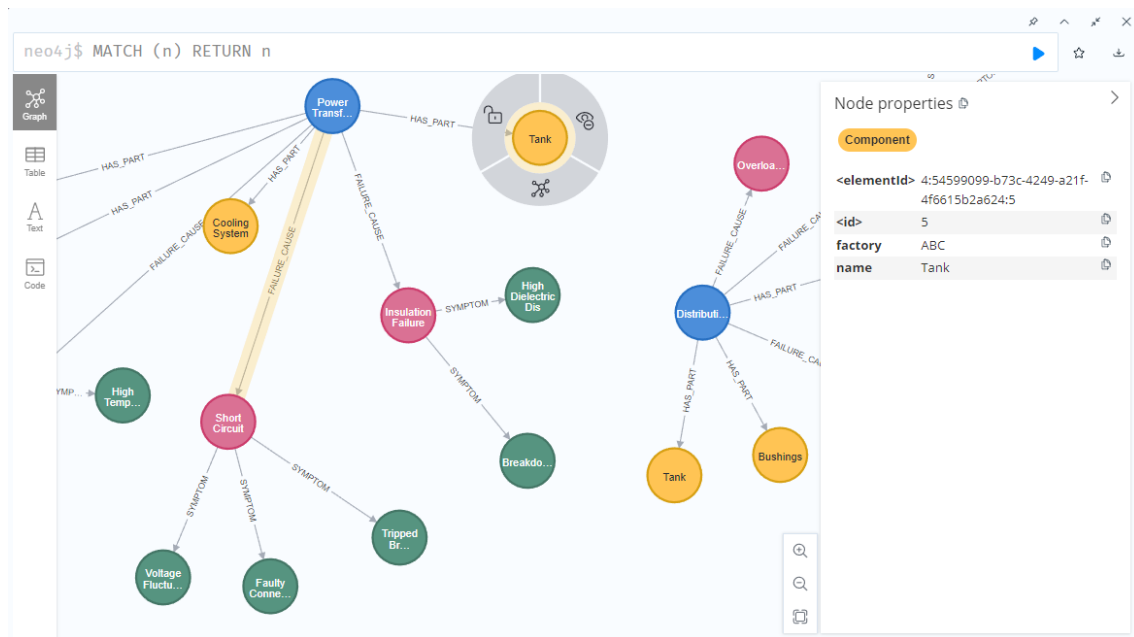


Рисунок 2 – Физическая реализация онтологической базы знаний в Neo4j

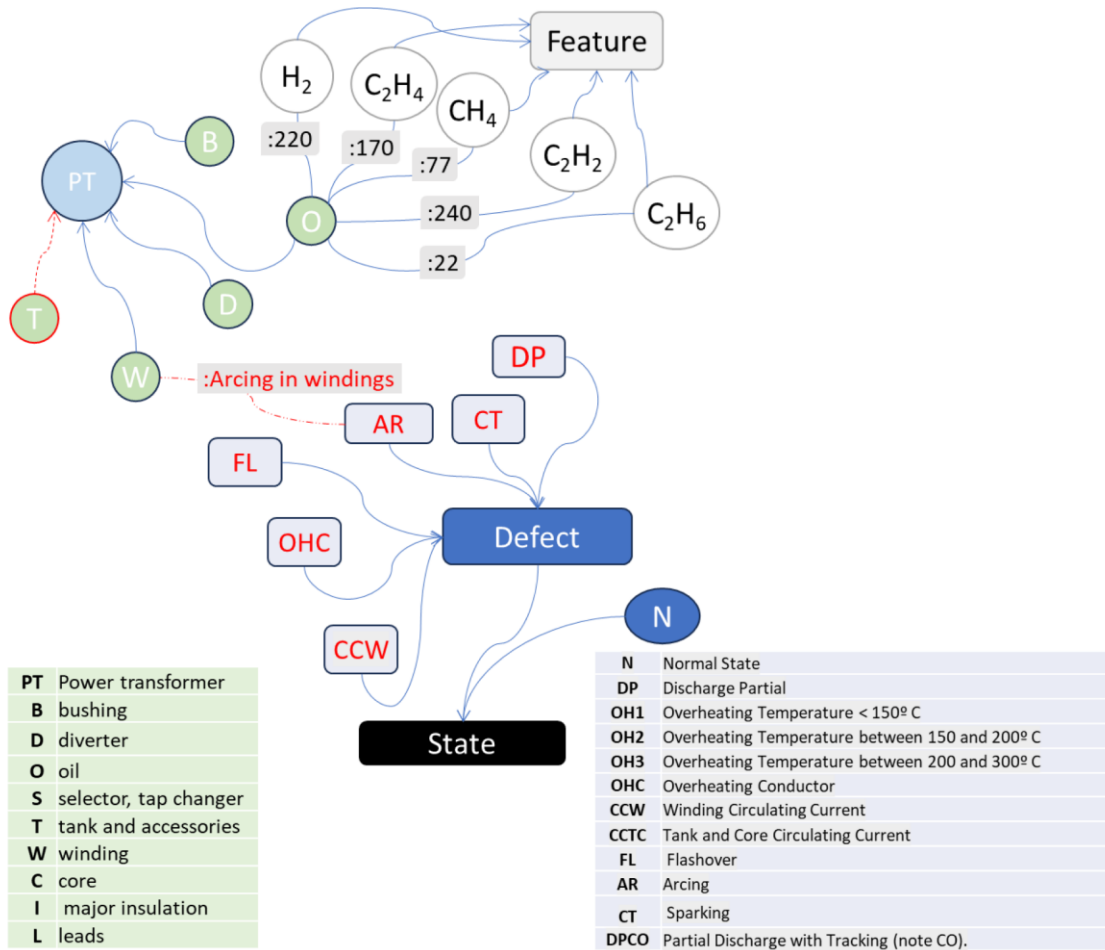


Рисунок 3 – Структурный поиск скрытых связей между узлами графа

Для базовой формы модели мы предполагаем, что существует конечное число классов, к которым могут принадлежать субъекты и объекты, и что эти классы полностью определяют структуру графа, причем вероятность существования связи между субъектом и объектом зависит только от того, к каким классам они принадлежат.

Субъектный узел графа, описывающий определенный класс оборудования, можно разложить на набор фрагментов, каждый из которых представляет собой конструктивные элементы (части) этого оборудования (рис. 3). То есть разбиение множества  $\mathcal{S}$  – это множество  $\mathcal{P}$  непустых, непересекающихся подмножеств  $\mathcal{S}$ , таких, что объединение множеств  $\mathcal{P}$  равно  $\mathcal{S}$ . Каждому фрагменту  $\mathcal{P}$  присваивается уникальный идентификатор, и совокупность идентификаторов субъектного узла хэшируется в битовый вектор фиксированной длины для построения структурной модели.

Объектный узел графа представляет собой определенный класс состояния конкретного оборудования.

Предикаты представляют собой параметризованные связи между субъектным и объектным узлами и в совокупности являются индексом исправности конкретного экземпляра оборудования.

### СТРУКТУРНЫЙ ПОИСК СКРЫТЫХ СВЯЗЕЙ МЕЖДУ УЗЛАМИ ГРАФА

На рисунке 3 мы видим грань и узел, выделенные красным цветом. Это недостающая информация, которую необходимо найти.

Восстановление связей связано с прогнозированием новых фактов (т. е. тройки (субъект, предикат, объект)). Формально онтологическая база знаний дефектных состояний электрооборудования представлена в виде направленного маркированного графа  $\mathcal{G} = (\mathcal{S}, \mathcal{R}, \mathcal{O})$ , в котором связь между конкретным экземпляром оборудования и дефектом представляет собой неполное подмножество  $\mathcal{R}$ . Задача присвоить веса  $f(s, t, o)$  возможным ребрам  $(s, t, o)$ , чтобы определить, насколько вероятно, что эти ребра принадлежат  $\mathcal{R}$ .

Мы предполагаем, каждый субъектный узел  $s \in \mathcal{S}$  связан со случайной переменной  $x_i \in X$ , каждый объектный узел  $o \in \mathcal{O}$  связан с функцией  $f_o$ , и предикат может соединить переменную  $x_i$  с объектной вершиной  $o$ , если  $f_o$  зависит от  $x_i$ .

Пусть  $x_\sigma$ , от которых зависит  $f_\sigma$ , является подмножеством всех переменных  $X$ . Обозначим множество индексов переменных  $x_\sigma$  через  $\mathcal{F}(\sigma)$ . В этом случае дискретная вероятность существования связи между субъектным узлом и объектным можно выразить формулой (1):

$$p(x) = \frac{1}{Z} \exp \left[ \sum_{\sigma \in \mathcal{O}} f_\sigma(x_\sigma) + \sum_{s \in \mathcal{S}} f_s(x_s) \right], \quad (1)$$

где  $\exp(f_\sigma(\cdot))$ ,  $\exp(f_s(\cdot))$ , – положительные функции, называемые потенциалами (где  $f_\sigma(\cdot)$ ,  $f_s(\cdot)$  – соответствующие логарифмические потенциалы), а  $Z$  – нормализующая константа. Когда полное байесовское решение непрактично, мы можем прибегнуть к использованию максимальной апостериорной оценки [42] (формула (2)), т. е. ищем параметры, которые максимизируют апостериорную вероятность  $p(x)$ :

$$\hat{x} = \arg \max_x \prod_{\sigma \in \mathcal{O}} f_\sigma(x_\sigma) = \arg \max_x \sum_{s \in \mathcal{S}} \log f_s(x_s). \quad (2)$$

Поскольку наша цель – прогнозировать существование связей между признаками, характеризующими электрооборудование, и признаками, характеризующими определенное состояние электрооборудования, то мы должны учитывать как существующие, так и несуществующие связи, т. е. формировать набор потенциальных связей между этими узлами.

Среди подходов, которые предсказывают связи отдельных объектов, можно выделить подходы, основанные на классах, и подходы, основанные на признаках. Большинство подходов, предлагаемых в этом направлении, ориентированы на разработку необучаемых алгоритмов. Разработки алгоритмов машинного обучения в основном сосредоточены на моделях скрытых классов, адаптируя байесовские непараметрические методы для определения количества скрытых классов и изучения того, какие объекты принадлежат к каждому классу.

Сходство графов обычно определяется либо точными совпадениями (изоморфизм всего графа или подграфа) [3–6], либо некоторой степенью структурного сходства [7, 8]. В работах [9, 10] был предложен метод классификации графов на основе сходства ядра графа. К популярным графическим ядрам относятся те, которые измеряют сходство между связями в графах с подструктурами ограниченного размера [11, 12], и ядра, основанные на поддеревовидных структурах [13]. Графовые ядра обычно используются в моделях, которые могут содержать обучаемые компоненты, но сами ядра разрабатываются вручную и основаны на теории графов. Обычно они могут быть сформулированы следующим образом: сначала вычисляются векторы признаков для каждого графа (встраивание в ядро), а затем берется внутреннее произведение между этими векторами для вычисления значения ядра. Большинство графовых ядер и графовых нейронных сетей (GNN) созданы на основе теста Вайсфайлера – Лемана [14] (WL-алгоритм) из-за его алгоритмической простоты, однако они работают только на графах с дискретными (категориальными) атрибутами, что ограничивает их использование. Суть WL-алгоритма сводится к агрегированию окрестностей. Так, в работе [16] описана графовая нейронная сеть GraphSAGE, в которой применяется простое усреднение. В работе [15] предложена графовая сверточная нейронная сеть GCN, которая использует в качестве агрегирования усреднения, нормализованного по степени. В работе [17] вместо среднего значения используется сумма. В других моделях GNN, таких как GAT [18], GatedGCN [19] и MoNet [20], перед агрегированием присваивают соседним элементам разные веса в зависимости от их атрибутов. Все упомянутые выше методы, основанные на свертке графов, используют одну и ту же стратегию: сначала агрегируют, а затем преобразуют. В статье [21] исследован проблемный аспект стратегии агрегат-преобразования применительно к графам с узлами, имеющими непрерывные вектора атрибутов, и предложено решение, основанное на объединении агрегатной информации с информацией из центрального узла.

Структурный поиск скрытых связей представляет собой фундаментальную задачу восстановления (или прогнозирования) отсутствующих связей в графе. Эта задача имеет множество решений, например, поиск на основе эвристики графа [28, 29], матричная факторизация [30], вероятностные методы [22, 23], а также стохастический процесс, индуцированный тензорным взаимодействием нескольких гауссовских процессов [24]. Поиск скрытых связей нашел свое применение в таких практических задачах, как рекомендательные системы, предсказание будущих связей в социальных сетях, функциональной геномике и системной биологии, прогнозировании эволюции сложных сетей, изучение недостающей или частичной информации в сетях и т. д. Так, авторы статьи [22] применяют структуру реляционной сети Маркова для определения общей вероятностной модели для всего графа связей атрибутов и связей объектов социальной сети и веб-страниц. Авторы применяют вероятностную модель, одновременно пытаясь предсказать и классифицировать весь набор связей в графе связей. Были выявлены два типа шаблонов, значительно повышающие точность классификации по сравнению с плоскими моделями: шаблон подобия, определяющий классификацию связей или объектов, обладающих определенным свойством, основанным на графах (например, ссылки,

имеющие общую конечную точку); шаблон транзитивности, определяющий тройки объектов и связей, организованные в виде треугольника. Авторы статьи [24] с помощью стохастических реляционных моделей и иерархических гауссовских процессов [27] обучения модели на неориентированных графах решают вычислительную проблему, которая имеет решающее значение для предсказания связей, поскольку количество потенциальных связей растет экспоненциально с увеличением размера графа. Вероятностные реляционные модели уже давно используются в машинном обучении. Так, чтобы избежать трудностей структурного обучения, авторы [25, 26] ввели бесконечную (скрытую) реляционную модель, объясняя связи через потенциально бесконечное число скрытых состояний сущностей. Методы встраивания сети, такие как DeepWalk [31], LINE [32] и node2vec [33], а также TransE [34], DistMult [35] и RotatE [36] для графов знаний, в которых для предсказания связей изучаются как латентные признаки, так и явные признаки узлов графа. Последние работы посвящены обучению на графах, изменяющихся во времени (динамические графы) [37, 38]. Несмотря на богатое разнообразие методов и моделей GNN, понимание графов знаний, которые применяются в области прогнозирования состояния объекта, довольно ограничено.

### **ЗАКЛЮЧЕНИЕ**

Разработка и использование информационных систем, методов и средств контроля на основе искусственного интеллекта, а также переход к интеллектуальному управлению и инжинирингу позволит уменьшить влияние человеческого фактора, уйти от устаревших методик диагностики технического состояния электрооборудования, своевременно принимать решение о необходимости технического обслуживания, ремонте или замене оборудования, правильно расставить приоритеты при обслуживании, а также спрогнозировать работу и избежать сбоев.

Техническая инфраструктура электроэнергетики имеет сложную структуру и включает объекты различных типов, которые связаны друг с другом сетью различных типов связей. Такие данные создают новые проблемы для машинного обучения. Одна из проблем возникает в связи с задачей предсказать, какие объекты связаны с какими-либо другими и каковы типы этих связей.

Формирование базы знаний о дефектных состояниях электрооборудования играет важную роль в развитии новой научной тематики. Это позволяет исследователям и инженерам более глубоко понять причины возникновения дефектов, разрабатывать эффективные методы и технологии их обнаружения и предотвращения. Одной из основных проблем, с которыми сталкиваются специалисты в области электрооборудования, является недостаток информации о типичных дефектах и способах их устранения. База знаний о дефектных состояниях электрооборудования позволяет собрать и систематизировать такую информацию, делая ее доступной для всех заинтересованных сторон. Формирование базы знаний требует обширного исследования и сбора информации о различных типах оборудования, их характеристиках, возможных дефектах и их причинах, методах диагностики и ремонта и т. д. Данные для формирования базы знаний могут быть получены из различных источников, таких как научные статьи, технические фотчеты, руководства по эксплуатации, опытные данные от специалистов и т. д. Это требует сбора, анализа и систематизации информации из разных источников. Также необходимо учитывать, что для обработки разного типа данных, например, таких как электрические параметры, температура, вибрация, звуковые сигналы и другие, требуется разработка специализированных алгоритмов и моделей машинного обучения. Состояние электросетевого оборудования может изменяться со временем, и могут появляться новые типы дефектов. Поэтому база знаний должна быть постоянно обновляемой и адаптируемой, чтобы учитывать новые данные и изменения в оборудовании. Это требует постоянного мониторинга и внесения новой информации в базу данных. Данные о дефектах электросетевого оборудования требуют классификации и структурирования для удобного использования. Это может включать создание системы категорий и тегов для классификации дефектов и связанных с ними информации. Для эффективной диагностики и оценки состояния оборудования требуются экспертные знания в области электротехники. В целом задача формирования базы знаний о дефектах электросетевого оборудования и применение искусственного интеллекта в диагностике неисправностей и оценке состояния оборудования представляют собой сложную и масштабную задачу, требующую глубоких знаний в области электротехники, разработки алгоритмов машинного обучения и обработки больших объемов данных. Такое исследование будет иметь практическую значимость, поскольку позволит повысить надежность и эффективность работы электросетевого оборудования, снизить риски аварийных ситуаций и сократить затраты на обслуживание и ремонт. Кроме того, разработанные методы и алгоритмы могут быть применимы не только в энергетической отрасли, но и в других областях, где требуется диагностика и оценка состояния технических систем.

Традиционные подходы работают со связанными структурами данных, используя этап предварительной обработки, который преобразует граф в набор плоских векторов. Однако таким образом может быть потеряна важная топологическая информация, и достигнутые результаты могут сильно зависеть от стадии предварительной обработки.

Структурный поиск скрытых связей – это область исследований, направленная на поиск граней графа, представляющих причинно-следственную связь или зависимость между переменными. Эта область связана с нашей работой, поскольку мы разделяем идею обнаружения сложных взаимодействий между наблюдаемыми или скрытыми переменными. Фактически онтологическую базу знаний дефектных состояний электрооборудования можно представить в виде многослойного перцептрона, в котором узлы соответствуют нейронам, а грани – связям между нейронами. Прямой проход через многослойный перцептрон равнозначен прямому проходу асинхронной GNN. Такое графовое представление дефектных состояний электрооборудования открывает новое направление разработки неразрушающих методов диагностики электрооборудования, используя хорошо разработанную модель GNN.

#### Список источников

1. Marco, Gori. A new model for learning in graph domains / Marco Gori, Gabriele Monfardini, and Franco Scarselli // IEEE International Joint Conference on Neural Networks (IJCNN). – IEEE, 2005. – Vol. 2. – P. 729–734. – URL: [https://www.researchgate.net/publication/4202380\\_A\\_new\\_model\\_for\\_learning\\_in\\_graph\\_domains](https://www.researchgate.net/publication/4202380_A_new_model_for_learning_in_graph_domains).
2. Franco, Scarselli. The graph neural network model / Franco Scarselli, Marco Gori, Ah Chung Tsoi, Markus Hagenbuchner, and Gabriele Monfardini // IEEE Transactions on Neural Networks. – 2009. – № 20 (1). – P. 61–80. – URL: <https://pubmed.ncbi.nlm.nih.gov/19068426/>.
3. Stefano, Berretti. Efficient matching and indexing of graph models in content-based retrieval / Stefano Berretti, Alberto Del Bimbo and Enrico Vicario // IEEE Transactions on Pattern Analysis and Machine Intelligence. – 2001. – № 23 (10). – P. 1089–1105. – URL: <https://ieeexplore.ieee.org/document/954600>.
4. Dennis, Shasha. Algorithmics and applications of tree and graph searching / Dennis Shasha, Jason TL Wang, and Rosalba Giugno // Proceedings of the twenty-first ACM SIGMOD-SIGACT symposium on Principles of database systems. – ACM, 2002. – P. 39–52. – URL: [https://www.researchgate.net/publication/2488274\\_Algorithmics\\_and\\_Applications\\_of\\_Tree\\_and\\_Graph\\_Searching](https://www.researchgate.net/publication/2488274_Algorithmics_and_Applications_of_Tree_and_Graph_Searching).
5. Srinath, Srinivasa. A platform based on the multi-dimensional data model for analysis of bio-molecular structures / Srinath Srinivasa and Sujit Kumar // Proceedings of the VLDB Conference. – Elsevier, 2003. – P. 975–986. URL: [https://www.researchgate.net/publication/221309832\\_A\\_Platform\\_Based\\_on\\_the\\_Multi-dimensional\\_Data\\_Model\\_for\\_Analysis\\_of\\_Bio-Molecular\\_Structures](https://www.researchgate.net/publication/221309832_A_Platform_Based_on_the_Multi-dimensional_Data_Model_for_Analysis_of_Bio-Molecular_Structures).
6. Xifeng, Yan. Graph indexing: a frequent structure-based approach / Xifeng Yan, Philip S. Yu, and Jiawei Han // Proceedings of the ACM SIGMOD international conference on Management of data. – ACM, 2004. – P. 335–346. – URL: [https://www.researchgate.net/publication/2915152\\_Graph\\_Indexing\\_A\\_Frequent\\_Structure-based\\_Approach](https://www.researchgate.net/publication/2915152_Graph_Indexing_A_Frequent_Structure-based_Approach).
7. Peter, Willett. Chemical similarity searching / Peter Willett, John M Barnard, and Geoffrey M Downs // Journal of Chemical Information and Computer Sciences. – 1998. – № 38 (6). – P. 983–996. – URL: <https://www.sci-hub.ru/10.1021/ci9800211>.
8. John, W. Raymond, Rascal: Calculation of graph similarity using maximum common edge subgraphs / John W. Raymond, Eleanor J. Gardiner, and Peter Willett // The Computer Journal. – 2002. – № 45 (6). – P. 631–644. – URL: [https://www.researchgate.net/publication/220457922\\_RASCAL\\_Calculation\\_of\\_Graph\\_Similarity\\_using\\_Maximum\\_Common\\_Edge\\_Subgraphs](https://www.researchgate.net/publication/220457922_RASCAL_Calculation_of_Graph_Similarity_using_Maximum_Common_Edge_Subgraphs).
9. Vishwanathan, S. V. N. Graph kernels / S. V. N. Vishwanathan, Nicol N. Schraudolph, Risi Kondor, and Karsten M. Borgwardt // Journal of Machine Learning Research. – 2010. – № 11 (Apr.). – P. 1201–1242. – URL: <https://jmlr.org/papers/volume11/vishwanathan10a/vishwanathan10a.pdf>.
10. Nino, Shervashidze. Weisfeiler-lehman graph kernels / Nino Shervashidze, Pascal Schweitzer, Erik Jan van Leeuwen, Kurt Mehlhorn, and Karsten M. Borgwardt // Journal of Machine Learning Research. – 2011. – № 12 (Sep.). – P. 2539–2561. – URL: <https://jmlr.org/papers/volume12/shervashidze11a/shervashidze11a.pdf>.
11. Karsten, M. Borgwardt. Shortest-path kernels on graphs / Karsten M. Borgwardt and Hans-Peter Kriegel // Fifth IEEE International Conference. – IEEE, 2005. – P. 8. – URL: [https://www.researchgate.net/publication/4207736\\_Shortest-Path\\_Kernels\\_on\\_Graphs](https://www.researchgate.net/publication/4207736_Shortest-Path_Kernels_on_Graphs).
12. Hisashi, Kashima. Marginalized kernels between labeled graphs / Hisashi Kashima, Koji Tsuda, and Akihiro Inokuchi // Proceedings of the 20th International Conference on Machine Learning (ICML-03). – 2003. – P. 321–328. – URL: [https://www.researchgate.net/publication/221345917\\_Marginalized\\_Kernels\\_Between\\_Labeled\\_Graphs](https://www.researchgate.net/publication/221345917_Marginalized_Kernels_Between_Labeled_Graphs).
13. Nino, Shervashidze. Fast subtree kernels on graphs / Nino Shervashidze and Karsten M. Borgwardt // Advances in neural information processing systems. – 2009. – P. 1660–1668. – URL: [https://www.researchgate.net/publication/221617867\\_Fast\\_subtree\\_kernels\\_on\\_graphs](https://www.researchgate.net/publication/221617867_Fast_subtree_kernels_on_graphs).
14. Вайсфайлер, Б. Приведение графа к канонической форме и алгебра, которая в нем появляется / Б. Вайсфайлер, А. А. Леман // Серия Научно-Техническая Информация. – 1968. – № 2 (9). – С. 12–16.
15. Kipf, Thomas N. Semi-supervised classification with graph convolutional networks / Thomas N. Kipf, & Max Welling // Fifth International Conference on Learning Representations (ICLR 2017). – 2017. – URL: <https://arxiv.org/abs/1609.02907>.
16. Hamilton, William L. Inductive representation learning on large graphs / William L. Hamilton, Rex Ying, & Jure Leskovec // Advances in Neural Information Processing Systems (NIPS 2017). – 2017. – № 30. – URL: <https://arxiv.org/abs/1706.02216>.
17. Xu, Keyulu. How powerful are graph neural networks? / Keyulu Xu, Weihua Hu, Jure Leskovec, & Stefanie Jegelka // Seventh International Conference on Learning Representations (ICLR 2019). – 2019. – URL: <https://arxiv.org/abs/1810.00826>.

18. Veličkovič, Petar. Graph attention networks / Petar Veličkovič, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Liò, & Yoshua Bengio // Sixth International Conference on Learning Representations (ICLR 2018). – 2018. – URL: <https://typeset.io/papers/graph-attention-networks-1xed6mcmzg>.
19. Bresson, Xavier. Residual gated graph ConvNets / Xavier Bresson, & Thomas Laurent // arXiv preprint arXiv:1711.07553. – 2017. – URL: <https://arxiv.org/abs/1711.07553>.
20. Monti, Federico. Geometric deep learning on graphs and manifolds using mixture model CNNs / Federico Monti, Davide Boscaini, Jonathan Masci, Emanuele Rodolà, Jan Svoboda, & Michael M. Bronstein // 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). – 2017. – URL: <https://arxiv.org/abs/1611.08402>.
21. Seongmin, Ok. A graph similarity for deep learning / Ok. Seongmin // Advances in Neural Information Processing Systems. – 2020. – URL: <https://paperswithcode.com/paper/a-graph-similarity-for-deep-learning>.
22. Ben, Taskar. Link Prediction in Relational Data / Ben Taskar, Ming-Fai Wong, Pieter Abbeel, Daphne Koller // Advances in Neural Information Processing Systems 16 [Neural Information Processing Systems, NIPS 2003, December 8–13, 2003, Vancouver and Whistler, British Columbia, Canada]. – URL: [https://www.researchgate.net/publication/221618061\\_Link\\_Prediction\\_in\\_Relational\\_Data](https://www.researchgate.net/publication/221618061_Link_Prediction_in_Relational_Data).
23. Taskar, B. Probabilistic classification and clustering in relational data / B. Taskar, E. Segal, and D. Koller. // Proc. IJCAI. – 2001. – P. 870–876. – URL: [https://www.researchgate.net/publication/220812979\\_Probabilistic\\_Classification\\_and\\_Clustering\\_in\\_Relational\\_Data](https://www.researchgate.net/publication/220812979_Probabilistic_Classification_and_Clustering_in_Relational_Data).
24. Kai, Yu. Stochastic Relational Models for Discriminative Link Prediction / Kai Yu, Wei Chu, Shipeng Yu, Volker Tresp, Zhao Xu // Advances in Neural Information Processing Systems. – 2007. – № 19. – P. 1553–1560. – URL: [https://www.researchgate.net/publication/371869717\\_Stochastic\\_Relational\\_Models\\_for\\_Discriminative\\_Link\\_Prediction](https://www.researchgate.net/publication/371869717_Stochastic_Relational_Models_for_Discriminative_Link_Prediction).
25. Kemp, C. Learning systems of concepts with an infinite relational model / C. Kemp, J. B. Tenenbaum, T. L. Griffiths, T. Yamada, and N. Ueda // Proceedings of the 21st National Conference on Artificial Intelligence (AAAI). – 2006. – URL: [https://www.researchgate.net/publication/215991886\\_Learning\\_Systems\\_of\\_Concepts\\_with\\_an\\_Infinite\\_Relational\\_Model](https://www.researchgate.net/publication/215991886_Learning_Systems_of_Concepts_with_an_Infinite_Relational_Model).
26. Xu, Z. Infinite hidden relational models / Z. Xu, V. Tresp, K. Yu, and H.-P. Kriegel // Proceedings of the 22nd International Conference on Uncertainty in Artificial Intelligence (UAI). – 2006. – URL: <https://arxiv.org/abs/1206.6864>.
27. Chu, W., Relational learning with gaussian processes / W. Chu, V. Sindhwani, Z. Ghahramani, and S. S. Keerthi. // Advances in Neural Information Processing Systems 19 // Proceedings of the Twentieth Annual Conference on Neural Information Processing Systems, Vancouver, British Columbia, Canada, December 4–7, 2006. – URL: [https://www.researchgate.net/publication/221620080\\_Relational\\_Learning\\_with\\_Gaussian\\_Processes](https://www.researchgate.net/publication/221620080_Relational_Learning_with_Gaussian_Processes).
28. Liben-Nowell, D. The link-prediction problem for social networks / D. Liben-Nowell and J. Kleinberg // Journal of the American society for information science and technology. – 2007. – № 58 (7). – P. 1019–1031. – URL: <https://www.sci-hub.ru/10.1002/asi.20591>.
29. Sarkar, P. Theoretical justification of popular link prediction heuristics / P. Sarkar, D. Chakrabarti, and A. W. Moore // International Conference on Learning Theory (COLT). – 2010. – P. 295–307. – URL: [https://www.researchgate.net/publication/220815677\\_Theoretical\\_Justification\\_of\\_Popular\\_Link\\_Prediction\\_Heuristics](https://www.researchgate.net/publication/220815677_Theoretical_Justification_of_Popular_Link_Prediction_Heuristics).
30. Koren, Y. Factorization meets the neighborhood: a multifaceted collaborative filtering model / Y. Koren // Proceeding of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining. – ACM, 2008. – P. 426–434.
31. Bryan, Perozzi, Deepwalk: Online learning of social representations / Bryan Perozzi, Rami Al-Rfou, and Steven Skiena // Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining. – ACM, 2014. – P. 701–710. – URL: <https://arxiv.org/abs/1403.6652v2>.
32. Jian, Tang. Line: Large-scale information network embedding / Jian Tang, Meng Qu, Mingzhe Wang, Ming Zhang, Jun Yan, and Qiaozhu Mei // Proceedings of the 24th International Conference on World Wide Web. – 2015. – P. 1067–1077. – URL: <https://paperswithcode.com/paper/line-large-scale-information-network>.
33. Aditya, Grover. node2vec: Scalable feature learning for networks / Aditya Grover and Jure Leskovec // Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining. – ACM, 2016. – P. 855–864. – URL: <https://arxiv.org/abs/1607.00653>.
34. Antoine, Bordes, Translating embeddings for modeling multi-relational data / Antoine Bordes, Nicolas Usunier, Alberto Garcia-Duran, Jason Weston, and Oksana Yakhnenko // Advances in Neural Information Processing Systems. – 2013. – P. 1–9. – URL: <https://paperswithcode.com/paper/translating-embeddings-for-modeling-multi>.
35. Bishan, Yang, Embedding entities and relations for learning and inference in knowledge bases / Bishan Yang, Wen-tau Yih, Xiaodong He, Jianfeng Gao, and Li Deng // International Conference on Learning Representations. – 2015. – URL: <https://arxiv.org/abs/1412.6575>.
36. Zhiqing, Sun. Rotate: Knowledge graph embedding by relational rotation in complex space / Zhiqing Sun, Zhi-Hong Deng, Jian-Yun Nie, and Jian Tang // International Conference on Learning Representations. – 2019. – URL: <https://arxiv.org/abs/1902.10197v1>.
37. Jiahao, Liu. Parameter-free Dynamic Graph Embedding for Link Prediction / Jiahao Liu, Dongsheng Li, Hansu Gu, Tun Lu, Peng Zhang, Ning Gu. – 2022. – URL: [https://www.researchgate.net/publication/364438325\\_Parameter-free\\_Dynamic\\_Graph\\_Embedding\\_for\\_Link\\_Prediction](https://www.researchgate.net/publication/364438325_Parameter-free_Dynamic_Graph_Embedding_for_Link_Prediction).
38. Farimah, Poursafaei. Towards Better Evaluation for Dynamic Link Prediction / Farimah Poursafaei, Shenyang Huang, Kellin Pelrine, Reihaneh Rabbany. – 2022. – URL: <https://arxiv.org/abs/2207.10128>.

## References

1. Marco, Gori, Gabriele, Monfardini, and Franco, Scarselli. A new model for learning in graph domains. *IEEE International Joint Conference on Neural Networks (IJCNN)*, 2005, vol. 2, pp. 729–734. URL: [https://www.researchgate.net/publication/4202380\\_A\\_new\\_model\\_for\\_learning\\_in\\_graph\\_domains](https://www.researchgate.net/publication/4202380_A_new_model_for_learning_in_graph_domains).
2. Franco, Scarselli, Marco, Gori, Ah Chung, Tsoi, Markus, Hagenbuchner, and Gabriele, Monfardini. The graph neural network model. *IEEE Transactions on Neural Networks*, 2009, no. 20 (1), pp. 61–80. URL: <https://pubmed.ncbi.nlm.nih.gov/19068426/>.
3. Stefano, Berretti, Alberto, Del Bimbo, and Enrico, Vicario. Efficient matching and indexing of graph models in content-based retrieval. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2001, no. 23 (10), pp. 1089–1105. URL: <https://ieeexplore.ieee.org/document/954600>.
4. Dennis, Shasha, Jason, TL Wang, and Rosalba, Giugno. Algorithmics and applications of tree and graph searching. *Proceedings of the twenty-first ACM SIGMOD-SIGACT symposium on Principles of database systems*, ACM, 2002, pp. 39–52. URL: [https://www.researchgate.net/publication/2488274\\_Algorithmics\\_and\\_Applications\\_of\\_Tree\\_and\\_Graph\\_Searching](https://www.researchgate.net/publication/2488274_Algorithmics_and_Applications_of_Tree_and_Graph_Searching).
5. Srinath, Srinivasa and Sujit, Kumar. A platform based on the multi-dimensional data model for analysis of bio-molecular structures. *Proceedings of the VLDB Conference*, 2003, pp. 975–986. URL: [https://www.researchgate.net/publication/221309832\\_A\\_Platform\\_Based\\_on\\_the\\_Multi-dimensional\\_Data\\_Model\\_for\\_Analysis\\_of\\_Bio-Molecular\\_Structures](https://www.researchgate.net/publication/221309832_A_Platform_Based_on_the_Multi-dimensional_Data_Model_for_Analysis_of_Bio-Molecular_Structures).
6. Xifeng, Yan, Philip, S. Yu, and Jiawei, Han. Graph indexing: a frequent structure-based approach. *Proceedings of the ACM SIGMOD international conference on Management of data*, 2004, pp. 335–346. URL: [https://www.researchgate.net/publication/2915152\\_Graph\\_Indexing\\_A\\_Frequent\\_Structure-based\\_Approach](https://www.researchgate.net/publication/2915152_Graph_Indexing_A_Frequent_Structure-based_Approach).
7. Peter, Willett, John, M. Barnard, and Geoffrey, M. Downs. Chemical similarity searching. *Journal of Chemical Information and Computer Sciences*, 1998, no. 38 (6), pp. 983–996. URL: <https://www.sci-hub.ru/10.1021/ci9800211>.
8. John, W. Raymond, Eleanor, J. Gardiner, and Peter, Willett. Rascal: Calculation of graph similarity using maximum common edge subgraphs. *The Computer Journal*, 2002, no. 45 (6), pp. 631–644. URL: [https://www.researchgate.net/publication/220457922\\_RASCAL\\_Calculation\\_of\\_Graph\\_Similarity\\_using\\_Maximum\\_Common\\_Edge\\_Subgraphs](https://www.researchgate.net/publication/220457922_RASCAL_Calculation_of_Graph_Similarity_using_Maximum_Common_Edge_Subgraphs).
9. Vishwanathan, S. V. N., Nicol, N. Schraudolph, Risi, Kondor, and Karsten, M. Borgwardt. Graph kernels. *Journal of Machine Learning Research*, 2010, no. 11 (Apr.), pp. 1201–1242. URL: <https://jmlr.org/papers/volume11/vishwanathan10a/vishwanathan10a.pdf>.
10. Nino, Shervashidze, Pascal Schweitzer, Erik Jan van Leeuwen, Kurt Mehlhom, and Karsten M Borgwardt. Weisfeiler-lehman graph kernels. *Journal of Machine Learning Research*, 2011, no. 12 (Sep.), pp. 2539–2561. URL: <https://jmlr.org/papers/volume12/shervashidze11a/shervashidze11a.pdf>.
11. Karsten, M. Borgwardt and Hans-Peter, Kriegel. Shortest-path kernels on graphs. *Fifth IEEE International Conference*, 2005, p. 8. URL: [https://www.researchgate.net/publication/4207736\\_Shortest-Path\\_Kernels\\_on\\_Graphs](https://www.researchgate.net/publication/4207736_Shortest-Path_Kernels_on_Graphs).
12. Hisashi, Kashima, Koji, Tsuda, and Akihiro, Inokuchi. Marginalized kernels between labeled graphs. *Proceedings of the 20th international conference on machine learning (ICML-03)*, 2003, pp. 321–328. URL: [https://www.researchgate.net/publication/221345917\\_Marginalized\\_Kernels\\_Between\\_Labeled\\_Graphs](https://www.researchgate.net/publication/221345917_Marginalized_Kernels_Between_Labeled_Graphs).
13. Nino, Shervashidze and Karsten, M. Borgwardt. Fast subtree kernels on graphs. *Advances in neural information processing systems*, 2009, pp. 1660–1668. URL: [https://www.researchgate.net/publication/221617867\\_Fast\\_subtree\\_kernels\\_on\\_graphs](https://www.researchgate.net/publication/221617867_Fast_subtree_kernels_on_graphs).
14. Weisfeiler, B. Y., & Leman, A. A. The reduction of a graph to canonical form and the algebra which appears therein. *Scientific and Technical Information*, 1968, no. 2 (9), pp. 12–16. URL: [https://www.itl.zcu.cz/wl2018/pdf/wl\\_paper\\_translation.pdf](https://www.itl.zcu.cz/wl2018/pdf/wl_paper_translation.pdf) (In Russ.).
15. Kipf, Thomas N., & Welling, Max. Semi-supervised classification with graph convolutional networks. *Fifth International Conference on Learning Representations (ICLR 2017)*, 2017. URL: <https://arxiv.org/abs/1609.02907>.
16. Hamilton, William L., Ying, Rex, & Leskovec, Jure. Inductive representation learning on large graphs. *Advances in Neural Information Processing Systems 30 (NIPS 2017)*, 2017. URL: <https://arxiv.org/abs/1706.02216>.
17. Xu, Keyulu, Hu, Weihua, Leskovec, Jure, & Jegelka, Stefanie. How powerful are graph neural networks? *Seventh International Conference on Learning Representations (ICLR 2019)*, 2019. URL: <https://arxiv.org/abs/1810.00826>.
18. Veličković, Petar, Cucurull, Guillem, Casanova, Arantxa, Romero, Adriana, Liò, Pietro, & Bengio, Yoshua. Graph attention networks. *Sixth International Conference on Learning Representations (ICLR 2018)*, 2018. URL: <https://typeset.io/papers/graph-attention-networks-1xed6mcmzg>.
19. Bresson, Xavier, & Laurent, Thomas. Residual gated graph ConvNets. *arXiv preprint arXiv:1711.07553*, 2017. URL: <https://arxiv.org/abs/1711.07553>.
20. Monti, Federico, Boscaini, Davide, Masci, Jonathan, Rodolà, Emanuele, Svoboda, Jan, & Bronstein, Michael M. Geometric deep learning on graphs and manifolds using mixture model cnns. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017. URL: <https://arxiv.org/abs/1611.08402>.
21. Seongmin, Ok. A graph similarity for deep learning. *Advances in Neural Information Processing Systems 33 (NeurIPS 2020)*, 2020. URL: <https://paperswithcode.com/paper/a-graph-similarity-for-deep-learning>.
22. Ben, Taskar, Ming-Fai, Wong, Pieter, Abbeel, Daphne, Koller. Link Prediction in Relational Data. *Advances in Neural Information Processing Systems 16 [Neural Information Processing Systems, NIPS 2003, December 8–13, 2003, Vancouver and Whistler, British Columbia, Canada]*. URL: [https://www.researchgate.net/publication/221618061\\_Link\\_Prediction\\_in\\_Relational\\_Data](https://www.researchgate.net/publication/221618061_Link_Prediction_in_Relational_Data).

23. Taskar, B., Segal, E. and Koller, D. Probabilistic classification and clustering in relational data. *Proc. IJCAI*, 2001, pp. 870–876. URL: [https://www.researchgate.net/publication/220812979\\_Probabilistic\\_Classification\\_and\\_Clustering\\_in\\_Relational\\_Data](https://www.researchgate.net/publication/220812979_Probabilistic_Classification_and_Clustering_in_Relational_Data).
24. Kai, Yu, Wei, Chu, Shipeng, Yu, Volker, Tresp, Zhao, Xu. Stochastic Relational Models for Discriminative Link Prediction. *Advances in Neural Information Processing Systems*, 2007, no. 19, pp.1553–1560. URL: [https://www.researchgate.net/publication/371869717\\_Stochastic\\_Relational\\_Models\\_for\\_Discriminative\\_Link\\_Prediction](https://www.researchgate.net/publication/371869717_Stochastic_Relational_Models_for_Discriminative_Link_Prediction).
25. Kemp, C., Tenenbaum, J. B., Griffiths, T. L., Yamada, T. and Ueda, N. Learning systems of concepts with an infinite relational model. *Proceedings of the 21st National Conference on Artificial Intelligence (AAAI)*, 2006. URL: [https://www.researchgate.net/publication/215991886\\_Learning\\_Systems\\_of\\_Concepts\\_with\\_an\\_Infinite\\_Relational\\_Model/](https://www.researchgate.net/publication/215991886_Learning_Systems_of_Concepts_with_an_Infinite_Relational_Model/).
26. Xu, Z., Tresp, V., Yu, K. and Kriegel, H.-P. Infinite hidden relational models. *Proceedings of the 22nd International Conference on Uncertainty in Artificial Intelligence (UAI)*, 2006. URL: <https://arxiv.org/abs/1206.6864>.
27. Chu, W., Sindhvani, V., Ghahramani, Z. and Keerthi, S. S. Relational learning with gaussian processes. *Advances in Neural Information Processing Systems 19, Proceedings of the Twentieth Annual Conference on Neural Information Processing Systems, Vancouver, British Columbia, Canada, December 4–7, 2006*. URL: [https://www.researchgate.net/publication/221620080\\_Relational\\_Learning\\_with\\_Gaussian\\_Processes](https://www.researchgate.net/publication/221620080_Relational_Learning_with_Gaussian_Processes).
28. Liben-Nowell, D. and Kleinberg, J. The link-prediction problem for social networks. *Journal of the American society for Information Science and Technology*, 2007, no. 58 (7), pp. 1019–1031. URL: <https://www.sci-hub.ru/10.1002/asi.20591>.
29. Sarkar, P., Chakrabarti, D. and Moore, A. W. Theoretical justification of popular link prediction heuristics. *International Conference on Learning Theory (COLT)*, 2010, pp. 295–307. URL: [https://www.researchgate.net/publication/220815677\\_Theoretical\\_Justification\\_of\\_Popular\\_Link\\_Prediction\\_Heuristics](https://www.researchgate.net/publication/220815677_Theoretical_Justification_of_Popular_Link_Prediction_Heuristics).
30. Koren, Y. Factorization meets the neighborhood: a multifaceted collaborative filtering model. *Proceeding of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*, 2008, pp. 426–434.
31. Bryan, Perozzi, Rami, Al-Rfou, and Steven, Skiena. Deepwalk: Online learning of social representations. *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*, 2014, pp. 701–710. URL: <https://arxiv.org/abs/1403.6652v2>.
32. Jian, Tang, Meng, Qu, Mingzhe, Wang, Ming, Zhang, Jun, Yan, and Qiaozhu, Mei. Line: Large-scale information network embedding. *Proceedings of the 24th International Conference on World Wide Web*, 2015, pp. 1067–1077. URL: <https://paperswithcode.com/paper/line-large-scale-information-network>.
33. Aditya, Grover and Jure, Leskovec. node2vec: Scalable feature learning for networks. *Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining*, 2016, pp. 855–864. URL: <https://arxiv.org/abs/1607.00653>.
34. Antoine, Bordes, Nicolas, Usunier, Alberto, Garcia-Duran, Jason, Weston, and Oksana, Yakhnenko. Translating embeddings for modeling multi-relational data. *Advances in Neural Information Processing Systems*, 2013, pp. 1–9. URL: <https://paperswithcode.com/paper/translating-embeddings-for-modeling-multi>.
35. Bishan, Yang, Wen-tau, Yih, Xiaodong, He, Jianfeng, Gao, and Li, Deng. Embedding entities and relations for learning and inference in knowledge bases. *International Conference on Learning Representations*, 2015. URL: <https://arxiv.org/abs/1412.6575>.
36. Zhiqing, Sun, Zhi-Hong, Deng, Jian-Yun, Nie, and Jian, Tang. Rotate: Knowledge graph embedding by relational rotation in complex space. *International Conference on Learning Representations*, 2019. URL: <https://arxiv.org/abs/1902.10197v1>.
37. Jiahao, Liu, Dongsheng, Li, Hansu, Gu, Tun, Lu, Peng, Zhang, Ning, Gu. *Parameter-free Dynamic Graph Embedding for Link Prediction*, 2022. URL: [https://www.researchgate.net/publication/364438325\\_Parameter-free\\_Dynamic\\_Graph\\_Embedding\\_for\\_Link\\_Prediction](https://www.researchgate.net/publication/364438325_Parameter-free_Dynamic_Graph_Embedding_for_Link_Prediction).
38. Farimah, Poursafaei, Shenyang, Huang, Kellin, Pelrine, Reihaneh, Rabbany. *Towards Better Evaluation for Dynamic Link Prediction*, 2022. URL: <https://arxiv.org/abs/2207.10128>.

Статья поступила в редакцию 15.05.2024; одобрена после рецензирования 10.06.2024; принята к публикации 18.06.2024.

The article was submitted 15.05.2024; approved after reviewing 10.06.2024; accepted for publication 18.06.2024.

УДК 004.021

## КЛАССИФИКАЦИЯ МЕТОДОВ РАСПОЗНАВАНИЯ КОНСТРУКТИВНЫХ ЭЛЕМЕНТОВ НА ЧЕРТЕЖАХ ТЕХНИЧЕСКИХ УСТРОЙСТВ

*Глазунов Тимур Алексеевич*, Волгоградский государственный технический университет, Российская Федерация, 400005, Волгоград, пр. им. В. И. Ленина, 28,

магистрант, ORCID: 0009-0008-3447-7663, e-mail: timurglazunov@mail.ru

*Соколов Александр Александрович*, АО «ВНИКТИнефтехимоборудование», Российская Федерация, 400078, Волгоград, пр. имени В. И. Ленина, 98Б, главный,

специалист лаборатории моделирования и анализа данных, ORCID: 0000-0003-3895-0522, e-mail: alexander.sokolov.it@gmail.com

*Маликов Виталий Павлович*, Волгоградский государственный технический университет, Российская Федерация, 400005, Волгоград, пр. им. В. И. Ленина, 28,

младший научный сотрудник, ORCID: 0000-0002-5053-2318, e-mail: axalter20@gmail.com

На сегодняшний день в нефтеперерабатывающей промышленности большое количество технических устройств находятся в эксплуатации с пятидесятих годов прошлого века. Данные о них до сих пор не существуют в цифровом виде, что замедляет работу по их дальнейшему обслуживанию и ремонту. Ручная оцифровка первичных данных технических устройств в кратчайшие сроки требует большого числа компетентных в работе с системами САПР сотрудников. В силу их отсутствия необходимо обобщить данные об конструктивных элементах на чертежах технических устройств и использовать их в качестве признаков для методов распознавания образов на изображениях. В этой статье представлен обзор существующих методов распознавания образов на изображениях, выявлены их особенности, области применения, а также их применение в решении задач чтения наборов измерений на чертежах технических устройств. Методы распознавания образов на изображении в данной статье разделены на три группы: традиционные методы, работающие напрямую с цветовыми свойствами изображения; методы машинного обучения, масштабирующие традиционные методы и позволяющие выделять и классифицировать на изображениях конкретные объекты; нейросетевые методы, работающие с отдельными областями изображения и позволяющие распознавать множество объектов на одном изображении. Цель статьи – систематизировать методы, которые можно использовать для распознавания конструктивных элементов на чертежах технических устройств, и предложить рекомендации для разработчиков метода интеллектуального распознавания конструктивных элементов технических устройств и сооружений на чертежах.

**Ключевые слова:** компьютерное зрение, распознавание образов, обнаружение образов, методы машинного обучения, методы глубокого обучения

## CLASSIFICATION OF METHODS FOR RECOGNIZING STRUCTURAL ELEMENTS IN DRAWINGS OF TECHNICAL DEVICES

*Glazunov Timur A.*, Volgograd State Technical University, 28 Lenin Ave., Volgograd, 400005, Russian Federation,

undergraduate student, e-mail: timurglazunov@mail.ru

*Sokolov Alexander A.*, JSC "VNIKTIneftekhimoborudovanie", main, 98B Lenin Ave., Volgograd, 400078, Russian Federation,

expert of the Laboratory of modeling and data analysis, e-mail: alexander.sokolov.it@gmail.com

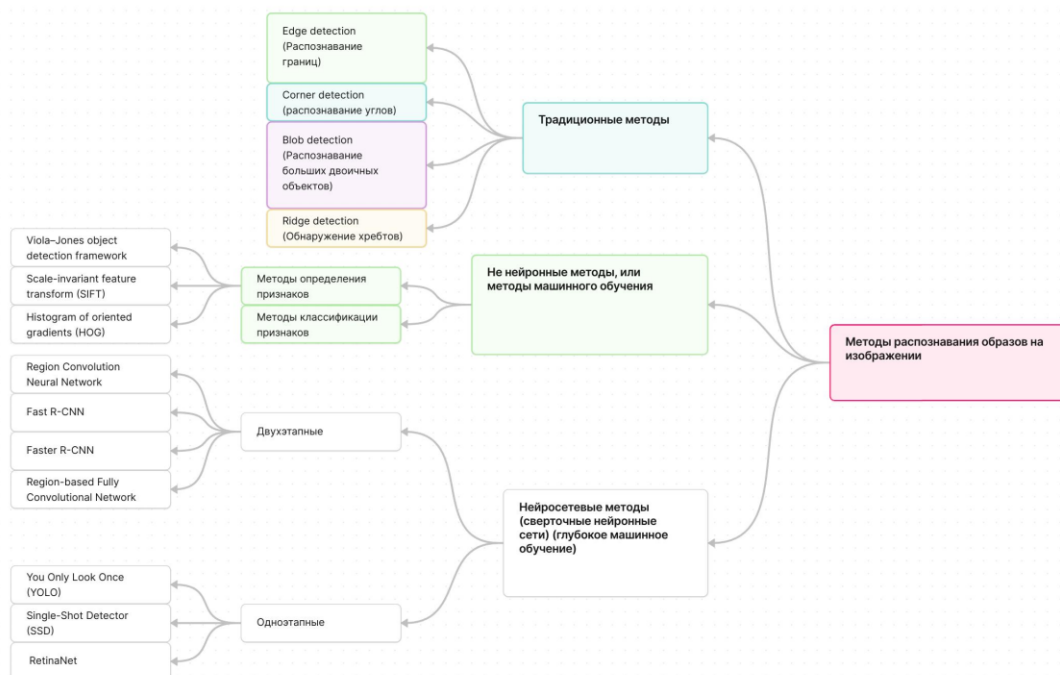
*Malikov Vitaly P.*, Volgograd State Technical University, 28 Lenin Ave., Volgograd, 400005, Russian Federation,

Junior Researcher, e-mail: axalter20@gmail.com

The oil refining industry have a large number of technical devices nowadays which are in operation since the fifties of the last century. Data of those still does not exist in digital form, that slows down the work on further maintenance and repair. Manual digitization of the primary data of technical devices as fast as possible requires a large number of employees who are competent in working with CAD systems. It is necessary to summarize the data on structural elements in the drawings of technical devices and use it as features for image recognition methods due to absence of employees. This article provides an overview of existing image recognition methods, identifies their features, areas of application, as well as their application in solving problems of reading measurement sets on drawings of technical devices. Image recognition methods in this article are divided into three groups: traditional methods that work directly with the features of an image; machine learning methods that scale traditional methods and allow you to identify and classify specific objects in images; neural network methods that work with individual areas of the image and allow you to recognize many objects in one image. The aim of the article is to systematize methods that can be used to recognize structural elements in drawings of technical devices, and to offer recommendations for developers of the method of intelligent recognition of structural elements of technical devices and structures in drawings.

**Keywords:** computer vision, pattern recognition, pattern detection, machine learning methods, deep learning methods

Graphical annotation (Графическая аннотация)



**ВВЕДЕНИЕ**

Технологии компьютерного зрения успешно применяются в таких важных отраслях экономики, как здравоохранение, логистика, заводская промышленность, кибербезопасность и финансы. Причиной тому стала необходимость обрабатывать огромные массивы визуальных данных для эффективного решения тех задач, где физических возможностей человека недостаточно. Особенно актуально это в тех задачах, где данных много, их качество варьируется от приемлемых до плохо читаемых, а извлечение информации является приоритетной бизнес-задачей. Распознавание образов на изображении является на сегодняшний день обширной и достаточно абстрактной темой. В разных контекстах [1] она имеет свою специфику. В сфере здравоохранения, например, для выявления злокачественных опухолей будет актуальна работа с цветовыми оттенками рентгеновских изображений, в то время как при оцифровке паспортов нефтеперерабатывающих заводов будет акцентировано внимание на геометрических свойствах искомым на чертежах конструктивных элементов. Для решения последней задачи в данной статье будут рассмотрены существующие методы распознавания образов на изображении.

Конструктивные элементы оборудования в контексте рассматриваемой задачи – это те составные элементы технического устройства нефтеперерабатывающего завода, которые позволяют собрать устройство, установить его на рабочей позиции и обеспечить коммуникацию с другими техническими устройствами. К техническим устройствам в рамках задачи относятся тысячи сосудов и теплообменников, построенных еще в пятидесятых годах прошлого века и чьи чертежи не существуют в цифровом виде до сих пор. Для приведения чертежей в цифровой вид необходимо большое количество специалистов, обладающих соответствующими компетенциями в обращении с САПР, таких как паоСАД. Организация, для которой разрабатывается метод интеллектуального распознавания конструктивных элементов технических устройств, не располагает такими ресурсами, поэтому задача по оцифровке чертежей ложится на большое число сотрудников, обладающих компетенциями в разных областях. Для оцифровки данных сотрудникам необходимо располагать технической информацией о конструктивных элементах технического устройства. К таким элементам можно отнести фланцевые соединения, штуцера, обечайки, змеевики и др. Каждый такой элемент в зависимости от устройства обладает своим набором технических характеристик, которые возможно перенести в САПР, однако в условиях достаточного количества специалистов. Как было ранее отмечено, у организации такого числа специалистов нет.

Конструктивные элементы, с одной стороны, являются набором примитивов, к распознаванию которых можно подойти с математической точки зрения, где использование методов машинного обучения будет избыточным, как, например, на рисунке 1.

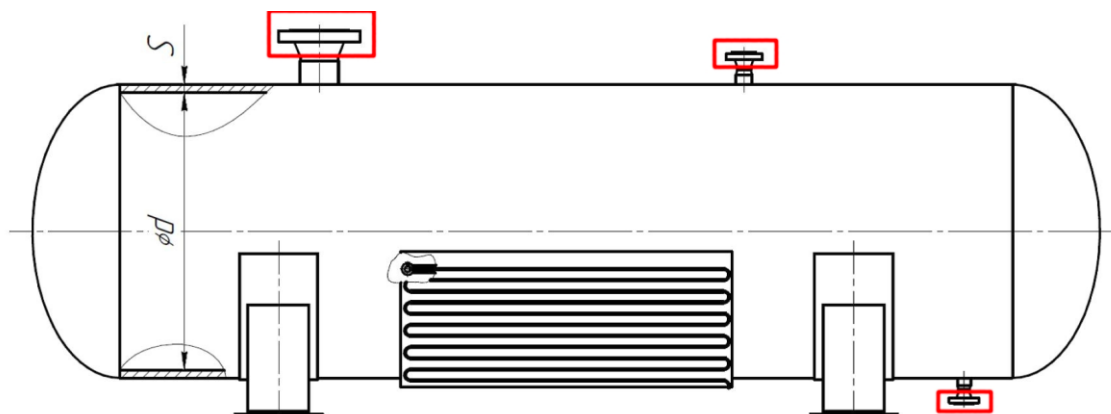


Рисунок 1 – Чертеж сосуда, где отмеченные фланцевые соединения изображены в виде двух прямоугольников

С другой стороны, как видно на рисунке 2, примитивы могут образовывать собой комплексные элементы, для распознавания которых будет труднее описать математическую модель.

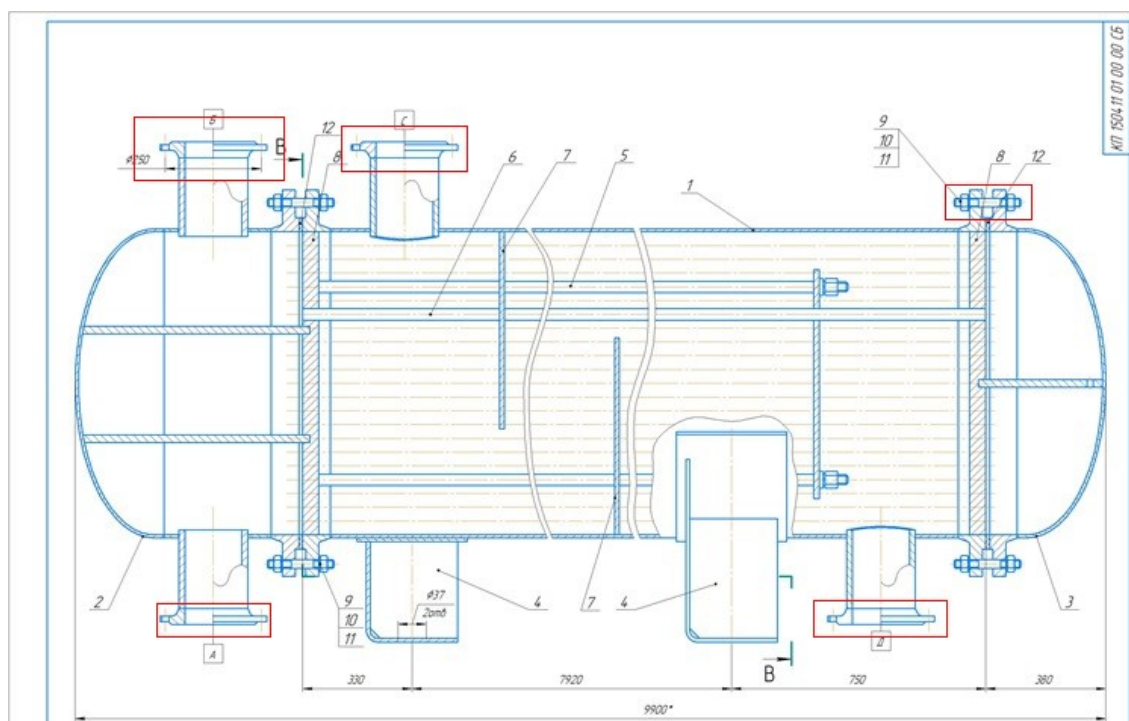


Рисунок 2 – Чертеж теплообменника с отмеченными фланцевыми соединениями

Ее точность будет неизбежно падать в случаях, если будут встречаться чертежи с одними и теми же элементами, но изображенными в разных чертежах по-разному. К тому же, если будет стоять задача классифицировать несколько элементов на чертеже, то для каждого такого элемента будет необходимо описывать математическую модель или приводить к стандартизированному виду через программы САПР, такие как nanoCAD, на что уйдет много времени. Эффективнее в данном случае обобщить данные об элементах на чертежах и использовать их в качестве признаков для распознавания образов на изображении. Чтобы выбрать подход к распознаванию конструктивных элементов на чертежах технических устройств, необходимо рассмотреть известные подходы к обнаружению объектов в компьютерном зрении.

Существующие методы обнаружения образов можно разделить на три группы:

- традиционные методы распознавания, или методы обработки изображений, решающие задачу распознавания объектов (англ. Object Recognition);
- методы машинного обучения, или нейросетевые методы, включающие в себя задачи выявления признаков и их дальнейшую классификацию;

– методы глубокого обучения, или нейронные методы, позволяющие проводить обнаружение объектов (англ. Object Detection) с использованием соответствующих алгоритмов, чтобы классифицировать объекты.

В статье будут кратко рассмотрены наиболее популярные методы в каждой из групп, а после будут прокомментированы работы, где данные методы использовались в задачах обнаружения элементов на технических чертежах.

### **ТРАДИЦИОННЫЕ МЕТОДЫ**

Традиционные методы основаны на алгоритмах, способных описать признаки изображения с помощью математической модели. Стоит отметить, что на сегодняшний день данные методы чаще всего используются для предварительной обработки изображений с целью устранения шумов, повышения контрастности, выделения контуров. Это позволяет разделить данные подходы на следующие группы:

– методы обнаружения краев, или границ (Edge detection) – данное понятие включает в себя множество математических методов, направленных на идентификацию границ, определяемых как кривые на цифровом изображении, на которых яркость изображения резко меняется или имеет разрывы. К наиболее популярным методам можно отнести детектор Кенни [2].

– методы обнаружения углов (Corner Detection) – это подход, используемый в системах компьютерного зрения для выделения углов – точек, где пересекаются границы. То есть, где происходит резкий переход из одной части изображения в другую. Распознавание углов часто используется при обнаружении движения, регистрации изображений, отслеживании видео и т. п. Наиболее популярным можно назвать детекторы Харриса и Стивенсона [3].

– обнаружение больших двоичных объектов – методы обнаружения больших объектов направлены на обнаружение областей цифрового изображения, которые отличаются по своим свойствам, таким как яркость или цвет, по сравнению с окружающими областями. Разность по Гауссу является одним из методов для выделения таких областей.

– обнаружение гребня [4] – с помощью программного обеспечения осуществляется поиск местоположения гребней на изображении, определяемых как кривые, точки которых являются локальными максимумами функции, сродни географическим гребням.

Перечисленные методы активно использовались, когда область компьютерного зрения только начинала свое развитие. В рамках рассматриваемой задачи оцифровки данных с бумажных чертежей технических устройств их применение необходимо в целях предварительной обработки изображения для устранения шумов и повышения качества чертежного рисунка. Изображения тысячи чертежей представляют собой фотографии, сделанные работниками завода на камеры смартфонов в разных условиях, от чего качество изображения чертежного рисунка варьируется от четкого до трудно читаемого, на что сотрудник организации, занимающийся оцифровкой данных, прикладывает дополнительные усилия в своей работе.

### **МЕТОДЫ МАШИНОГО, НЕНЕЙРОСЕТЕВОГО ОБУЧЕНИЯ**

Данные методы не относятся к области глубокого обучения, а являются переходными от традиционных к более автоматизированным. Они применяются в задаче определения признаков на изображении, которые затем классифицируются такими методами, как метод опорных векторов (англ. support vector machine (SVM)).

К существующим методам обнаружения образов на изображении можно отнести следующие примеры:

– Viola-Jones object detection framework – это алгоритм обнаружения объектов машинного обучения, предложенный в 2001 г. Полом Виолой и Майклом Джонсом. Он был разработан в первую очередь для решения проблемы распознавания лиц, где показывает себя наилучшим образом, хотя может быть адаптирован для обнаружения других классов объектов;

– Scale-invariant feature transform (SIFT) [5] – это алгоритм компьютерного зрения для обнаружения, описания и сопоставления локальных особенностей на изображениях, изобретенный Дэвидом Лоу в 1999 г. Наибольшую эффективность показывает при поиске дубликатов и сопоставлении идентичных изображений. Меньшую эффективность он показывает при обобщении объектов;

– Histogram of oriented gradients (HOG) [6] – это дескриптор признака, используемый в компьютерном зрении и обработке изображений для обнаружения объектов. Данный метод позволяет подсчитывать случаи ориентации градиента в локализованных частях изображения. Метод аналогичен методу гистограмм ориентации краев, дескрипторов SIFT, не зависящих от масштаба и контекстов формы, но отличается тем, что он вычисляется на плотной сетке из равномерно расположенных ячеек и использует перекрывающую нормализацию локального контраста для повышения точности.

## МЕТОДЫ ГЛУБОКОГО, НЕЙРОСЕТЕВОГО ОБУЧЕНИЯ

Методы нейросетевого обучения делятся на двухэтапные и одноэтапные, но они все основаны на технологии сверточных нейронных сетей (англ. convolutional neural network, CNN). Создание сверточных сетей было вдохновлено биологическими процессами [7], в которых структура связей между нейронами напоминает организацию зрительной коры головного мозга. Отдельные нейроны коры реагируют на стимулы только в ограниченной области поля зрения, известной как рецептивное поле. Рецептивные поля различных нейронов частично перекрываются таким образом, что они охватывают все поле зрения. Название архитектура сети получила из-за наличия операции свёртки, суть которой в том, что каждый фрагмент изображения умножается на матрицу (ядро) свёртки поэлементно, а результат суммируется и записывается в аналогичную позицию выходного изображения.

Двухэтапные методы, как следует из названия, описывают решение распознавания изображений в два этапа. На первом этапе с помощью специального слоя нейронной сети выделяются области, с высокой вероятностью содержащие внутри себя объекты. На втором этапе выбранные регионы рассматриваются классификатором для определения принадлежности к исходным классам и регрессором, уточняющим местоположение ограничивающих рамок. К таким методам относятся [8]:

- R-CNN. Модель R-CNN сочетает в себе метод выборочного поиска для обнаружения предложений регионов и глубокое обучение для обнаружения объекта в этих регионах. Размер каждого предложения региона изменяется, чтобы соответствовать входным данным CNN. Вектор признаков передается в несколько классификаторов для получения вероятностей принадлежности к каждому классу;

- Fast R-CNN. Цель быстрой сверточной сети на основе регионов состоит в том, чтобы сократить затраты времени, связанные с большим количеством моделей, необходимых для анализа всех предложений регионов;

- Faster R-CNN. Предложения регионов, обнаруженные с помощью метода выборочного поиска, по-прежнему были необходимы в предыдущей модели, которая требовала значительных вычислительных ресурсов. В 2016 г. представили Сеть предложений по регионам (RPN) для прямого создания предложений по регионам, прогнозирования ограничивающих рамок и обнаружения объектов. Более быстрая сверточная сеть на основе регионов (Faster R-CNN) представляет собой комбинацию между RPN и моделью Fast R-CNN;

- Region-based Fully Convolutional Network (R-FCN). Методологии Fast и Faster R-CNN заключаются в обнаружении предложений регионов и распознавании объекта в каждом регионе. Региональная полностью сверточная сеть (R-FCN) представляет собой модель только со сверточными слоями, обеспечивающую полное обратное распространение для обучения и логического вывода. Авторы объединили два основных шага в одну модель, чтобы одновременно учитывать обнаружение объекта (инвариант местоположения) и его положение (вариант местоположения).

Одноэтапные методы не используют отдельный алгоритм для генерации регионов. Вместо этого в них предсказывают координаты определённого количества ограничивающих рамок с различными характеристиками, такими как результаты классификации и степень уверенности, и в дальнейшем корректируют местоположение этих рамок. К ним относятся [9]:

- You Only Look Once (YOLO) [10]. Алгоритм, широко используемый для обнаружения объектов на изображениях. Основное отличие YOLO от других алгоритмов сверточной нейронной сети (CNN), используемых для обнаружения объектов, заключается в том, что он очень быстро распознает объекты в режиме реального времени. Принцип работы YOLO заключается в том, что вначале вводят изображение и проводят через сверточную нейронную сеть только один раз. На сегодняшний день один из самых быстрых алгоритмов распознавания объектов на изображении;

- Single-Shot Detector (SSD) [11]. В данной модели используются преимущества архитектуры YOLO и добавляются новые для более эффективной работы нейросети. В качестве входных данных модель принимает изображение, которое проходит через несколько сверточных слоев с различными размерами фильтров. Карты объектов из сверточных слоев в разных положениях сети используются для прогнозирования ограничивающих рамок;

- RetinaNet [12]. Это мощная одноступенчатая модель обнаружения объектов, которая эффективно идентифицирует объекты на изображениях с высокой точностью, принимает на вход изображение и выделяет из него важные признаки. Состоит из 4 разных частей: Backbone, Feature Pyramid Net (FPN), Classification Subnet, Regression Subnet.

## ПРИМЕРЫ ИСПОЛЬЗОВАНИЯ МЕТОДОВ РАСПОЗНАВАНИЯ ОБРАЗОВ НА ЧЕРТЕЖАХ

Из статей на данную тематику можно выделить работу [13], где из чертежей извлекают текстовые данные путем решения двух задач: обнаружение образов измерительных обозначений и оптическое распознавание текста на них. Для обнаружения образов использовалась модель Faster-RCNN ResNet50, предоставляемая Python библиотекой PyTorch. Выбор метода Faster-RCNN обусловлен

хорошими показателями при работе с реальными изображениями [14]. После получения нужных сегментов они попадают в Keras-OCR модель, откуда извлекаются текстовые обозначения, в том числе и специфические символы, как, например, диаметры и радиусы. В качестве метрики использовалась оценка средней точности (mAP), так как она показывает, насколько хорошо классифицирован объект и локализован. Эксперимент проводился тремя способами: с обучением на реальных чертежах, с обучением на большой выборке с использованием генератора изображений и с использованием стандартной модели Keras-OCR. В результате эксперимента точность распознавания образов была наивысшей при использовании модели с большой обучающей выборкой, чуть меньшей при использовании стандартной модели Keras-OCR и самой низкой на реальных чертежах.

Похожие методы использовали [15] и для больших, комплексных чертежей. Автор акцентировал внимание на факте, что точность выявленных объектов при обработке изображений выше, чем и без нее.

В другой работе [16], где на чертежах распознавали наборы измерений, на этапе предварительной обработки изображения применяли специально разработанный метод Флетчера и Кастури [17], рассматривающий текст как множество отдельных сегментов изображения, которые необходимо отделить друг от друга. Далее на выявленных сегментах для обнаружения измерений анализировали изображение с точки зрения свойств тех примитивов, что составляют искомое измерение.

В другой работе [18] распознавание поверхностей на конструктивных элементах чертежа отнесли к задачам компьютерной обработки и анализа изображений, применяя методы машинного обучения, а именно классификатор случайных лесов, классификатор K-ближайших соседей и классификатор опорных векторов для распознавания фигур по таким признакам, как координаты точек, наклоны линий, площадь, периметр и центр тяжести.

В следующей работе автор расширяет возможности популярной модели Keras-OCR одноименной библиотеки на языке Python для выявления наборов измерений на чертежах [19]. К стандартному набору текстовых и числовых обозначений добавляются специальные знаки измерений, такие как длины с высотами, радиусы с диаметрами и др.

На фоне рассмотренных выше работ выделяется та [20], где кроме специальных текстовых обозначений американского инженерного стандарта распознают сами конструкторские элементы и используют для этого метод глубокого машинного обучения Yolo с помощью Python-реализации OpenCV. Задача распознавания образов, содержащих текстовые обозначения не изменилась, но в рассматриваемой работе она была выделена и рассмотрена в разделе обучения модели.

### **ЗАКЛЮЧЕНИЕ**

В статье были рассмотрены существующие подходы к обнаружению объектов на изображении и их применение на примере нескольких работ, исследующих задачу обнаружения текста и конструкторских элементов на чертежах.

Нахождение текста часто встречалось в рассмотренных работах и являлось основной решаемой задачей. Но она являлась продолжением задачи сегментации тех участков чертежей, где присутствуют замеры и специальные символы. Модели нейронных сетей обучались распознавать эти символы, но последняя работа, где применялся алгоритм Yolo, показала сегментацию сначала элементов, а уже затем замеров. Это говорит о том, что задача обнаружения конструкторских элементов может быть решена нейросетевыми методами, в том числе и методом Yolo.

Традиционные методы сегодня применяются в качестве инструментов предварительной обработки изображений с целью четче выделить искомые объекты на изображении. Использование этих методов для обнаружения объектов, как было рассмотрено в одной из работ [16], позволяет лишь приблизительно обозначить область, где находится искомый объект. Методы машинного обучения, не использующие нейросетевые технологии, постепенно вытесняются более эффективными нейросетевыми методами, разработанными специально для решения задач компьютерного зрения, в том числе для обнаружения объектов. В рассмотренных работах чаще всего применялась Python-библиотека OpenCV в связке с моделью Keras-OCR, Faster R-CNN и Yolo. Авторы, использовавшие в своей работе алгоритм Yolo [20] для распознавания наборов измерений на чертежах технических устройств, показали возможность классифицировать отдельные конструктивные элементы, дабы на них в дальнейшем распознавать наборы измерений. На основе этой работы будет разрабатываться метод для решения задачи распознавания конструкторских элементов на чертежах сосудов и аппаратов нефтеперерабатывающих заводов.

### **Список источников**

1. Forsyth, David. Computer vision: a modern approach. Pearson / David Forsyth, Jean Ponce. – 2012.
2. Canny, J. A computational approach to edge detection / J. Canny // IEEE Transactions on pattern analysis and machine intelligence. – 1986. – № 6. – С. 679–698.
3. Dey, N. A comparative study between Moravec and Harris corner detection of noisy images using adaptive wavelet thresholding technique / N. Dey et al. – 2012. – URL: arXiv preprint arXiv:1209.1558.

4. Damon, J. Properties of ridges and cores for two-dimensional images / J. Damon // *Journal of Mathematical Imaging and Vision*. – 1999. – Vol. 10. – P. 163–174.
5. Lowe, D. G. Object recognition from local scale-invariant features / D. G. Lowe // *Proceedings of the seventh IEEE international conference on computer vision*. – IEEE, 1999. – Vol. 2. – P. 1150–1157.
6. Dalal, N. Histograms of oriented gradients for human detection / N. Dalal, B. Triggs // *2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05)*. – IEEE, 2005. – Vol. 1. – P. 886–893.
7. Matsugu, M. et al. Subject independent facial expression recognition with robust face detection using a convolutional neural network / M. Matsugu et al. // *Neural networks*. – 2003. – Vol. 16, № 5–6. – P. 555–559.
8. Распознавание образов с помощью искусственного интеллекта // Хабр. – URL: <https://habr.com/ru/articles/709432/> (дата обращения: 25.04.2024).
9. Задача нахождения объектов на изображении // Конспекты ИТМО. – URL: [https://neerc.ifmo.ru/wiki/index.php?title=Задача\\_нахождения\\_объектов\\_на\\_изображении](https://neerc.ifmo.ru/wiki/index.php?title=Задача_нахождения_объектов_на_изображении) (дата обращения: 25.04.2024).
10. Jiang, P. et al. A Review of Yolo algorithm developments / P. Jiang et al. // *Procedia Computer Science*. – 2022. – Vol. 199. – P. 1066–1073.
11. Liu, W. SSD: Single Shot MultiBox Detector. / W. Liu et al. // *Computer Vision – ECCV 2016. Lecture Notes in Computer Science*. – Springer, Cham., 2016. – Vol. 9905. – URL: [https://doi.org/10.1007/978-3-319-46448-0\\_2](https://doi.org/10.1007/978-3-319-46448-0_2).
12. Lin, T. Y. Focal loss for dense object detection / T. Y. Lin et al. // *Proceedings of the IEEE international conference on computer vision*. – 2017. – P. 2980–2988.
13. Schlagenhauf, T. Text detection on technical drawings for the digitization of brown-field processes / T. Schlagenhauf, M. Netzer, J. Hillinger // *Procedia CIRP*. – 2023. – Vol. 118. – P. 372–377.
14. Ren, S. Faster r-cnn: Towards real-time object detection with region proposal networks / S. Ren et al. // *Advances in neural information processing systems*. – 2015. – Vol. 28.
15. Nguyen, T. Object Detection and Text Recognition in Large-scale Technical Drawings / T. Nguyen, L. Van Pham, C. Nguyen, and V. Van Nguyen // *Proceedings of the 10th International Conference on Pattern Recognition Applications and Methods – ICPRAM*. – 2021. – P. 612–619. – DOI: 10.5220/0010314406120619.
16. Lai, C. P. Detection of dimension sets in engineering drawings / C. P. Lai, R. Kasturi // *IEEE transactions on pattern analysis and machine intelligence*. – 1994. – Vol. 16, no. 8. – P. 848–855.
17. Fletcher, L. A. A robust algorithm for text string separation from mixed text/graphics images / L. A. Fletcher, R. Kasturi // *IEEE transactions on pattern analysis and machine intelligence*. – 1988. – Vol. 10, no. 6. – P. 910–918.
18. Касимов, Д. П. Автоматическая идентификация и описание поверхностей на 2D чертежах / Д. П. Касимов, М. Р. Королева, А. А. Чернова // *Графикон-конференции по компьютерной графике и зрению*. – 2023. – Т. 33. – С. 881–890.
19. Villena Toro J., Optical character recognition on engineering drawings to achieve automation in production quality control / Toro J. Villena, A. Wiberg, M. Tarkian // *Frontiers in Manufacturing Technology*. – 2023. – Vol. 3. – P. 1154132.
20. Lin, Y. H. Integration of Deep Learning for Automatic Recognition of 2D Engineering Drawings / Y. H. Lin et al. // *Machines*. – 2023. – Vol. 11, no. 8. – P. 802.

#### References

1. Forsyth, David, Ponce, Jean. *Computer vision: a modern approach*. Pearson. 2012.
2. Canny, J. A computational approach to edge detection. *IEEE Transactions on pattern analysis and machine intelligence*, 1986, no. 6, pp. 679–698.
3. Dey, N., Nandi, P., Barman, N., Das, D., & Chakraborty, S. (2012). A comparative study between Moravec and Harris corner detection of noisy images using adaptive wavelet thresholding technique. *arXiv preprint arXiv:1209.1558*.
4. Damon, J. Properties of ridges and cores for two-dimensional images. *Journal of Mathematical Imaging and Vision*, 1999, no. 10, pp. 163–174.
5. Lowe, D. G. Object recognition from local scale-invariant features. *Proceedings of the seventh IEEE international conference on computer vision*. IEEE, 1999, Sept., vol. 2, pp. 1150–1157.
6. Dalal, N., & Triggs, B. Histograms of oriented gradients for human detection. *2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05)*. IEEE, 2005, June, vol. 1, pp. 886–893.
7. Matsugu, M., Mori, K., Mitari, Y., & Kaneda, Y. Subject independent facial expression recognition with robust face detection using a convolutional neural network. *Neural networks*, 2003, no. 16 (5–6), pp. 555–559.
8. Pattern recognition using artificial intelligence. *Habr*. URL: <https://habr.com/ru/articles/709432/> (accessed 25.04.2024) (In Russ.).
9. The problem of finding objects in an image. *ITMO Notes*. URL: [https://neerc.ifmo.ru/wiki/index.php?title=Задача\\_нахождения\\_объектов\\_на\\_изображении](https://neerc.ifmo.ru/wiki/index.php?title=Задача_нахождения_объектов_на_изображении) (accessed 25.04.2024) (In Russ.).
10. Jiang, P., Ergu, D., Liu, F., Cai, Y., & Ma, B. A Review of Yolo algorithm developments. *Procedia Computer Science*, 2022, vol. 199, pp. 1066–1073.
11. Liu, W. et al. SSD: Single Shot MultiBox Detector. *Computer Vision – ECCV 2016. Lecture Notes in Computer Science*. Springer, Cham., 2016, vol. 9905. URL: [https://doi.org/10.1007/978-3-319-46448-0\\_2](https://doi.org/10.1007/978-3-319-46448-0_2).
12. Lin, T. Y. et al. Focal loss for dense object detection. *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2980–2988.

13. Schlagenhaut, T., Netzer, M., Hillinger, J. Text detection on technical drawings for the digitization of brown-field processes. *Procedia CIRP*, 2023, vol. 118, pp. 372–377.
14. Ren, S. et al. Faster r-cnn: Towards real-time object detection with region proposal networks. *Advances in neural information processing systems*, 2015, vol. 28.
15. Nguyen, T., Pham, L. Van, Nguyen C. and Nguyen, V. Van. Object Detection and Text Recognition in Large-scale Technical Drawings. *Proceedings of the 10th International Conference on Pattern Recognition Applications and Methods – ICPRAM*, 2021, pp. 612–619. DOI: 10.5220/0010314406120619.
16. Lai, C. P., Kasturi, R. Detection of dimension sets in engineering drawings. *IEEE transactions on pattern analysis and machine Intelligence*, 1994, vol. 16, no. 8, pp. 848–855.
17. Fletcher, L. A., Kasturi, R. A robust algorithm for text string separation from mixed text/graphics images. *IEEE transactions on pattern analysis and machine intelligence*, 1988, vol. 10, no. 6, pp. 910–918.
18. Kasimov, D. R., Koroleva, M. R., & Chernova, A. A. Automatic Identification and Description of Surfaces in 2D Drawings. *Graphicon-Conference on Computer Graphics and Vision*, 2023, vol. 33, pp. 881–890 (In Russ.).
19. Villena, Toro, J., Wiberg, A., & Tarkian, M. Optical character recognition on engineering drawings to achieve automation in production quality control. *Frontiers in Manufacturing Technology*, 2023, vol. 3, p. 1154132.
20. Lin, Y. H., Ting, Y. H., Huang, Y. C., Cheng, K. L., & Jong, W. R. Integration of Deep Learning for Automatic Recognition of 2D Engineering Drawings. *Machines*, 2023, vol. 11, no. 8, p. 802.

Статья поступила в редакцию 19.05.2024; одобрена после рецензирования 24.06.2024; принята к публикации 01.07.2024.

The article was submitted 19.05.2024; approved after reviewing 24.06.2024; accepted for publication 01.07.2024.

УДК 004.89

## СРАВНИТЕЛЬНЫЙ АНАЛИЗ МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ С ПРИМЕНЕНИЕМ МЕТОДОВ СНИЖЕНИЯ РАЗМЕРНОСТИ ДЛЯ КЛАССИФИКАЦИИ ЖЕСТОВ РУК

**Захарова Оксана Игоревна**, Поволжский государственный университет телекоммуникаций и информатики, 443090, Российская Федерация, г. Самара, ул. Московское шоссе, 77, кандидат технических наук, доцент, ORCID: 0003\_3371\_4344, e-mail: o.zaharova@psuti.ru  
**Бедняк Светлана Геннадьевна**, Поволжский государственный университет телекоммуникаций и информатики, 443090, Российская Федерация, г. Самара, ул. Московское шоссе, 77, кандидат педагогических наук, доцент, ORCID: 0001\_5678\_3196, e-mail: s.bednyak@psuti.ru  
**Шамрин Даниил Александрович**, Поволжский государственный университет телекоммуникаций и информатики, 443090, Российская Федерация, г. Самара, ул. Московское шоссе, 77, магистрант, ORCID: 0009\_0001\_1264\_4300, e-mail: tyvm14@mail.ru

Данная работа посвящена анализу различных методов машинного обучения, применяемых для классификации жестов рук с использованием техник предобработки данных и снижения размерности. В рамках исследования обучены и протестированы пять видов классификаторов: наивный байесовский классификатор, метод опорных векторов, случайный лес, метод k ближайших соседей и логистическая регрессия. Основное внимание уделяется сравнению их эффективности при решении задачи распознавания жестов рук. Цель исследования заключается в создании программы, которая сравнивает классификаторы машинного обучения для распознавания жестов рук, применяя методы предобработки и уменьшения размерности данных. В работе используются методы контролируемого и неконтролируемого обучения. В ходе эксперимента был подробно изучен набор данных, используемый для классификации различных жестов рук, и проанализированы результаты нескольких классификаторов машинного обучения для решения данной задачи. Для сравнения результатов на языке программирования Python было написано приложение с графическим пользовательским интерфейсом. Исследование продемонстрировало эффективность различных классификаторов машинного обучения в задаче распознавания жестов рук, выявив сильные и слабые стороны каждого метода. Область применения результатов: разработанная программа и результаты могут быть использованы в системах распознавания жестов для различных приложений, включая интерфейсы человек-машина, виртуальную реальность и системы безопасности.

**Ключевые слова:** машинное обучение, снижение размерности, метод главных компонент, линейный дискриминантный анализ, классификация жестов рук

## COMPARATIVE ANALYSIS OF MACHINE LEARNING METHODS WITH APPLICATION OF SIZE REDUCTION METHODS FOR HAND GESTURE CLASSIFICATION

**Zaharova Oksana I.**, Povolzhskiy State University of Telecommunications and Informatics, 77 Moskovskoe shosse, Samara, 443090, Russian Federation, Candidate of Technical Sciences, Associate Professor, ORCID: 0003\_3371\_4344, e-mail: o.zaharova@psuti.ru  
**Bednyak Svetlana G.**, Povolzhskiy State University of Telecommunications and Informatics, 77 Moskovskoe shosse, Samara, 443090, Russian Federation, Cand. Sci. (Pedagogical), Associate Professor, ORCID: 0001\_5678\_3196, e-mail: s.bednyak@psuti.ru  
**Shamrin Daniil A.**, Povolzhskiy State University of Telecommunications and Informatics, 77 Moskovskoe shosse, Samara, 443090, Russian Federation, undergraduate student, ORCID: 0009\_0001\_1264\_4300, e-mail: tyvm14@mail.ru

Pattern recognition mainly focuses on recognizing patterns in data. In most cases, pattern recognition systems are trained using labeled training sample data (supervised learning or learning with a teacher). Various learning algorithms can also be used to find patterns in unlabeled data (unsupervised learning or learning without a teacher). The goal of the work is to create a program that performs a comparison of machine learning classifiers to solve the hand gesture recognition problem using preprocessing and dimensionality reduction techniques. The work will train and test five different types of classifiers, namely naive Bayesian classifier, support vector method, random forest, k nearest neighbors method, and logistic regression. During the experiment, the dataset used to classify different hand gestures was studied in detail and the results of non-few machine learning classifiers for this task were analyzed. To compare the results, an application with a graphical user interface was written in the Python programming language.

**Keywords:** machine learning, dimensionality reduction, principal component method, linear discriminant analysis, classification of hand gestures

## ВВЕДЕНИЕ

Современные системы распознавания образов ставят перед собой две основные цели: повышение скорости обработки и улучшение точности распознавания. Однако эти параметры часто противоречат друг другу: увеличение скорости может приводить к снижению точности и наоборот [7]. В 1961 г. Ричард Беллман назвал эту проблему «проклятием размерности» [5]. Для решения этой проблемы используются методы предобработки данных, которые включают удаление менее значимых признаков и тем самым сокращают время обучения и тестирования, одновременно повышая точность.

Для исследования был использован необработанный набор данных о жестах рук, взятый из репозитория машинного обучения UCI [9]. Экспериментальные данные были собраны от 12 пользователей, 9 из которых использовались для обучения модели, а 3 – для тестирования. Данные хранились в формате CSV и были преобразованы в массив для дальнейшей обработки.

Задачи исследования:

1. Реализовать программу для предобработки данных, конструирования признаков и снижения размерности с последующим обучением моделей машинного обучения.
2. Оценить эффективность использованных методов и провести сравнительный анализ результатов классификации.

## МЕТОДОЛОГИЯ

Общая схема алгоритма классификации [3] представлена ниже (рис. 1).

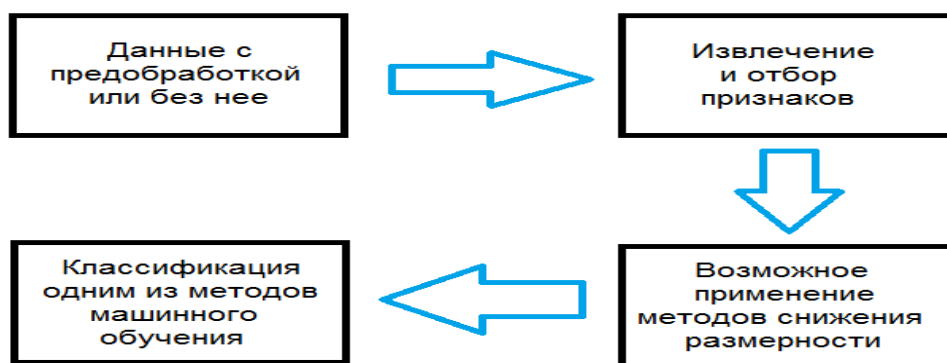


Рисунок 1 – Общая схема решения задачи

В процессе предобработки данных выполнялось извлечение признаков, среди которых: количество записанных маркеров, средние значения и стандартные отклонения по осям X, Y и Z, а также максимальные и минимальные значения координат маркеров.

Для выбора наилучших признаков использовался метод одномерного отбора признаков, а для предварительной обработки применялись стандартизация и нормализация данных. Уменьшение размерности осуществлялось методами главных компонент и линейного дискриминантного анализа Фишера. Кросс-валидация использовалась для предотвращения переобучения моделей.

## РЕЗУЛЬТАТЫ

В таблицах 1 и 2 приведены результаты классификации при использовании стандартизации и нормализации данных.

Таблица 1 – Сравнение точности классификаторов со стандартизацией и без сокращения размерности

Классификатор	Точность на обучающей выборке	Точность на тестовой выборке	F1-Score
Наивный байесовский классификатор	0,89088	0,80202	0,78998
Метод опорных векторов с RBF	0,99844	0,95781	0,95784
Случайный лес	1	0,68491	0,70052
Метод k ближайших соседей с k = 5	0,99896	0,79899	0,79384
LDA	0,9388	0,9272	0,9283
QDA	0,99	0,8490	0,8356
Логистическая регрессия	0,96088	0,9121	0,91125

Таблица 2 – Сравнение точности классификаторов с нормализацией и сокращением размерности

Классификатор	Точность на обучающей выборке	Точность на тестовой выборке	F1-Score
Наивный байесовский классификатор	0,78881	0,66813	0,62527
Метод опорных векторов с RBF	0,79807	0,56125	0,48786
Случайный лес	1	0,64458	0,64004
Метод k ближайших соседей с k = 5	0,99785	0,64723	0,63453
LDA	0,8968	0,8227	0,8223
QDA	0,9866	0,7166	0,6738
Логистическая регрессия	0,8377	0,6371	0,6159

На рисунках 2–4 показаны графики точности классификаторов при применении различных методов снижения размерности.

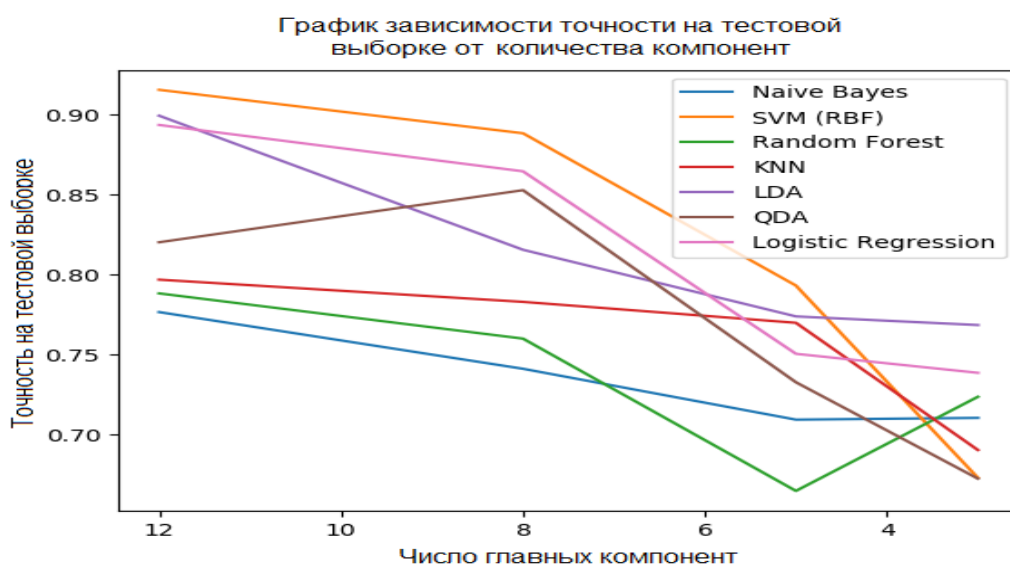


Рисунок 2 – График для PCA

Как следует из таблиц 1 и 2, точность классификации падает, если используется нормализация вместо стандартизации.

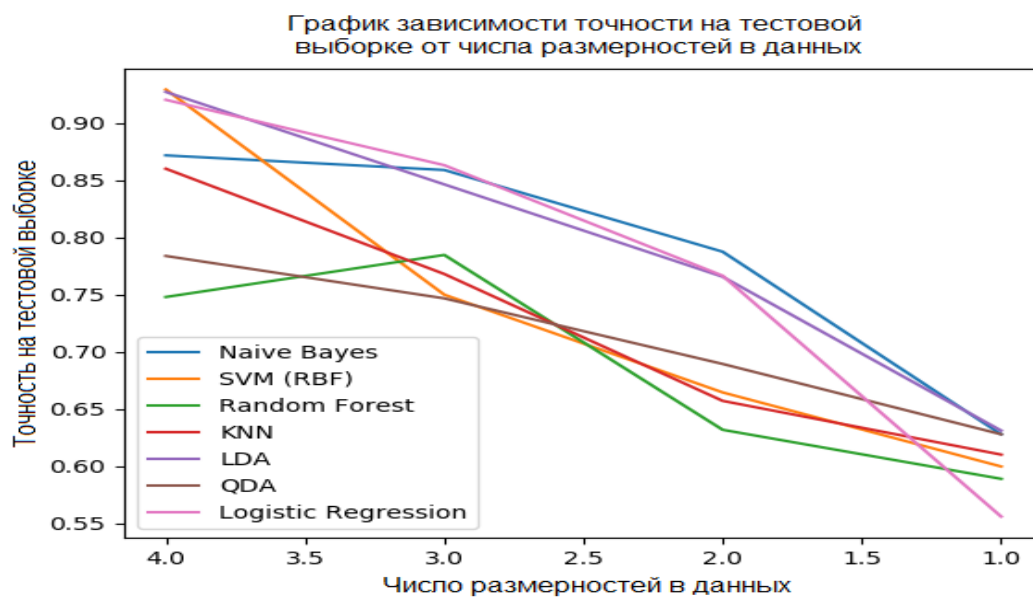


Рисунок 3 – График для FLD (LDA)

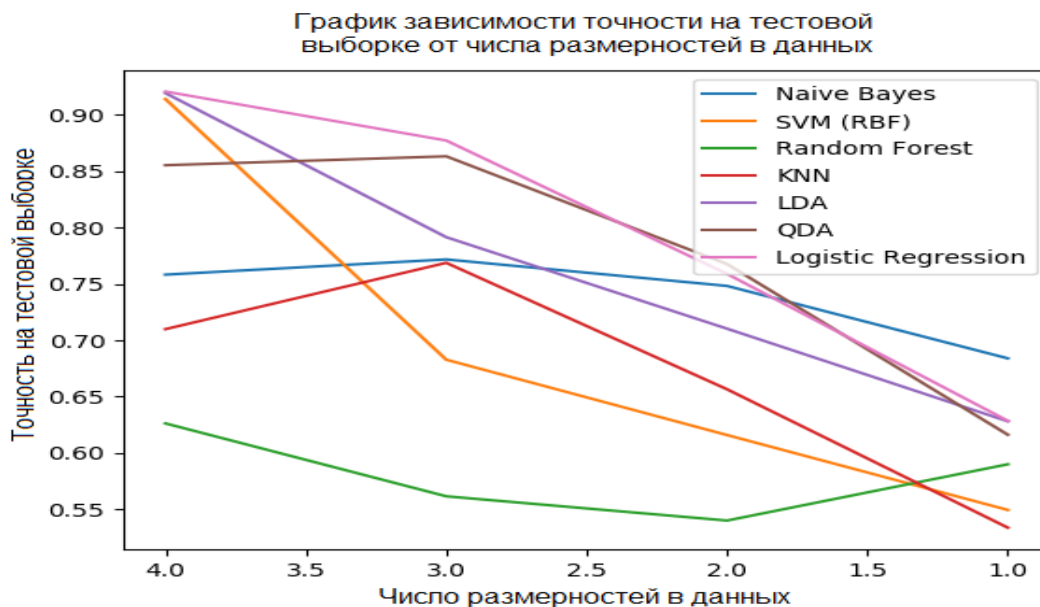


Рисунок 4 – График для одномерного отбора признаков

На основе проведенных экспериментов было установлено, что уменьшение числа признаков приводит к снижению точности классификации. Метод линейного дискриминантного анализа (LDA) показал наилучшие результаты в плане дискриминационной способности, что делает его оптимальным выбором для данного набора данных.

В таблице 3 приведены результаты сравнения.

Пример работы реализованного приложения, позволяющего выполнять сравнение классификаторов, представлен на рисунке 5.

Статья

**Распознавание положения руки на основе захвата движения**

**Результаты**

Классификатор	Точность обучающей выборки	Точность тестовой выборки	F-Score
Наивный байесовский классификатор	0.9378518518518518	0.7765296933503958	0.764293177291642
Метод опорных векторов (ядро - RBF)	0.9985185185185185	0.9156832077349637	0.915760060944161
Случайный лес	1.0	0.788236409308498	0.7776671780327827
Метод k ближайших соседей	0.9989629629629629	0.7968150149296175	0.791907846172628
Линейный дискриминантный анализ	0.9372592592592592	0.8994739087160529	0.8993767778091111
Квадратичный дискриминантный анализ	0.9843703703703703	0.8202284468458221	0.8109815674759522
Логистическая регрессия	0.9592592592592593	0.8936916441537514	0.8923454747926842

Главное меню

Рисунок 5 – Пример работы программы

Стандартизация данных, в сочетании с методом LDA, продемонстрировала наивысшую точность среди всех протестированных комбинаций. Среди классификаторов наилучшие результаты показали метод опорных векторов и логистическая регрессия, значительно превосходя другие модели. В то же время случайный лес продемонстрировал наихудшие результаты, вероятно, из-за недостаточной оптимизации параметров модели.

Итоговые результаты подтверждают, что выбор методов предобработки и снижения размерности играет ключевую роль в повышении точности классификации. Эти находки указывают на необходимость дальнейших исследований в области подбора наиболее значимых признаков и оптимизации параметров моделей для улучшения качества распознавания жестов.

Таблица 3 – Сравнение точности классификаторов со стандартизацией и FLD с 4 размерностями

Классификатор	Точность на обучающей выборке	Точность на тестовой выборке	F1-Score
Наивный байесовский классификатор	0,9521	0,8717	0,8744
Метод опорных векторов с RBF	0,9976	0,9293	0,9292
Случайный лес	1,0	0,7478	0,7149
Метод k ближайших соседей с k = 5	0,9979	0,8600	0,8588
LDA	0,9388	0,9272	0,9283
QDA	0,9571	0,7837	0,7553
Логистическая регрессия	0,9430	0,9203	0,9196

### ЗАКЛЮЧЕНИЕ

В данной работе проведено исследование различных методов машинного обучения и снижения размерности в контексте задачи классификации жестов рук. Результаты экспериментов показали, что стандартизация данных в сочетании с методом линейного дискриминантного анализа обеспечивает наибольшую точность классификации. Среди протестированных алгоритмов наибольшую эффективность продемонстрировали метод опорных векторов и логистическая регрессия.

Метод случайного леса показал сравнительно низкую точность, что может быть связано с неэффективной настройкой параметров модели.

Выводы исследования подтверждают важность тщательного выбора методов предобработки данных и алгоритмов снижения размерности для достижения высокой точности в задачах классификации. Для дальнейшего улучшения результатов можно рассмотреть добавление новых признаков, обладающих более высокой дискриминационной способностью, таких как расстояние маркеров до начальной точки координат и медианные значения координат.

Эти результаты подчеркивают важность дальнейших исследований в области оптимизации моделей машинного обучения и выбора наиболее информативных признаков для повышения точности распознавания жестов рук.

### Список источников

1. Бенгфорт, Б. Прикладной анализ текстовых данных на Python. Машинное обучение и создание приложений обработки естественного языка / Б. Бенгфорт. – Санкт-Петербург : Питер, 2019. – 368 с.
2. Мюллер, А. Введение в машинное обучение с помощью Python / А. Мюллер. – Москва : Альфа-книга, 2017.
3. Николенко, С. Глубокое обучение / С. Николенко, А. Кадурин, Е. Архангельская. – Санкт-Петербург : Питер, 2018. – 480 с.
4. Плас, Д. Python для сложных задач. Наука о данных и машинное обучение. Руководство / Плас Джейк Вандер. – Москва : Питер, 2018. – 759 с.
5. Рашка, С. Python и машинное обучение / С. Рашка. – ДМК Пресс, 2017. – 418 с.
6. Ясницкий, Л. Н. Интеллектуальные системы : учебник / Л. Н. Ясницкий. – Москва : Лаборатория знаний, 2016. – 221 с.
7. Goodfellow, I. Deep Learning / I. Goodfellow, Y. Bengio, A. Courville. – MIT Press, 2016.
8. LeCun, Y. Deep learning / Y. LeCun, Y. Bengio, G. Hinton // Nature. – 2015. – Vol. 521, no. 7553. – P. 436.
9. UCI Machine Learning Repository. – URL: <https://archive.ics.uci.edu/ml/index.php>.

### References

1. Bengfort, B. *Applied analysis of text data in Python. Machine learning and the creation of natural language processing applications*. St. Petersburg, Peter, 2019. 368 p. (In Russ.).
2. Muller, A. *Introduction to machine learning using Python*. Moscow, Alpha-book, 2017. (In Russ.).
3. Nikolenko, S., Kadurin, A., Arkhangelskaya, E. *Deep learning*. St. Petersburg, Peter, 2018. 480 p. (In Russ.).
4. Plas, D. *Python for complex tasks. Data science and machine learning. Manual*. Moscow, Peter, 2018. 759 p. (In Russ.).
5. Rashka, S. *Python and machine learning*. DMK Press, 2017. 418 p. (In Russ.).
6. Yasnitsky, L. N. *Intelligent systems : textbook*. Mosciow, Laboratory of Knowledge, 2016. 221 p. (In Russ.).
7. Goodfellow, I., Bengio, Y., Courville, A. *Deep Learning*. MIT Press, 2016.
8. LeCun, Y., Bengio, Y., Hinton, G. Deep learning. *Nature*, 2015, vol. 521, no. 7553, p. 436,
9. *UCI Machine Learning Repository*. URL: <https://archive.ics.uci.edu/ml/index.php>.

Статья поступила в редакцию 11.06.2024; одобрена после рецензирования 03.07.2024; принята к публикации 10.07.2024.

The article was submitted 11.06.2024; approved after reviewing 03.07.2024; accepted for publication 10.07.2024.

УДК 004.89

**ИНСТРУМЕНТЫ ПРОГНОЗИРОВАНИЯ ТЕХНОЛОГИЧЕСКОГО РАЗВИТИЯ  
НА ОСНОВЕ ДАННЫХ ИЗ ОТКРЫТЫХ ИСТОЧНИКОВ:  
СИСТЕМАТИЧЕСКОЕ ИССЛЕДОВАНИЕ РУССКОЯЗЫЧНЫХ ДОКУМЕНТОВ**

*Бондаренко Артём Геннадьевич*, Волгоградский государственный технический университет, 400005, Российская Федерация, г. Волгоград, проспект им. В. И. Ленина, 28, магистрант, ORCID: 0009-0004-4485-7169, e-mail: temdit01@yandex.ru

*Кравец Алла Григорьевна*, Волгоградский государственный технический университет, 400005, г. Волгоград, проспект им. В. И. Ленина, 28, доктор технических наук, профессор, ORCID: 0000-0003-1675-8652, e-mail: AllaGKravets@yandex.ru

Прогнозирование технологического развития приобрело решающее значение для процессов принятия стратегических решений на различных уровнях. Открытые данные становятся ценным ресурсом для анализа и прогнозирования технологических тенденций благодаря их объему и доступности. Прогнозирование технологического развития учитывает темпы и направление трендов, а также будущие изменения. Существует множество подходов к прогнозированию технологического развития, каждый со своими сильными и слабыми сторонами. Среди них выделяются методы экспертных оценок, математической статистики, интеллектуального анализа данных, стратегического и нормативного прогнозирования. Каждый из этих методов использует различные подходы и инструменты для анализа данных и предсказания будущих технологических трендов. В данной работе проводится систематический обзор 53 русскоязычных документов по методам прогнозирования технологического развития на основе данных из открытых источников за период 2014–2024 гг. Исследование охватывает анализ типов открытых данных, инструментов их обработки, а также ограничений и вызовов, существующих при использовании открытых данных для прогнозирования технологического развития. Интеллектуальный анализ данных определен как наиболее эффективный и востребованный метод прогнозирования технологического развития. Применение инструментов интеллектуального анализа данных, таких как машинное обучение и нейронные сети, позволяет анализировать большие объемы данных, выявлять сложные закономерности и строить точные прогнозные модели развития технологий.

**Ключевые слова:** прогнозирование технологического развития, открытые данные, интеллектуальный анализ данных, машинное обучение, нейронные сети, методы прогнозирования, системный обзор литературы

**OPEN DATA-BASED TOOLS FOR TECHNOLOGICAL DEVELOPMENT FORECASTING:  
SYSTEMATIC MAPPING STUDY OF RUSSIAN-LANGUAGE SOURCES**

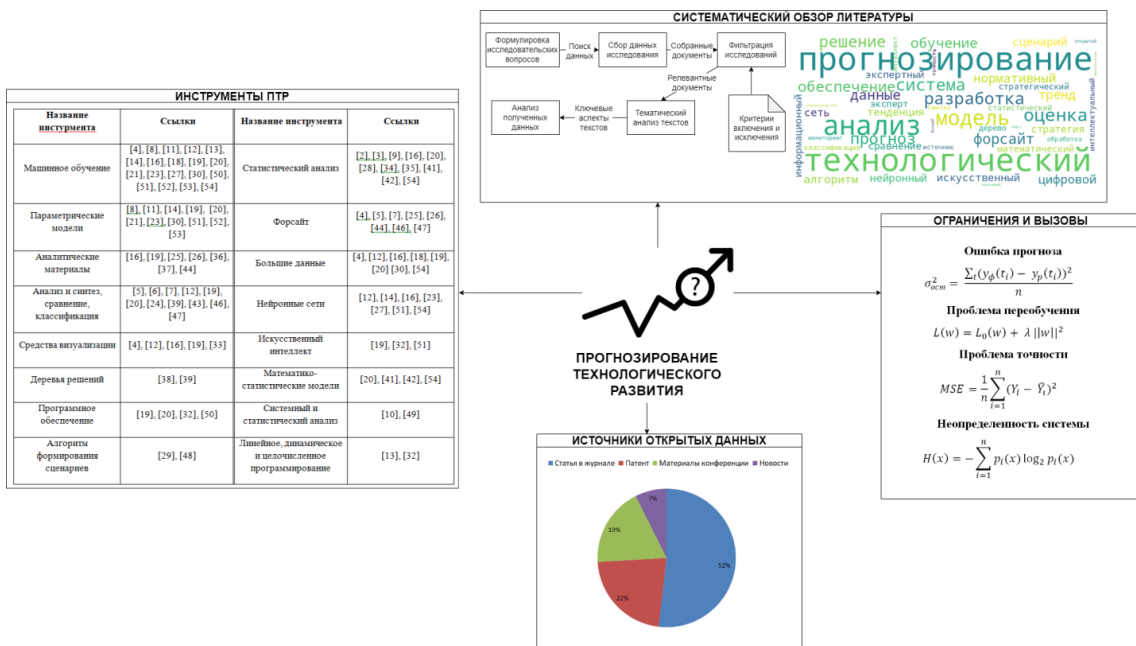
*Bondarenko Artem G.*, Volgograd State Technical University, 28 V. I. Lenin Ave., Volgograd, 400005, Russian Federation, master student, ORCID: 0009-0004-4485-7169, e-mail: temdit01@yandex.ru

*Kravets Alla G.*, Volgograd State Technical University, 28 V. I. Lenin Ave., Volgograd, 400005, Russian Federation, Doct. Sci. (Engineering), Professor, ORCID: 0000-0003-1675-8652, e-mail: AllaGKravets@yandex.ru

Technological development forecasting has become crucial for strategic decision-making processes at various levels. Open data is a valuable resource for analyzing and forecasting technology trends due to its volume and accessibility. Technological development forecasting takes into account the pace and direction of trends, as well as future changes. There are many approaches to technological development forecasting, each with its own strengths and weaknesses. Among them are methods of expert assessments, mathematical statistics, data mining, strategic and normative forecasting. Each of these methods uses different approaches and tools to analyze data and predict future technology trends. This paper conducts a systematic mapping study of 53 Russian-language documents on methods for technological development forecasting based on data from open sources for the period 2014–2024. The study covers an analysis of the types of open data, tools for processing them, as well as the limitations and challenges that exist in using open data to predict technological development. Data mining has been identified as the most effective and popular method for predicting technological development. The use of data mining tools, such as machine learning and neural networks, allows analyzing large volumes of data, identifying complex patterns and building accurate predictive models of technology development.

**Keywords:** technological development forecasting, open data, data mining, machine learning, neural networks, forecasting methods, systematic mapping study

## Graphical annotation (Графическая аннотация)



## ВВЕДЕНИЕ

Новые технологии стремительно развиваются и характеризуются быстрым техническим прогрессом, прогнозирование технологического развития приобретает все большее значение. Оно играет важную роль в принятии стратегических решений на разных уровнях – от государственного планирования до корпоративной стратегии. В этом случае открытые данные являются ценным ресурсом для анализа и прогнозирования развития технологий. Большой объем информации, находящейся в открытом доступе, открывает новые возможности для исследования технологического развития с помощью методов анализа данных [1].

Технологическое развитие происходит за счет инноваций, исследований и разработок, а также адаптации существующих технологий. Прогнозирование технологического развития (ПТР) включает в себя анализ тенденций [2–3], экспертное мнение [4–7], использование математических моделей [8–10], а также средств интеллектуального анализа данных [11–15] для предсказания будущих изменений в технологических областях.

Актуальность данного исследования заключается в том, что быстро растущий объем открытых данных создает благоприятную основу для разработки новых методов анализа и прогнозирования технологического развития. Современные методы машинного обучения и искусственного интеллекта предоставляют значительные возможности для обработки этих больших объемов данных [16] и извлечения из них ценной информации, необходимой для точного и объективного прогнозирования технологических тенденций. В условиях стремительного технологического прогресса такие инструменты становятся все более востребованными для принятия обоснованных решений в различных областях.

Цель данной работы – исследовать инструменты прогнозирования технологического развития на основе данных из открытых источников на русском языке. В работе проанализированы 53 документа из различных источников, в основном из Российской научной электронной библиотеки eLIBRARY, за последние 10 лет 2014–2024 гг. с соблюдением каждого шага в соответствии с методом систематического картирования (Systematic Mapping Study, систематический обзор литературы, СОЛ). В рамках исследования определены наиболее частые и эффективные инструменты ПТР и типы источников данных, а также исследованы значимые результаты, подчеркивающие развитие в исследованиях в течение 10 лет.

Работа структурирована следующим образом. В разделе 2 представлен основной подход к проведению данного систематического исследования и подробно описаны процессы в рамках исследования и его результаты. В разделе 3 кратко изложены основные идеи наиболее эффективного метода, используемого в исследовании ПТР, который был раскрыт в разделе 2. Раздел 4 содержит сформулированные выводы.

## СИСТЕМАТИЧЕСКИЙ ОБЗОР ИНСТРУМЕНТОВ ПТР НА ОСНОВЕ ДАННЫХ ИЗ ОТКРЫТЫХ ИСТОЧНИКОВ

Существующие исследования в области прогнозирования технологического развития на основе открытых данных (ОД) характеризуются фрагментарностью и разнообразием подходов, что затрудняет для исследователей и практиков получение целостной картины развития данной области. Несмотря на возрастающий интерес к ОД как источнику информации для прогнозирования ПТР, сами методы и модели решения этой задачи все еще остаются недостаточно изученными. Для восполнения этого пробела, предлагается проведение систематического обзора литературы, целью которого является анализ и систематизация существующих методов и моделей прогнозирования, основанных на ОД, а также выявление актуальных исследовательских вопросов и перспективных направлений развития данной области.

СОЛ позволит сформировать основу для дальнейшего развития методологии технологического прогнозирования на основе ОД и повысить эффективность ее применения в различных областях.

Основная цель СОЛ – сбор научных исследований и их классификация. Весь процесс включает в себя формирование исследовательских вопросов, сбор данных, фильтрацию документов, тематический анализ текстов и анализ полученных данных [17]. Важно отметить, что в работе [17] представлен СОЛ на основе англоязычных источников, в то время как настоящее исследование фокусируется на анализе русскоязычных источников по данной теме. В последующих разделах каждый этап будет разобран подробнее (рис. 1).

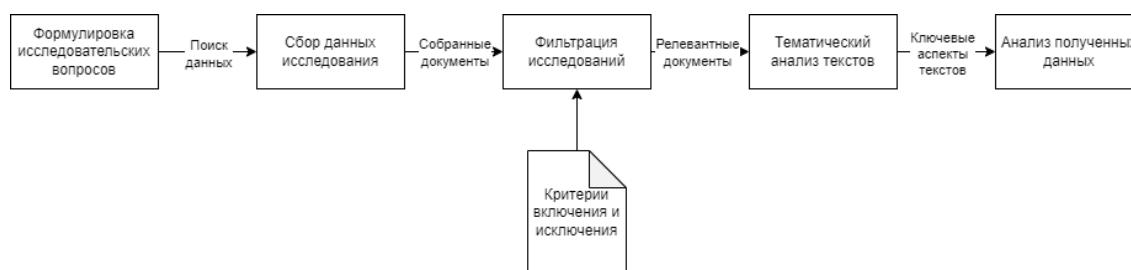


Рисунок 1 – Этапы систематического обзора литературы

### ФОРМУЛИРОВКА ИССЛЕДОВАТЕЛЬСКИХ ВОПРОСОВ

Данный этап является начальным в контексте исследовательского процесса. Ключевые вопросы исследования (ВИ), отражающие поставленные цели СОЛ, описаны ниже:

**ВИ1.** Какие типы открытых данных используются для прогнозирования технологического развития?

**ВИ2.** Какие инструменты используются для сбора, анализа, прогнозирования и визуализации открытых данных в контексте прогнозирования технологического развития?

**ВИ3.** Какие существуют ограничения и вызовы при использовании открытых данных для прогнозирования технологического развития?

### СБОР ДАННЫХ ИССЛЕДОВАНИЯ

Для поиска материалов применяется комплексный подход, включающий в себя как автоматизированные методы, так и ручную обработку информации. Исследование проводится с использованием поисковых запросов в научных базах данных, а также путем анализа публикаций в научных журналах и материалах конференций. Дополнительно для обеспечения полноты охвата анализируются патентные базы данных и новостные источники. Вследствие этого сформированные вопросы исследования помогают создать поисковый запрос. Поисковые запросы применялись в русскоязычных источниках, установка периода публикации документа ограничена 10 годами, т. е. с 2014–2024 гг. с указанием таких типов публикаций, как статья (eLIBRARY.RU), материалы конференции (eLIBRARY.RU), патенты (eLIBRARY.RU и Роспатент). В итоге количество извлеченных документов составляет 203.

Поисковые запросы:

– (прогнозирование технологий) И (открытые данные) И (машинное обучение ИЛИ глубокое обучение ИЛИ статистические методы ИЛИ интеллектуальный анализ данных);

– (прогнозирование технологий) И (открытые данные) И (патентные данные ИЛИ научные публикации ИЛИ данные социальных сетей ИЛИ экономические данные).

### ФИЛЬТРАЦИЯ ДОКУМЕНТОВ

Процесс отбора публикаций, соответствующих ВИ, требует тщательного анализа. Каждый документ проходит оценку на соответствие ВИ и определенным критериям включения (КВ). Документы оцениваются путем изучения заголовков и аннотаций публикаций. В случае если информации, представленной в заголовке и аннотации документа, оказывается недостаточно для полноценного анализа

его соответствия теме исследования, то необходимо углубиться в содержание и прочитать текст документа, чтобы принять решение о его включении или исключении из списка источников.

При формировании окончательного списка публикаций для анализа учитываются следующие критерии, которым должна соответствовать каждая работа, чтобы быть включенной в исследование:

**KB1.** В публикации рассматриваются поставленные ВИ.

**KB2.** Доступен полный текст публикации для детального анализа.

**KB3.** В публикации рассматриваются методы сбора, анализа или интерпретации данных из открытых источников, направленные на прогнозирование технологического развития.

Помимо KB, существует также ряд факторов, которые могут привести к исключению публикации из списка источников. Документ не будет включен в анализ, если он соответствует хотя бы одному из следующих критериев исключения (КИ):

**КИ1.** Публикация выходит за рамки выбранного временного периода исследования.

**КИ2.** Публикация не соответствует выбранному языку исследования.

После тщательного анализа и отсеивания публикаций, не соответствующих KB, изначальный список из 203 публикаций сократился до 53. Эти 53 документа представляют собой отобранный массив информации для дальнейшего углубленного анализа и формирования ответов на поставленные ВИ.

### ТЕМАТИЧЕСКИЙ АНАЛИЗ ТЕКСТОВ

Тематический анализ текстов направлен на извлечение доминирующих слов на основе изучения полных текстов литературы. В контексте систематического обзора литературы, посвященного технологическому прогнозированию на основе открытых данных, тематический анализ текстов позволяет выявлять ключевые слова для ответа на ВИ.

В ходе анализа аннотаций, ключевых слов и названий публикаций были выделены доминирующие термины, представленные на рисунке 2 в виде облака слов.



Рисунок 2 – Облако ключевых слов

Размер каждого слова отражает его частоту встречаемости в исследованных материалах, что позволяет визуально оценить значимость различных тематических направлений в области технологического прогнозирования на основе открытых данных.

### ТИПЫ ОТКРЫТЫХ ДАННЫХ

На рисунке 3 и в таблице 1 представлен ответ на **ВИ1** (Какие типы открытых данных используются для прогнозирования технологического развития?).

Выявлено четыре типа данных в исследованиях ПТР. На рисунке 3 показано процентное соотношение типов данных.

■ Статья в журнале ■ Патент ■ Материалы конференции ■ Новости

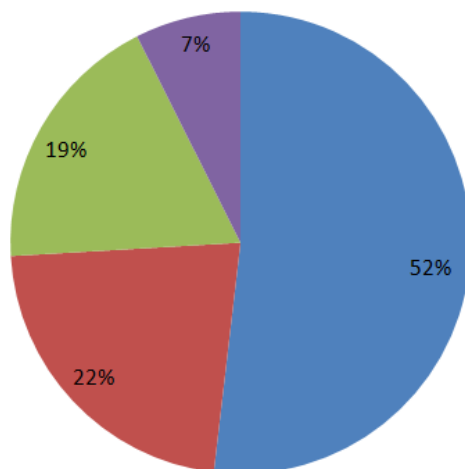


Рисунок 3 – Распределение типов данных

К этим типам относятся: 1) статьи из журнала; 2) патенты; 3) материалы конференций; 4) новостные ресурсы. При анализе установлено, что статьи из журналов использовались наиболее часто и составили больше половины публикаций (52 %). Следующие типы данных: патенты (22 %), материалы конференций (19 %) и новости (7 %).

В таблице 1 приведены соответствующие исследования различных типов данных для ПТР.

Таблица 1 – Описание типов данных

Тип данных	Описание	Ссылки
Статья в журнале	Научная электронная библиотека eLIBRARY.RU	[3], [5], [6], [7], [10], [11], [12], [14], [18], [19], [23–29], [39–44], [46–49], [54]
Патенты	Научная электронная библиотека eLIBRARY.RU, Роспатент	[1], [4], [8], [13], [15], [20] [30–34], [45]
Материалы конференций	Сборники статей международных и всероссийских научно-практических конференций	[2], [9], [16], [21], [22], [35–38]
Новости	Вестник ГЛОНАСС, ТАСС Наука, ТАСС	[50–53]

### ИНСТРУМЕНТЫ СБОРА, АНАЛИЗА И ВИЗУАЛИЗАЦИИ ОД

В таблице 2 представлен ответ на **ВИ2** (Какие инструменты используются для сбора, анализа и визуализации открытых данных в контексте прогнозирования технологического развития?).

Инструменты прогнозирования технологического развития помогают организациям и лицам принимать обоснованные решения, предвидеть будущие тенденции, управлять рисками и находить возможности. Они позволяют анализировать текущее состояние технологий, исследовать потенциальные пути развития, оценивать влияние новых технологий на общество и экономику, а также разрабатывать стратегии для адаптации к изменяющемуся технологическому ландшафту.

Список инструментов, используемых в ПТР, велик. Необходимо его ограничить и ввести порог отображения, равный 2, т. е. если данный инструмент встречается 2 и более раз, то он будет выведен.

Как видно из таблицы 2, наиболее эффективными и частыми инструментами, применяемыми в исследованиях ПТР, являются: машинное обучение, статистический анализ, параметрические модели, форсайт, нейронные сети, анализ, синтез, сравнение, классификация и т. д. Это вполне объяснимо, ведь большинство этих инструментов ПТР являются методами интеллектуального анализа данных (ИАД).

Таблица 2 – Инструменты ПТР с указанием ссылок (порог отображения = 2)

Название инструмента	Ссылки	Название инструмента	Ссылки
Машинное обучение	[4], [8], [11], [12], [13], [14], [16], [18], [19], [20], [21], [23], [27], [30], [50], [51], [52], [53], [54]	Статистический анализ	[2], [3], [9], [16], [20], [28], [34], [35], [41], [42], [54]
Параметрические модели	[8], [11], [14], [19], [20], [21], [23], [30], [51], [52], [53]	Форсайт	[4], [5], [7], [25], [26], [44], [46], [47]
Аналитические материалы	[16], [19], [25], [26], [36], [37], [44]	Большие данные	[4], [12], [16], [18], [19], [20], [30], [54]
Анализ и синтез, сравнение, классификация	[5], [6], [7], [12], [19], [20], [24], [39], [43], [46], [47]	Нейронные сети	[12], [14], [16], [23], [27], [51], [54]
Средства визуализации	[4], [12], [16], [19], [33],	Искусственный интеллект	[19], [32], [51]
Деревья решений	[38], [39]	Математико-статистические модели	[20], [41], [42], [54]
Программное обеспечение	[19], [20], [32], [50]	Системный и статистический анализ	[10], [49]
Алгоритм формирования сценариев	[29], [48],	Линейное, динамическое и целочисленное программирование	[13], [32]

### ОГРАНИЧЕНИЯ И ВЫЗОВЫ ПРИ ИСПОЛЬЗОВАНИИ ОД ДЛЯ ПТР

В данном разделе представлен ответ на **ВИЗ** (Какие существуют ограничения и вызовы при использовании открытых данных для прогнозирования технологического развития?).

Прогнозирование технологического развития – сложная задача, сталкивающаяся с рядом ограничений и вызовов. Проблемы и ограничения были сформированы по текстам документов. Ограничений так же много, как и инструментов ПТР, поэтому рассмотрим наиболее частые из них. Список и их описание можно увидеть ниже.

**Ошибка прогноза** [11, 31, 32]. Немодельная составляющая ошибки прогноза  $\sigma_n$  свидетельствует о том, что вне зависимости от вида выбираемой прогнозной модели точность оценки, в конечном счете её качество будут определяться остаточной дисперсией, описываемой соотношением:

$$\sigma_{ост}^2 = \frac{\sum_t (y_{\phi}(t_i) - y_p(t_i))^2}{n}, \quad (1)$$

где  $y_{\phi}(t_i)$  – фактическое значение оцениваемого параметра  $y$  в точке  $t_i$ ;

$y_p(t_i)$  – расчетное значение оцениваемого параметра  $y$  в точке  $t_i$ , получаемой с использованием избранной модели регрессии.

**Проблема переобучения** [13, 21, 37, 42]. Переобучение (overfitting) – это распространенная проблема в машинном обучении, которая возникает, когда модель слишком хорошо подстраивается к обучающим данным, включая случайные флуктуации и шумы. В результате модель плохо обобщается на новые, невиданные данные, что приводит к снижению точности прогнозирования:

$$L(w) = L_0(w) + \lambda ||w||^2, \quad (2)$$

где  $L(w)$  – функция потерь с регуляризацией;

$L_0(w)$  – исходная функция потерь;

$\lambda$  – коэффициент регуляризации (гиперпараметр);

$w$  – вектор весов модели;

$||w||^2$  – L2-норма вектора весов (сумма квадратов всех весов).

**Проблема точности** [30, 37, 44]. Прогнозирование технологического развития сопряжено с проблемой точности, так как будущее технологий подвержено множеству неопределенностей и влиянию различных факторов. Достижение высокой точности прогнозов является сложной задачей, требующей комплексного подхода. Метрики как функции ошибки позволяют оценить точность прогнозирования. Этих формул несколько, выведем одну из них, а именно среднеквадратичной ошибки:

$$MSE = \frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2, \quad (3)$$

где  $n$  – количество наблюдений;

$Y_i$  – фактическое значение для  $i$ -го наблюдения;

$\hat{Y}_i$  – прогнозируемое значение для  $i$ -го наблюдения.

Чем ближе MSE к нулю, тем точнее модель предсказывает фактические значения.

**Неопределенность системы** [8, 21, 28, 40, 41]. Энтропия является мерой неопределенности или случайности в системе. В контексте прогнозирования технологического развития, энтропия используется для оценки степени неопределенности, связанной с будущим направлением и темпами технологического прогресса. Для расчёта и оценки неопределенности применяется формула Шеннона. Часто используется в методах математической статистики:

$$H(x) = - \sum_{i=1}^n p_i(x) \log_2 p_i(x), \quad (4)$$

где  $H(x)$  – энтропия случайной переменной  $x$ ;

$p_i(x)$  – вероятность значения  $x$ ;

$\Sigma$  – суммирование по всем возможным значениям  $x$ .

#### **КЛЮЧЕВЫЕ ПРИНЦИПЫ ИАД В ПТР**

Обзор выявил интеллектуальный анализ данных как наиболее эффективный и востребованный метод. Это объясняется применением таких инструментов, как машинное обучение, нейронные сети и большие данные. Вместе с параметрическими моделями они упрощают анализ данных, особенно при работе с большими объёмами данных. Раздел рассматривает 8 исследований, иллюстрирующих ключевые концепции и общий алгоритм интеллектуального анализа данных, подчёркивая его значимость и актуальность.

В статье [12] (2022) предлагается метод прогнозирования технологических трендов, основанный на анализе разнородных данных. В работе выдвигается гипотеза, что восходящий тренд цены акций высокотехнологичной компании свидетельствует о перспективности разрабатываемых ею технологий. Метод включает анализ веб-новостей для выявления значимых компаний, прогнозирование тренда цен их акций с использованием моделей машинного и глубокого обучения (с байесовской оптимизацией гиперпараметров) и кластерный анализ патентных заявок этих компаний для определения направлений их разработок. В качестве примеров рассматривается прогнозирование динамики цен акций Tesla и Samsung и выявление технологических трендов на основе анализа их патентных заявок. Результаты показывают высокую точность прогнозов, подтверждая эффективность предложенного метода.

В исследовании [14] (2020) рассматривается проблема прогнозирования трендов технологического развития. В работе проводится анализ данных патентов, трендов и по итогу формируются наборы данных для обучения нейронной сети. В документе описываются методы машинного обучения, такие как градиентный бустинг, деревья решений и глубокое обучение. Такие методы помогают проводить более точный анализ данных. По итогу авторы разрабатывают модель машинного обучения, описывают работу алгоритмов и демонстрируют формирование прогнозов, т. е. результат работы модели.

Исследование [18] (2018) посвящено инструментам интеллектуального анализа «больших данных» с целью выявления важных тенденций в развитии научно-технических инноваций в промышленности. В центре внимания авторов – алгоритмы, реализующие методы и модели решения задач классификации данных, такие как деревья решений, случайные леса, «ленивая» классификация, ассоциативные классификаторы и нейронные сети, включая глубокое обучение, а также методы опорных векторов. Методом опорных векторов строится классифицирующая функция:

$$F(x) = \text{sign}(\langle w, x \rangle + b), \quad (5)$$

где  $\text{sign} \langle w, x \rangle$  – скалярное произведение;

$w$  – нормаль к разделяющей гиперплоскости;

$b$  – вспомогательный параметр.

Эти методы используются в процессе составления прогнозов научно-технического развития, на основе которых делаются выводы и принимаются обоснованные решения.

В исследовании [19] (2022) предлагается новый метод анализа и прогнозирования технологических трендов (АПТТ) на основе анализа научных статей и патентных заявок из открытых источников. Сначала извлекаются значимые ключевые слова из 20 влиятельных статей в области ИИ из базы данных Web of Science. Затем с помощью регрессии гауссовского процесса прогнозируются тренды изменения импакт-оценки этих ключевых слов. Далее, используя предсказанные ключевые слова, собираются патентные заявки из платформы The Lens. Из этих заявок извлекаются ключевые слова и коды совместной патентной классификации, на основе которых создаются матрицы совместного появления. Визуализация и кластеризация этих матриц с помощью VOSviewer позволяет выявить основные технологические тренды. Точность метода оценивается путем сравнения предсказанных трендов с 25 наиболее важными технологическими трендами, описанными экспертом Bernard Marr. Результаты показывают высокую точность прогнозов, подтверждая эффективность предложенного метода АПТТ.

Зарегистрированная программа [20] (2022) «Предиктивное моделирование технологических трендов» анализирует большие данные из патентных заявок, классифицированных по Международной патентной классификации, чтобы предсказать будущие тенденции развития технологий. Она

строит временные ряды для каждой технологической области и применяет методы машинного обучения, такие как наивное прогнозирование, экспоненциальное сглаживание и ARIMA, чтобы предсказать будущий тренд. Общая формула ARIMA(p, d, q) выглядит следующим образом:

$$\Delta^d Y(t) = c + \varphi_1 * \Delta^d Y(t-1) + \dots + \varphi_p * \Delta^d Y(t-p) + \theta_1 * \varepsilon(t-1) + \dots + \theta_q * \varepsilon(t-q) + \varepsilon(t), \quad (6)$$

где  $Y(t)$  – значение временного ряда в момент времени  $t$ ;

$c$  – константа;

$p$  – порядок авторегрессии (AR);

$\varphi_1 \dots \varphi_p$  – коэффициенты авторегрессии;

$d$  – порядок интегрированности (I);

$q$  – порядок скользящего среднего (MA);

$\theta_1 \dots \theta_q$  – коэффициенты скользящего среднего;

$\varepsilon(t)$  – ошибка (шум) в момент времени  $t$ .

Затем, оценивая точность моделей с помощью RMSE, программа выбирает лучшую модель для каждой области и идентифицирует «угасающие», перспективные и «прорывные» технологии, что помогает принимать обоснованные решения в различных сферах. Формула RMSE имеет такой вид:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}, \quad (7)$$

где  $y_i$  – фактическое значение временного ряда в момент времени  $i$ ;

$\hat{y}_i$  – прогнозируемое значение временного ряда в момент времени  $i$ ;

$n$  – количество точек данных во временном ряде;

$\Sigma$  – суммирование по всем точкам данных.

Материалы статьи [21] (2020) отражают вопрос прогнозирования развития энергетики в условиях высокой конкурентности рынка. В данном исследовании описывается применение такого инструмента, как параметрическая модель. Применение модели позволяет провести прогноз развития энергетики в контексте тепловой электростанции для повышения эффективности ее работы. Как результат демонстрируются такие перспективные решения, как: аккумуляторы воды, многоступенчатый нагрев воды, модификация и внедрение предварительного подогрева воды для улучшения эффективности характеристик станции и выработки электроэнергии.

Документ [22] (2020) демонстрирует использование такого инструмента интеллектуального анализа данных, как обработка естественного языка (NLP). Данный инструмент применяется для прогнозирования развития энергетической инфраструктуры России. В работе представлено подробное описание работы семантического анализа данных, а также инструкция, как необходимо использовать модуль для автоматизации прогнозирования. Инструмент отлично работает с большими массивами данных, что и демонстрирует исследование, ведь такое количество данных позволяет повысить качество и точность прогнозирования.

В исследовании [23] (2021) рассматривается проблема инновационного развития региональной экономики в контексте экономической безопасности Российской Федерации. Авторы анализируют и прогнозируют развитие региональной экономики с помощью кластерного анализа на основе нейросетевого моделирования с использованием официальных статистических данных. Нейросетевые инструменты, в частности самоорганизующаяся карта Кохонена, позволили провести кластеризацию данных и классифицировать регионы Российской Федерации по пяти кластерам, независимо от их принадлежности к федеральным округам. В ходе исследования выявлена дифференциация регионов по показателям инновационного развития и определены различные стратегии развития в инновационной сфере.

### ЗАКЛЮЧЕНИЕ

В проведенном систематическом обзоре литературы было проанализировано 53 документа, посвященных методам прогнозирования технологического развития на основе открытых данных. Среди типов данных наиболее популярными оказались статьи из журналов, которые составили более половины от общего количества документов. Рассмотрены также такие типы ОД, как патенты, включая свидетельства о регистрации программ для ЭВМ, материалы конференций и новостные источники. В отличие от имеющихся обзоров, использующих данные зарубежных источников, в представленном исследовании рассмотрены отечественные информационные ресурсы.

В ходе исследования выявлены ограничения при использовании открытых данных: ошибка прогноза, проблема переобучения, проблема точности и неопределенность системы с описанием этих ограничений.

Использование инструментов машинного обучения, нейронных сетей, искусственного интеллекта, параметрических моделей и т. д. позволяет анализировать большие объемы данных и строить

точные прогностические модели, а также выявлять закономерности, которые трудно или невозможно обнаружить с помощью традиционных методов. Несмотря на сложность настройки, модели ИАД при правильном применении способны выдавать более эффективные и точные прогнозы технологического развития.

Результаты систематического обзора подчеркивают важность дальнейшего развития методологии и инструментария для прогнозирования технологического развития на основе открытых данных. Особое внимание следует уделить повышению точности и достоверности прогнозов, а также учету неопределенности и сложности технологических систем.

#### Список источников

1. Патент РФ 2686820. Система и способ настройки моделей процессов / Я. Дж. Хорн, К. Роматье, П. Ковальчик, З. Алзейн. – Заявл. 30.03.2016 ; опубл. 30.04.2019.
2. Котов, А. В. Пространственное стратегирование в начале 1920-х годов и стратегия пространственного развития России на 2019–2025 годы.: подходы, институты, инструменты / А. В. Котов // Российские экономические реформы в региональном измерении : сборник материалов Всероссийской научной конференции, посвященной столетию начала НЭПа, Новосибирск, 16–17 сентября 2021 года / отв. редакторы В. А. Ильиных, В. М. Рынков. – Новосибирск : Параллель, 2021. – С. 149–156.
3. Глушенко, В. В. Формирование инновационной политики организаций в период перехода к шестому технологическому укладу / В. В. Глушенко // Современные научные исследования и инновации. – 2021. – № 8 (124).
4. Свидетельство о государственной регистрации программы для ЭВМ № 2022613485 Российская Федерация. Экспертно-аналитическая модель получения энергонасыщенных экструдатов из питательных смесей на основе зерна : № 2022612862 : заявл. 02.03.2022 : опубл. 14.03.2022 / И. А. Чаплыгина, В. В. Матюшев, А. В. Семенов, А. А. Беляков ; заявитель Федеральное государственное бюджетное образовательное учреждение высшего образования «Красноярский государственный аграрный университет».
5. Левина, Е. В. Цифровые технологии и технико-технологическое обеспечение как платформа устойчивого развития промышленных предприятий / Е. В. Левина // Экономика и социум: современные модели развития. – 2021. – Т. 11, № 1. – С. 59–70. – DOI: 10.18334/ecsoc.11.1.111733.
6. Данелян, Т. Я. Формальные методы экспертных оценок / Т. Я. Данелян // Экономика, статистика и информатика. Вестник УМО. – 2015. – № 1. – С. 183–187.
7. Петухова, М. С. Алгоритм разработки дорожной карты научно-технологического развития отрасли растениеводства с использованием когнитивных методов / М. С. Петухова // Экономический обзор. – 2020. – № 9–10 (9). – С. 11–14.
8. Свидетельство о государственной регистрации программы для ЭВМ № 2022610237 Российская Федерация. Инструменты оперативного контроля снижения производительности и выявления технологических нарушений в цикле производства алюминия : № 2021682044 : заявл. 29.12.2021 : опубл. 10.01.2022 / Т. Г. Пенькова, А. С. Михалев, А. В. Коробко, А. М. Метус, Н. М. Баранова, В. А. Баранов, А. А. Коробко, М. Ю. Сенашова ; заявитель Федеральное государственное бюджетное научное учреждение «Федеральный исследовательский центр «Красноярский научный центр Сибирского отделения Российской академии наук».
9. Лубнина, А. А. Моделирование и прогнозирование инновационного и технологического развития нефтехимических предприятий в условиях кооперации / А. А. Лубнина, М. В. Шинкевич // Международный форум «Kazan Digital Week – 2021» : сборник материалов, Казань, 21–24 сентября 2021 года. – Казань : ГБУ «НЦБЖД», 2021. – Т. 1. – С. 313–319.
10. Покровская, О. Д. Экономическое сравнение вариантов организации контейнерного терминала на сортировочной станции / О. Д. Покровская // Бюллетень результатов научных исследований. – 2020. – № 4. – С. 78–100. – DOI: 10.20295/2223-9987-2020-4-78-100.
11. Серяков, Г. Н. Развитие методологических аспектов прогнозирования процессов развития нового технологического уклада экономики / Г. Н. Серяков // Вестник Полоцкого государственного университета. Серия Д. Экономические и юридические науки. – 2018. – № 6. – С. 86–93.
12. Кравец, А. Г. Прогнозирование технологических тенденций на основе анализа разнородных данных / А. Г. Кравец, Т. В. Нгуен // Программные продукты и системы. – 2022. – № 3. – С. 396–412. – DOI: 10.15827/0236-235X.139.396-412.
13. Свидетельство о государственной регистрации программы для ЭВМ № 2023688399 Российская Федерация. Программа генерации сверх точных предиктивных моделей машинного интеллекта : № 2023687444 : заявл. 11.12.2023 : опубл. 21.12.2023 / И. А. Андреев.
14. Кравец, А. Г. Предсказательное моделирование трендов технологического развития / А. Г. Кравец, Н. А. Сальникова // Известия Санкт-Петербургского государственного технологического института (технического университета). – 2020. – № 55 (81). – С. 103–108. – DOI: 10.36807/1998-9849-2020-55-81-103-108.
15. Патент РФ 2727631. Системы и способы прогнозирования структуры и свойств атомарных веществ и их сплавов / С. Маити, Ш. Мишра, Б. С. Дваси, Б. Раи. – Заявл. 18.03.2019 ; опубл. 22.07.2020.
16. Отраслевое научно-технологическое прогнозирование: от построения образа к изучению неопределенного будущего / Н. В. Трифонова, М. С. Власова, И. А. Лобанова, А. С. Ковалева // Развивая энергетическую повестку будущего : сборник докладов Международной научно-практической конференции для представителей сообщества молодых инженеров ТЭК, Санкт-Петербург, 10–11 декабря 2021 года. – Санкт-Петербург : Санкт-Петербургский государственный электротехнический университет «ЛЭТИ» им. В.И. Ульянова (Ленина), 2021. – С. 241–245.

17. Viet, N. T. Data Mining Methods for Analysis and Forecast of an Emerging Technology Trend: A Systematic Mapping Study from SCOPUS Papers / N. T. Viet, A. Kravets, T. Duong Quoc Hoang // *Lecture Notes in Computer Science*. – 2021. – Vol. 12948 LNAI. – P. 81–101. – DOI: 10.1007/978-3-030-86855-0\_7.
18. Голубев, С. С. Методы и модели решения задач классификации при инновационном прогнозировании научно-технологического развития с использованием интеллектуального анализа «больших» данных / С. С. Голубев, Е. П. Дюндик, Е. В. Скубрий // *Вестник МГПУ. Серия: Экономика*. – 2018. – № 4 (18). – С. 8–16. – DOI: 10.25688/2312-6647.2018.18.4.1.
19. Вьет, Н. Т. Новый метод прогнозирования технологических трендов на основе анализа научных статей и патентов / Н. Т. Вьет, А. Г. Кравец // *International Journal of Open Information Technologies*. – 2022. – Т. 10, № 10. – С. 49–62.
20. Свидетельство о государственной регистрации программы для ЭВМ № 2022683319 Российская Федерация. Предиктивное моделирование технологических трендов : № 2022681820 : заявл. 16.11.2022 : опубл. 05.12.2022 / О. А. Хохлова, А. Н. Хохлова ; заявитель Федеральное государственное бюджетное образовательное учреждение высшего образования «Восточно-Сибирский государственный университет технологий и управления».
21. Лисин, Е. М. Обеспечение устойчивого развития энергетической отрасли на основе повышения конкурентоспособности ТЭЦ / Е. М. Лисин, И. С. Замшаева // *Устойчивое развитие цифровой экономики, промышленности и инновационных систем : сборник трудов научно-практической конференции с зарубежным участием, Санкт-Петербург, 20–21 ноября 2020 года / под ред. Д. Г. Родионова, А. В. Бабкина*. – Санкт-Петербург : ПОЛИТЕХ-ПРЕСС, 2020. – С. 120–123. – DOI: 10.18720/IEP/2020.7/34.
22. Копайгородский, А. Н. Семантический анализ Big Data в задаче прогнозирования инновационного развития энергетической инфраструктуры РФ / А. Н. Копайгородский, Е. П. Хайруллина, И. И. Хайруллин // *Физико-техническая информатика (СРТ2020) : материалы 8-й Международной конференции, Пушкино, Московская обл., 09–13 ноября 2020 года*. – Нижний Новгород : Автономная некоммерческая организация в области информационных технологий «Научно-исследовательский центр физико-технической информатики», 2020. – Т. 2. – С. 199–203. – DOI: 10.30987/conferencearticle\_5fd755c02b5fa9.82083943.
23. Исследование инновационного развития региональной экономики в условиях глобальных вызовов с использованием методов искусственного интеллекта / Е. Н. Летагина, В. И. Перова, С. Н. Яшин, С. А. Борисов // *Экономическая безопасность*. – 2021. – Т. 4, № 3. – С. 781–794. – DOI: 10.18334/есес.4.3.112025.
24. Авдзейко, В. И. Патентный анализ. Выявление перспективных и прорывных технологий / В. И. Авдзейко, В. И. Карнышев, Р. В. Мещеряков // *Вопросы инновационной экономики*. – 2018. – Т. 8, № 1. – С. 79–90. – DOI: 10.18334/vinec.8.1.38890.
25. Миронов, Н. А. Анализ опыта применения экспертных методов прогнозирования в информационной системе Федерального реестра экспертов научно-технической сферы для разработки направлений научно-технологического развития / Н. А. Миронов, Е. А. Марышев, Н. А. Лукашева, Н. А. Дивуева // *Инноватика и экспертиза : научные труды*. – 2022. – № 1 (33). – С. 55–64.
26. Шеломенцева, В. П. Использование метода «форсайт» для прогнозирования научно-технологического и экономического развития старопромышленного региона (на примере Павлодарской области Республики Казахстан) / В. П. Шеломенцева, Е. Б. Никитин, С. В. Беспалый, Е. А. Ифутина // *Вестник Омского университета. Серия: Экономика*. – 2015. – № 4. – С. 242–248.
27. Абдурахманов, А. М. Методы прогнозирования электропотребления в распределительных сетях (обзор) / А. М. Абдурахманов, М. В. Володин, Е. Ю. Зыбин, В. Н. Рябченко // *Электротехника*. – 2016. – Т. 3, № 1. – С. 3–23.
28. Шинкевич, А. И. Тенденции новационного развития обрабатывающих отраслей промышленности / А. И. Шинкевич, А. А. Лубнина, И. А. Райский // *Известия Самарского научного центра Российской академии наук*. – 2021. – Т. 23, № 4 (102). – С. 51–56. – DOI: 10.37313/1990-5378-2021-23-4-51-56.
29. Белошицкий, А. В. К вопросу формирования сценариев устойчивого развития нефтесервисной отрасли / А. В. Белошицкий, А. Е. Череповицын // *Вестник Южно-Российского государственного технического университета (НПИ). Серия: Социально-экономические науки*. – 2022. – Т. 15, № 2. – С. 95–106. – DOI: 10.17213/2075-2067-2022-2-95-106.
30. Патент РФ 2722525. Адаптивное управление процессами аддитивного производства в реальном масштабе времени с использованием машинного обучения / Э. Мехр, Т. Эллис, Дж. Нун. – Заявл. 23.05.2018 ; опубл. 01.06.2020.
31. Патент РФ 2663011. Система и способ выполнения операции интенсификации скважины / Б. Лесерф, Д. Усольцев, Т. Поуп. – Заявл. 01.12.2014 ; опубл. 01.08.2018.
32. Свидетельство о государственной регистрации программы для ЭВМ № 2015612989 Российская Федерация. Система методов управления развитием растениеводства (ПС «Методы») : № 2015610223 : заявл. 12.01.2015 : опубл. 27.02.2015 / А. С. Бахмут, Н. Ф. Гайворонская, Г. В. Григорьева, О. В. Егорова, В. В. Кузнецов ; заявитель Федеральное государственное бюджетное научное учреждение «Всероссийский научно-исследовательский институт экономики и нормативов».
33. Свидетельство о государственной регистрации программы для ЭВМ № 2023688714 Российская Федерация. Программа для нахождения параметров логистического уравнения : № 2023688468 : заявл. 18.12.2023 : опубл. 25.12.2023 / Д. Т. Кампос, И. В. Ильин, О. Ю. Ильяшенко, Е. М. Щеников ; заявитель федеральное государственное автономное образовательное учреждение высшего образования «Санкт-Петербургский политехнический университет Петра Великого».
34. Патент РФ 2783863. Способ комплексной оптимизации параметров энергоблока / М. В. Лифшиц. – Заявл. 03.09.2021 ; опубл. 21.11.2022.

35. Владимиров, И. С. О специфике методов научно-технологического прогнозирования / И. С. Владимиров, В. И. Азаров, А. В. Боровик, И. А. Мельников, А. Н. Ватлин // Инновационные научные исследования в современном мире: теория, методология, практика : сборник научных статей по материалам VII Международной научно-практической конференции, Уфа, 31 января 2022 года. – Уфа : Общество с ограниченной ответственностью «Научно-издательский центр «Вестник науки», 2022. – С. 49–54.
36. Дубишкин, Е. А. Институциональный механизм рынка технологий / Е. А. Дубишкин // Государственное управление и развитие России: вызовы и перспективы : сборник статей VI Всероссийской научно-практической конференции, Пенза, 25–26 ноября 2022 года / под науч. ред. А. В. Яшина, А. А. Грачева. – Пенза : Пензенский государственный аграрный университет, 2022. – С. 104–110.
37. Петухова, М. С. Теоретические основы прогнозирования научно-технологического развития отрасли растениеводства / М. С. Петухова // Приоритеты стратегии научно-технологического развития России и обеспечение воспроизводства инновационного потенциала высшей школы : материалы Всероссийской научной конференции, Ижевск, 27 февраля 2019 года / ответственный ред. А. М. Макаров. – Ижевск : Издательский дом «Удмуртский университет», 2019. – С. 121–124.
38. Владимиров, И. С. Перспективные методы научно-технологического прогнозирования / И. С. Владимиров, А. Ю. Вopilов, М. В. Орешкин, А. В. Боровик, М. А. Соловьев // Современные научные исследования: теория, методология, практика : сборник научных статей по материалам VII Международной научно-практической конференции, Уфа, 16 февраля 2022 года. – Уфа : Общество с ограниченной ответственностью «Научно-издательский центр «Вестник науки», 2022. – С. 13–18.
39. Глущенко, В. В. Функции и роли технологических укладов в управления технологическим и социально-экономическим развитием / В. В. Глущенко // Kazakhstan Science Journal. – 2020. – Т. 3, № 3 (16). – С. 80–93.
40. Тимофеева, Н. Ю. Регрессионный анализ и кривые роста экономических рядов динамики в прогнозировании развития сельхозпроизводителей Липецкой области / Н. Ю. Тимофеева // Известия высших учебных заведений. Поволжский регион. Общественные науки. – 2017. – № 4 (44). – С. 145–157. – DOI: 10.21685/2072-3016-2017-4-16.
41. Болдыревский, П. Б. Модели прогнозирования основных показателей инновационной деятельности промышленных предприятий / П. Б. Болдыревский, Л. А. Кистанова // Экономический анализ: теория и практика. – 2014. – № 29 (380). – С. 52–57.
42. Грабчак, Е. П. О подходах к определению перспектив развития электроэнергетики / Е. П. Грабчак // Проблемы рыночной экономики. – 2021. – № 4. – С. 8–19. – DOI: 10.33051/2500-2325-2021-4-8-19.
43. Комков, Н. И. Комплексное прогнозирование научно-технологического развития: опыт и уроки / Н. И. Комков // Проблемы прогнозирования. – 2014. – № 2 (143). – С. 3–17.
44. Пудовкина, О. Е. Технологические приоритеты развития промышленного сектора России в условиях глубокого проникновения цифровых технологий / О. Е. Пудовкина // Вестник университета. – 2021. – № 9. – С. 74–80. – DOI: 10.26425/1816-4277-2021-9-74-80.
45. Патент РФ 2686257. Способ и система удалённой идентификации и прогнозирования развития зарождающихся дефектов / Наумов С. А., Крымский А. В., Лифшиц М. В. – Заявл. 27.04.2018 ; опубл. 24.04.2019.
46. Дегтярева, В. В. Применение метода корпоративного форсайта для совершенствования системы управления инновационной деятельностью / В. В. Дегтярева, Е. Ю. Камчатова, Е. А. Какаева // Modern Economy Success. – 2019. – № 2. – С. 105–108.
47. Корнев, А. К. О повышении конкурентоспособности действующих производств обрабатывающей промышленности / А. К. Корнев, С. И. Максимцова // Проблемы прогнозирования. – 2019. – № 6 (177). – С. 83–94.
48. Сиптиц, С. О. Особенности моделирования и сценарного анализа влияния долгосрочных климатических изменений на агропродовольственные системы регионов / С. О. Сиптиц, И. А. Романенко, И. А. Ганиева, Н. Е. Евдокимова // Достижения науки и техники АПК. – 2020. – Т. 34, № 4. – С. 5–9. – DOI: 10.24411/0235-2451-2020-10401.
49. Костарев, А. С. Оценка резервов инновационного развития угледобывающего производственного объединения и влияние организационно-экономических отношений на их использование / А. С. Костарев // Известия Уральского государственного горного университета. – 2020. – № 1 (57). – С. 208–217. – DOI: 10.21440/2307-2091-2020-1-208-217.
50. Научные центры мирового уровня предложили решения для развития промышленности. – URL: <https://nauka.tass.ru/nauka/17000583> (дата обращения: 14.04.2024).
51. Конференция Сбера по искусственному интеллекту AIJ 2023. Текстовая трансляция второго дня. – URL: <https://tass.ru/online-conference/19352375> (дата обращения: 14.04.2024).
52. Билайн запатентовал новый метод рефарминга. – URL: [https://www.cnews.ru/news/line/2024-04-11\\_bilajn\\_zarpatentoval\\_novuj](https://www.cnews.ru/news/line/2024-04-11_bilajn_zarpatentoval_novuj) (дата обращения: 14.04.2024).
53. Кибергиена для систем искусственного интеллекта в промышленном Интернете вещей. – URL: <http://vestnik-ghonass.ru/news/intro/kibergigiena-dlya-sistem-iskusstvennogo-intellekta-v-promyshlennom-internete-veshchey> (дата обращения: 14.04.2024).
54. Нгуен, Т. В. Анализ и прогноз тенденций использования терминов в компьютерных науках на основе нейросетевых моделей / Т. В. Нгуен, К. Х. Т. Зьонг, А. Г. Кравец // Вестник компьютерных и информационных технологий. – 2021. – Т. 18, № 2 (200). – С. 24–38. – DOI: 10.14489/vkit.2021.02.pp.024-038.

## References

1. Patent RF 2686820. *System and method of process model customization*. Ya. Dzh. Horn, K. Romate, P. Kovalchik, Z. Alzejn. Decl. 30.03.2016 ; publ. 30.04.2019 (In Russ.).
2. Kotov, A. V. Spatial strategizing in the early 1920s and Russia's spatial development strategy for 2019–2025: approaches, institutions, and tools. *Russian economic reforms in the regional dimension : proceedings of the All-Russian scientific conference devoted to the centenary of the beginning of the NEP, Novosibirsk, September 16–17, 2021*. Editors V. A. Ilinykh, V. M. Rynkov. Novosibirsk, Parallel Publ., 2021, pp. 149–156 (In Russ.).
3. Glushhenko, V. V. Formation of innovation policy of organizations in the period of transition to the sixth technological mode. *Modern scientific research and innovations*, 2021, no. 8 (124) (In Russ.).
4. Chaplygina, I. A., Matyushev, V. V., Semenov, A. V., Belyakov, A. A. *Certificate of state registration of computer program No. 2022613485 Russian Federation. Expert-analytical model for obtaining energy-rich extrudates from grain-based nutrient mixtures* : no. 2022612862 : declared 02.03.2022 : published 14.03.2022 ; applicant Federal State Budgetary Educational Institution of Higher Education “Krasnoyarsk State Agrarian University” (In Russ.).
5. Levina, E. V. Digital technologies and technical and technological support as a platform for sustainable development of industrial enterprises. *Economy and society: modern models of development*, 2021, vol 11, no. 1, pp. 59–70. DOI: 10.18334/ecsoc.11.1.111733 (In Russ.).
6. Danelyan, T. Ya. Formal methods of expert assessments. *Economics, Statistics and Informatics. Bulletin of UMO*, 2015, no. 1, pp. 183–187.
7. Petukhova, M. S. Algorithm for the development of a roadmap for scientific and technological development of the crop production industry using cognitive methods. *Economic Review*, 2020, no. 9–10 (9), pp. 11–14 (In Russ.).
8. Penkova, T. G., Mikhalev, A. S., Korobko, A. V., Metus, A. M., Baranova, N. M., Baranov, V. A., Korobko, A. A., Senashova M. Yu. *Certificate of State Registration of Computer Program no. 2022610237 Russian Federation. Tools for operational control of productivity reduction and detection of technological violations in the aluminum production cycle* : no. 2021682044 : declared 29.12.2021 : published 10.01.2022 ; applicant Federal State Budgetary Scientific Institution “Federal Research Center Krasnoyarsk Research Center of the Siberian Branch of the Russian Academy of Sciences” (In Russ.).
9. Lubnina, A. A., Shinkevich, M. V. Modeling and forecasting of innovative and technological development of petrochemical enterprises in conditions of cooperation. *International Forum “Kazan Digital Week – 2021” : proceedings, Kazan, September 21–24, 2021*. Kazan, State Budgetary Institution “NCBWC”, 2021, vol. 1, pp. 313–319 (In Russ.).
10. Pokrovskaya, O. D. Economic comparison of options for organizing a container terminal at a marshalling yard]. *Research Bulletin*, 2020, no. 4, pp. 78–100. DOI: 10.20295/2223-9987-2020-4-78-100 (In Russ.).
11. Seryakov, G. N. Development of methodological aspects of forecasting the processes of development of a new technological mode of the economy. *Bulletin of Polotsk State University. Series D. Economic and Legal Sciences*, 2018, no. 6, pp. 86–93 (In Russ.).
12. Kravets, A. G., Nguen, T. V. Forecasting technology trends based on heterogeneous data analysis. *Software products and systems*, 2022, no. 3, pp. 396–412. DOI: 10.15827/0236-235X.139.396-412 (In Russ.).
13. Andreev, I. A. *Certificate of State Registration of Computer Program no. 2023688399 Russian Federation. Program for generation of super accurate predictive models of machine intelligence* : no. 2023687444 : declared 11.12.2023 : publ. 21.12.2023 (In Russ.).
14. Kravets, A. G., Salnikova, N. A. Predictive modeling of technological development trends. *Izvestiya of St. Petersburg State Technological Institute (Technical University)*, 2020, no. 55 (81), pp. 103–108. DOI: 10.36807/1998-9849-2020-55-81-103-108 (In Russ.).
15. Patent RF 2727631. *Systems and methods for predicting the structure and properties of atomic substances and their alloys*. S. Maiti, Sh. Mishra, B. S. Dvadas, B. Rai. Decl. 18.03.2019 ; publ. 22.07.2020 (In Russ.).
16. Trifonova, N. V., Vlasova, M. S., Lobanova, I. A., Kovaleva, A. S. Sectoral S&T forecasting: from image building to exploring an uncertain future. *Developing the Future Energy Agenda : Collection of Papers of the International Scientific and Practical Conference for Representatives of the Community of Young Engineers of Fuel and Energy Complex, St. Petersburg, December 10–11, 2021*. St. Petersburg: St. Petersburg State Electrotechnical University “LETI” named after V.I. Ulyanov (Lenin). St. Petersburg, 2021, pp. 241–245 (In Russ.).
17. Viet, N. T., Kravets, A., Duong, Quoc Hoang T. Data Mining Methods for Analysis and Forecast of an Emerging Technology Trend: A Systematic Mapping Study from SCOPUS Papers. *Lecture Notes in Computer Science*, 2021, vol. 12948 pp. 81–101. DOI: 10.1007/978-3-030-86855-0\_7.
18. Golubev, S. S., Dyundik, E. P., Skubriy, E. V. Methods and models for solving classification problems in innovative forecasting of scientific and technological development using intelligent analysis of “big” data. *Bulletin of Moscow State Pedagogical University. Series: Economics*, 2018, no. 4 (18), pp. 8–16. DOI: 10.25688/2312-6647.2018.18.4.1 (In Russ.).
19. Vet, N. T., Kravets, A. G. A new method for predicting technology trends based on the analysis of scientific articles and patents. *International Journal of Open Information Technologies*, 2022, vol. 10, no. 10, pp. 49–62 (In Russ.).
20. Khokhlova, O. A., Khokhlova, A. N. *Certificate of State Registration of Computer Program No. 2022683319 Russian Federation. Predictive modeling of technological trends* : no. 2022681820 : decl. 16.11.2022 : publ. 05.12.2022 ; applicant Federal State Budgetary Educational Institution of Higher Education “East-Siberian State University of Technology and Management” (In Russ.).
21. Lisin, E. M., Zameshaeva, I. S. Ensuring sustainable development of the energy sector through increased competitiveness of thermal power plants. *Sustainable development of digital economy, industry and innovation systems : proceedings of the scientific and practical conference with foreign participation, St. Petersburg, November 20–21, 2020*. Ed. by D. G. Rodionov, A. V. Babkin. St. Petersburg, POLYTEKH-PRESS Publ., 2020, pp. 120–123. DOI: 10.18720/IEP/2020.7/34 (In Russ.).

22. Kopyagorodskiy, A. N., Hayrullina E. P., Hayrullin, I. I. Semantic analysis of Big Data in the task of forecasting innovative development of the energy infrastructure of the Russian Federation. *Physics and Technology Informatics (CPT2020) : proceedings of the 8th International Conference, Pushchino, Moscow region, November 09–13, 2020*. Nizhny Novgorod: Autonomous non-profit organization in the field of information technologies “Research Center for Physical-Technical Informatics”, 2020, vol. 2, pp. 199–203. DOI: 10.30987/conferencearticle\_5fd755c02b5fa9.82083943 (In Russ.).
23. Letyagina, E. N., Perova, V. I., Yashin, S. N., Borisov, S. A. Study of innovative development of regional economy in the context of global challenges using artificial intelligence methods. *Economic Security*, 2021, vol. 4, no. 3, pp. 781–794. DOI: 10.18334/ecsec.4.3.112025 (In Russ.).
24. Avdzeyko, V. I., Karnyshev, V. I., Meshcheryakov, R. V. Patent Analysis. Identification of promising and breakthrough technologies. *Issues of Innovation Economy*, 2018, vol. 8, no. 1, pp. 79–90. DOI: 10.18334/vinec.8.1.38890 (In Russ.).
25. Mironov, N. A., Maryshev, E. A., Lukasheva, N. A., Divueva, N. A. Analysis of the experience of applying expert forecasting methods in the information system of the Federal Register of S&T experts for the development of S&T development directions. *Innovation and expertise: scientific works*, 2022, no. 1 (33), pp. 55–64 (In Russ.).
26. Shelomentseva, V. P., Nikitin, E. B., Bespalyy, S. V., Ifutina, E. A. Use of “foresight” method for forecasting scientific-technological and economic development of old industrial region (on the example of Pavlodar region of the Republic of Kazakhstan). *Bulletin of Omsk University. Series: Economics*, 2015, no. 4, pp. 242–248 (In Russ.).
27. Abdurakhmanov, A. M., Volodin, M. V., Zybin, E. Ju., Ryabchenko, V. N. Methods of electricity consumption forecasting in distribution networks (review). *Electrical Engineering*, 2016, vol. 3, no. 1, pp. 3–23 (In Russ.).
28. Shinkevich, A. I., Lubnina, A. A., Rayskiy, I. A. Trends of innovative development of manufacturing industries. *Izvestiya of Samara Scientific Center of the Russian Academy of Sciences*, 2021, vol. 23, no. 4 (102), pp. 51–56. DOI: 10.37313/1990-5378-2021-23-4-51-56 (In Russ.).
29. Beloshickiy A. V., Cherepovitsyn, A. E. Toward the formation of scenarios of sustainable development of the oilfield service industry. *Bulletin of South-Russian State Technical University (NPI). Series: Social and Economic Sciences*, 2022, vol. 15, no. 2, pp. 95–106. DOI 10.17213/2075-2067-2022-2-95-106 (In Russ.).
30. Patent RF 2722525. *Adaptive control of additive manufacturing processes in real time using machine learning*. E Mehr., T. Ellis, Dzh. Nun. Decl. 23.05.2018 ; publ. 01.06.2020 (In Russ.).
31. Patent RF 2663011. *System and method of performing well stimulation operation*. B. Leserf, D. Usoltsev, T. Poup. Decl. 01.12.2014 ; publ. 01.08.2018 (In Russ.).
32. Bakhmut, A. S., Gayvoronskaya, N. F., Grigoreva, G. V., Egorova, O. V., Kuznetsov, V. V. *Certificate of State Registration of Computer Program No. 2015612989 Russian Federation. System of methods for crop production development management (PS “Methods”)* : no. 2015610223 : decl. 12.01.2015 : publ. 27.02.2015 ; applicant Federal State Budgetary Scientific Institution “All-Russian Scientific Research Institute of Economics and Norms” (In Russ.).
33. Kampos, D. T., Ilin, I. V., Ilyashenko, O. Yu., Shchenikov E. M. *Certificate of State Registration of Computer Program No. 2023688714 Russian Federation. Program for finding parameters of the logistic equation* : no. 2023688468 : decl. 18.12.2023 : publ. 25.12.2023 ; applicant Federal State Autonomous Educational Institution of Higher Education “Peter the Great St. Petersburg Polytechnic University” (In Russ.).
34. Patent RF 2783863. *Method of complex optimization of power unit parameters*. M. V. Lifshic. Decl. 03.09.2021 ; publ. 21.11.2022 (In Russ.).
35. Vladimirov, I. S., Azarov, V. I., Borovik, A. V., Melnikov, I. A., Vatlin, A. N. On the specifics of methods of scientific and technological forecasting. *Innovative scientific research in the modern world: theory, methodology, practice : collection of scientific articles on the materials of VII International Scientific and Practical Conference, Ufa, January 31, 2022*. Ufa, Limited Liability Company “Scientific-Publishing Center ‘Vestnik Nauki’”, 2022, pp. 49–54 (In Russ.).
36. Dubishkin, E. A. Institutional mechanism of the technology market. *Public administration and development of Russia: challenges and prospects : collection of articles of the VI All-Russian Scientific and Practical Conference, Penza, November 25–26, 2022. By the scientific editorship of A. V. Yashin, A. A. Grachev*. Penza, Penza State Agrarian University, 2022, pp. 104–110 (In Russ.).
37. Petukhova, M. S. Theoretical bases of forecasting of scientific and technological development of crop production industry. *Priorities of the strategy strategy of scientific and technological development of Russia and ensuring the reproduction of innovation potential of higher education : proceedings of the All-Russian scientific conference, Izhevsk, February 27, 2019*. Editor-in-chief A. M. Makarov. Izhevsk, Publishing House “Udmurt University”, 2019, pp. 121–124 (In Russ.).
38. Vladimirov, I. S., Vopilov, A. Yu., Oreshkin, M. V., Borovik, A. V., Solovev, M. A. Perspective methods of scientific and technological forecasting. *Modern scientific research: theory, methodology, practice : collection of scientific articles on the materials of VII International Scientific and Practical Conference, Ufa, February 16, 2022*. Ufa, Limited Liability Company “Scientific-Publishing Center “Vestnik Nauki””, 2022, pp. 13–18 (In Russ.).
39. Glushchenko, V. V. Functions and roles of technological patterns in the management of technological and socio-economic development. *Kazakhstan Science Journal*, 2020, vol. 3, no. 3 (16), pp. 80–93 (In Russ.).
40. Timofeeva, N. Ju. Regression analysis and growth curves of economic series of dynamics in forecasting the development of agricultural producers in the Lipetsk region. *Proceedings of higher educational institutions. Volga region. Social Sciences*, 2017, no. 4 (44), pp. 145–157. DOI: 10.21685/2072-3016-2017-4-16 (In Russ.).
41. Boldyrevskiy, P. B., Kistanova, L. A. Models of forecasting the main indicators of innovation activity of industrial enterprises]. *Economic Analysis: Theory and Practice*, 2014, no. 29 (380), pp. 52–57 (In Russ.).
42. Grabchak, E. P. On approaches to determining the prospects for development of the electric power industry. *Problems of Market Economy*, 2021, no. 4, pp. 8–19. DOI: 10.33051/2500-2325-2021-4-8-19 (In Russ.).
43. Komkov, N. I. Integrated Forecasting of S&T Development: Experience and Lessons Learned. *Problems of forecasting*, 2014, no. 2 (143), pp. 3–17 (In Russ.).

44. Pudovkina, O. E. Technological priorities for the development of Russia's industrial sector in the context of deep penetration of digital technologies. *University Bulletin*, 2021, no. 9, pp. 74–80. DOI: 10.26425/1816-4277-2021-9-74-80 (In Russ.).
45. Patent RF 2686257. *Method and system of remote identification and prediction of incipient defects development*. S. A. Naumov, A. V. Krymskiy, M. V. Lifshits. Decl. 27.04.2018 ; publ. 24.04.2019 (In Russ.).
46. Degtyareva, V. V., Kamchatova, E. Ju., Kakaeva E. A. Application of the corporate foresight method to improve the innovation management system. *Modern Economy Success*, 2019, no. 2, pp. 105–108 (In Russ.).
47. Kornev, A. K., Maksimtsova, S. I. On improving the competitiveness of existing manufacturing industries. *Problems of forecasting*, 2019, no. 6 (177), pp. 83–94 (In Russ.).
48. Siptits, S. O., Romanenko, I. A., Ganieva, I. A., Evdokimova, N. E. Features of modeling and scenario analysis of the impact of long-term climatic changes on regional agrifood systems. *Achievements of science and technology of agro-industrial complex*, 2020, vol. 34, no. 4, pp. 5–9. DOI: 10.24411/0235-2451-2020-10401 (In Russ.).
49. Kostarev, A. S. Assessment of reserves of innovative development of coal-mining production association and the influence of organizational and economic relations on their use. *Proceedings of the Ural State Mining University*, 2020, no. 1 (57), pp. 208–217. DOI: 10.21440/2307-2091-2020-1-208-217 (In Russ.).
50. *World-class research centers have come up with solutions for industrial development*. Available at: <https://nauka.tass.ru/nauka/17000583> (accessed 14.04.2024) (In Russ.).
51. *Sber's AIJ 2023 Artificial Intelligence Conference. Text broadcast of the second day*. Available at: <https://tass.ru/online-conference/19352375> (accessed 14.04.2024) (In Russ.).
52. *Beeline has patented a new method of rearing*. Available at: [https://www.cnews.ru/news/line/2024-04-11\\_bilajn\\_zapatentoval\\_novyj](https://www.cnews.ru/news/line/2024-04-11_bilajn_zapatentoval_novyj) (accessed 14.04.2024) (In Russ.).
53. *Cyberhygiene for artificial intelligence systems in the industrial Internet of Things*. Available at: <http://vestnik-glonass.ru/news/intro/kibergigiena-dlya-sistem-iskusstvennogo-intellekta-v-promyshlennom-internete-veshchey> (accessed 14.04.2024) (In Russ.).
54. Nguen, T. V., Zyong, K. H. T., Kravets, A. G. Analysis and prediction of trends in the use of terms in computer science based on neural network models. *Bulletin of Computer and Information Technologies*, 2021, vol. 18, no. 2 (200), pp. 24–38. DOI: 10.14489/vkit.2021.02 (In Russ.).

Статья поступила в редакцию 08.05.2024; одобрена после рецензирования 27.05.2024; принята к публикации 14.07.2024.

The article was submitted 08.05.2024; approved after reviewing 27.05.2024; accepted for publication 14.07.2024.

## **МЕТОДЫ И СИСТЕМЫ ЗАЩИТЫ ИНФОРМАЦИИ, ИНФОРМАЦИОННАЯ БЕЗОПАСНОСТЬ**

УДК 004.896

### **МЕТОД ЗАЩИТЫ ОТ ИССЛЕДОВАНИЯ ЛОКАЛЬНОЙ ВЫЧИСЛИТЕЛЬНОЙ СЕТИ НА ОСНОВЕ РЕКОНФИГУРАЦИИ ТОПОЛОГИИ СЕТЕВОГО УРОВНЯ**

*Кушко Евгений Александрович*, Сибирский государственный университет науки и технологий, 660037, Российская Федерация, г. Красноярск, пр. им. газ. «Красноярский рабочий», 31, старший преподаватель, ORCID: 0000-0003-1290-7075, e-mail: evgeny.kushko@ya.ru

*Трофимычев Илья Игоревич*, Самарский государственный технический университет, 443100, Российская Федерация, г. Самара, ул. Молодогвардейская, 244, главный корпус, аспирант, ORCID: 0009-0002-9532-8924, e-mail: mr.trofimychev@mail.ru

В статье рассмотрена проблема защиты локальных вычислительных сетей от исследования. Авторами предложен метод защиты от исследования локальной вычислительной сети на основе реконфигурации топологии сетевого уровня. Данный метод направлен на ограничение длительности времени актуальности информации о топологии сетевого уровня локальной вычислительной сети, при котором злоумышленник не будет обладать полезной информацией о ней. В статье приведено алгоритмическое обеспечение предложенного метода, а также его математическая модель. При помощи математической модели дана оценка вероятности идентификации целевого узла при условии функционирования предложенного метода. Также авторами предложен алгоритм расчета длительности актуальности информации о топологии сетевого уровня, при котором вероятность идентификации целевого узла остается на прогнозируемом уровне. В результате действия предложенного метода злоумышленник вынужден атаковать локальную вычислительную сеть «вслепую», так как собранная им информация о сети утратила свою актуальность.

**Ключевые слова:** информационная безопасность, локальная вычислительная сеть, защита от исследования, технология защиты движущейся цели, распределенная инфраструктура ложных целей

### **LOCAL NETWORK EXPLORATION COUNTERING METHOD BASED ON NETWORK LAYER TOPOLOGY RECONFIGURATION**

*Kushko Evgenii A.*, Siberian State University of Science and Technology, 31 Krasnoyarsky Rabochy Ave., Krasnoyarsk, 660037, Russian Federation,

Senior Lecturer, ORCID: 0000-0003-1290-7075, e-mail: evgeny.kushko@ya.ru

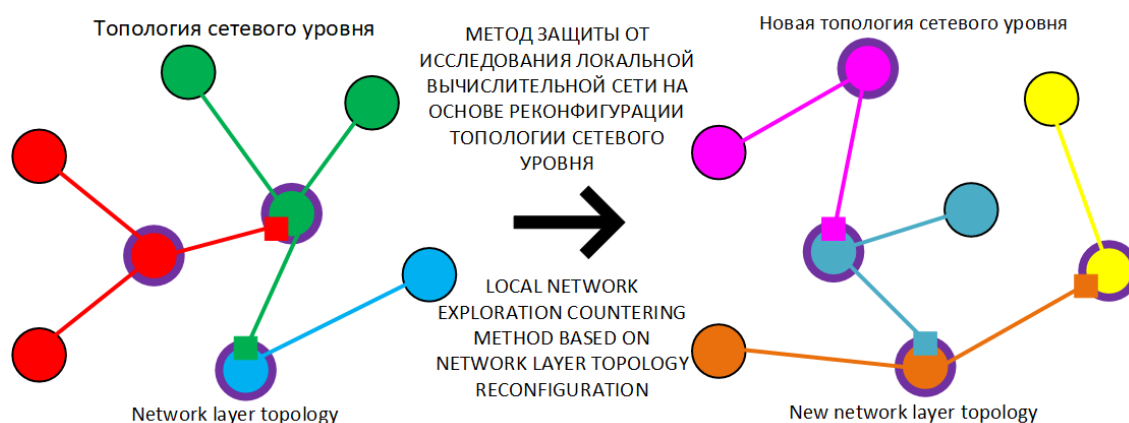
*Trofimychev Ilya I.*, Samara State Technical University, main building, 244 Molodogvardeiskaya St., Samara, 443100, Russian Federation,

graduate student, ORCID: 0009-0002-9532-8924, e-mail: mr.trofimychev@mail.ru

The paper considers the issue of protecting local computer networks from exploration. The authors propose a method for countering local network research based on network layer topology reconfiguration. The method is aimed at reducing local network topology relevance time so that an attacker does not have up-to-date information. The article provides algorithmic support and a mathematical model of the proposed method. Using the mathematical model, the authors estimate the probability of identifying the target node in the network protected by the proposed solution. The authors also propose an algorithm for calculating network topology information relevance time, ensuring the probability of identifying the target node to remain at the predicted level. As a result of the proposed method, the attacker is forced to go blind since the collected information has lost its relevance.

**Keywords:** information security, local area network, network scanning protection, moving target defense technology, distributed deception platform

## Graphical annotation (Графическая аннотация)

**ВВЕДЕНИЕ**

Исследование сети является неотъемлемой частью атаки злоумышленника. Основная цель исследования сети злоумышленником заключается в сборе и структуризации информации о сети и узлах, функционирующих в ней, таким образом выявляя потенциальные уязвимости. Пассивный сбор данных, различные техники сканирования, анализ служб и сервисов – все это используется для создания карты топологии сети, идентификации активных узлов и определения слабых мест локальной вычислительной сети. Эта информация является основой для построения возможных векторов атаки злоумышленником, именно она позволяет злоумышленникам создавать целенаправленные и сложные цепочки атак. Поэтому система защиты информации локальной вычислительной сети должна включать в себя механизмы по противодействию исследованию сети для того, чтобы потенциальные угрозы были предотвращены уже на ранних этапах атаки.

В настоящий момент для защиты локальной вычислительной сети от исследования рекомендуется:

- 1) сегментирование сети и построение дополнительных эшелонов защиты;
- 2) контроль передаваемого по сети трафика;
- 3) выявление паттернов сканирования;
- 4) ограничение определенной сетевой активности.

Однако существует риск того, что злоумышленник смог преодолеть данные меры. При этом злоумышленник в процессе исследования сети имеет неограниченное время для сбора данных, так как топология локальной вычислительной сети, её маршруты передачи, каналы взаимодействия, а также другие элементы статичны.

Для решения проблемы «статичности» исследователями в области информационной безопасности предложена технология защиты движущейся цели (moving target defense, MTD) [1]. В отличие от других технологий защиты, технология защиты движущейся цели привносит изменчивость в среду передачи. Эта изменчивость включает в себя регулярные изменения топологии сетевого уровня, которая включает в себя физические и логические адреса, маршруты передачи данных, параметры адресации и маршрутизации локальной вычислительной сети. Постоянное изменение топологии сетевого уровня позволяет технологии защиты движущейся цели повышать неопределенность для злоумышленника, тем самым повышая устойчивость сети к попыткам сканирования и пассивному наблюдению [2]. Технология защиты движущейся цели не ограничивает в действиях злоумышленника, а лишь изменяет информацию таким образом, чтобы на момент её сбора злоумышленником она перестала быть актуальной. В результате злоумышленник теоретически вынужден атаковать локальную вычислительную сеть «вслепую», так как собранная им информация о сети утратила свою актуальность. Однако решения по защите локальной вычислительной сети на основе технологии защиты движущейся цели имеют ряд существенных недостатков [3].

Кроме того, регуляторами Российской Федерации в области информационной безопасности рекомендуется реализация следующих мер защиты [4]:

1) создание (эмуляция) ложных информационных систем или их компонентов, предназначенных для обнаружения, регистрации и анализа действий нарушителей в процессе реализации угроз безопасности информации;

2) воспроизведение ложных и (или) скрывание истинных отдельных информационных технологий и (или) структурно-функциональных характеристик информационной системы или ее сегментов, обеспечивающее навязывание нарушителю ложного представления об истинных информационных технологиях и (или) структурно-функциональных характеристиках информационной системы.

Данные меры на практике реализуются при помощи распределенной инфраструктуры ложных целей (distributed deception platform, DDP) [5]. Решения данного класса направлены на реализацию воспроизведения ложных сущностей, на исследование и анализ которых злоумышленник потратит значительное количество ресурсов. Однако в процессе сбора данных злоумышленник не ограничен в исследовании истинной инфраструктуры, которая все также остается статичной. Но в защиту технологии DDP следует сказать, что инфраструктура ложных целей формируется так, чтобы быть более привлекательной для злоумышленника с точки зрения реализации им потенциальной атаки. В результате функционирования системы злоумышленник принимает неправильное решение, атакует ложный объект инфраструктуры и тем самым демаскирует себя. Однако существует вероятность того, что злоумышленник выберет истинный объект защиты, и оценить эту вероятность достаточно сложно, так как зависит она от многих факторов в каждом конкретном случае [6].

Поэтому с учетом вышесказанного разработка новых решений по защите локальной вычислительной сети от исследования злоумышленником является актуальной научно-технической задачей.

Целью исследования является ограничение длительности времени актуальности информации о топологии сетевого уровня локальной вычислительной сети, при котором злоумышленник не будет обладать полезной информацией при наступлении этапа осуществления атаки. Для достижения поставленной цели авторами предложен метод защиты от исследования локальной вычислительной сети на основе реконфигурации топологии сетевого уровня.

#### АЛГОРИТМИЧЕСКОЕ ОБЕСЕЧЕНИЕ МЕТОДА

Основная идея метода заключается в периодических изменениях топологии сетевого уровня локальной вычислительной сети, а именно следующих параметров: физические и логические адреса узлов сети, маршруты передачи данных. Также следует отметить, что рассматривается локальная вычислительная сеть, в которой не происходит изменений на физических и канальных уровнях в рамках структурной взаимосвязи узлов.

На рисунке 1 представлен алгоритм реконфигурации топологии сетевого уровня. Сеть управляется централизованно участником-сервером. Участник-сервер во время своего функционирования постоянно ожидает следующие события:

- 1) оповещения от участников взаимодействия;
- 2) истечение временного интервала текущей конфигурации сети.

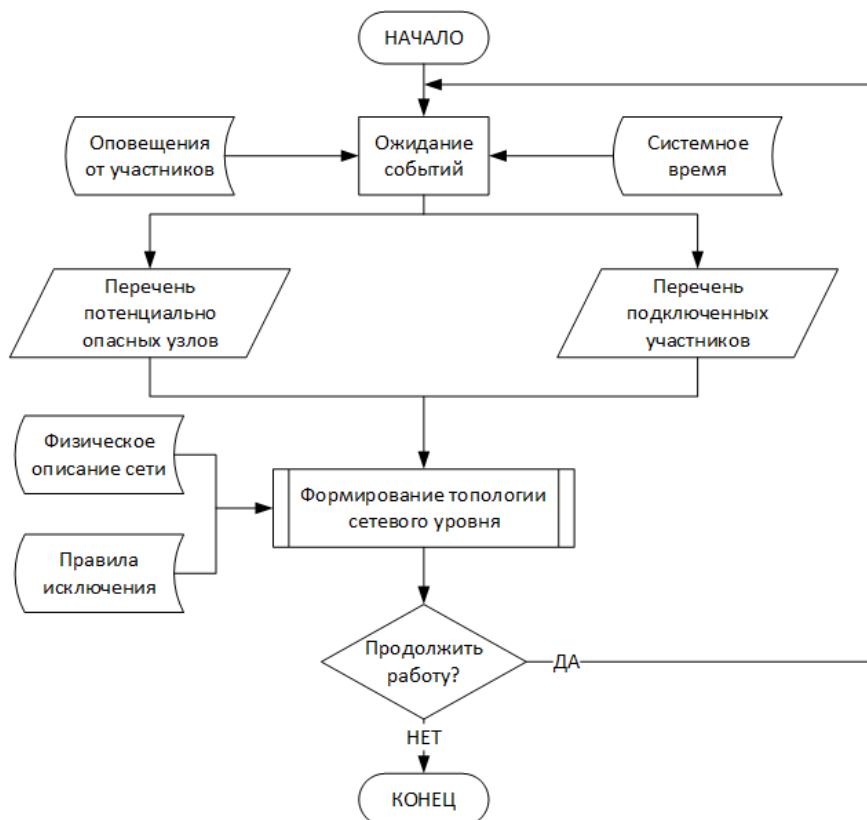


Рисунок 1 – Алгоритм реконфигурации топологии сетевого уровня

Также во время функционирования участник-сервер периодически посылает запросы участникам взаимодействия об их текущем статусе, чтобы таким образом удостовериться в том, что участник взаимодействия находится в сети и готов принимать новые конфигурации. В результате так формируется перечень подключенных участников взаимодействия участником-сервером.

В то же время некоторые из участников взаимодействия образуют распределенную инфраструктуру ложных целей сети. Когда злоумышленник использует активное сканирование для исследования сети, то узел детектирует такое событие и оповещает об этом участника-сервера. Участник-сервер принимает это оповещение, содержащее в себе информацию о сканирующем узле. При помощи этих оповещений участник-сервер формирует перечень потенциально опасных узлов. К потенциально опасным узлам могут быть применены определенные меры согласно заданным правилам исключения.

На основе сформированных перечней, правил исключения и физическом описании сети участник-сервер формирует топологию сетевого уровня в соответствии с алгоритмом на рисунке 2.

Физическое описание сети создает администратор локальной вычислительной сети и предоставляет его участнику-серверу. Данное описание должно включать в себя: перечень оборудования сетевой инфраструктуры, образующей каналы связи; перечень конечных устройств, участвующих во взаимодействии; описание физических подключений между конечными устройствами и сетевым оборудованием. Следует отметить то, что все участники взаимодействия должны быть достижимы участником-сервером. Кроме того, оборудование должно быть способно программно управляться через сеть, а участник-сервер должен иметь подключение к интерфейсам управления сетевым оборудованием.

При формировании топологии сетевого уровня участник-сервер формирует перечень участников взаимодействия. Для того чтобы локальная вычислительная сеть функционировала, должно быть как минимум два участника взаимодействия.

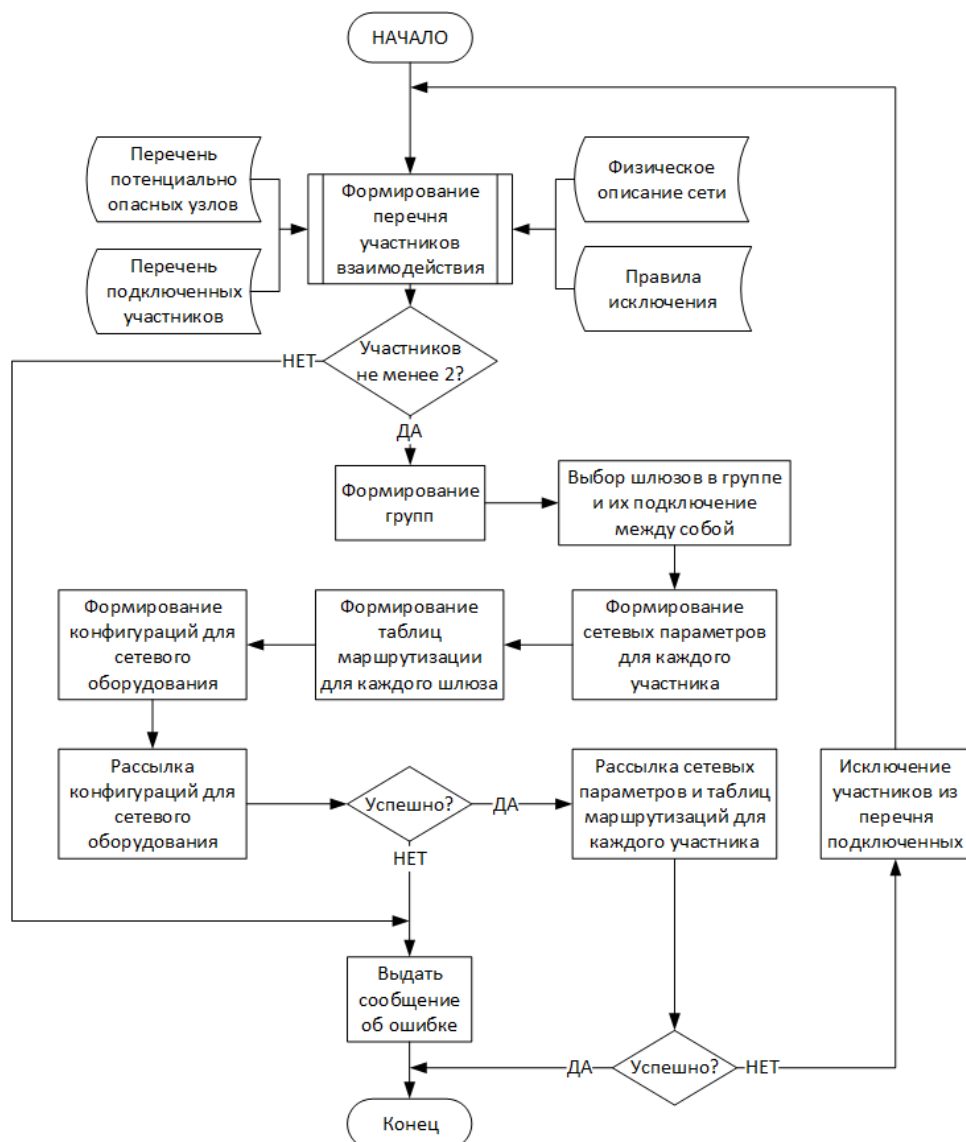


Рисунок 2 – Алгоритм формирования топологии сетевого уровня

Далее участник-сервер создает  $N$  групп, где  $N$  – это случайное число от 2 до  $M/2$ , а  $M$  – это количество участников взаимодействия. Каждому участнику взаимодействия случайно присваивается номер группы, который указывает на принадлежность к ней. Каждая группа вмещает в себя не более чем  $M/N$  участников. Все участники равномерно распределяются по группам. Те участники, которым не досталось места в какой-либо группе, помещаются в случайную группу, превышая заданное ограничение.

Каждая группа образует виртуальную локальную сеть. Внутри каждой группы один участник взаимодействия выбирается в качестве внутреннего участника-шлюза. Соответственно, весь трафик виртуальной вычислительной сети группы, к которой выбранный внутренний участник-шлюз принадлежит, будет проходить через него. Внутренний участник-шлюз выбирается случайным образом при каждой новой реконфигурации.

Для каждой группы выделяется своя метка виртуальной локальной сети и собственное адресное подпространство из адресного пространства всей сети, а каждому участнику взаимодействия – адрес в этом подпространстве. Соответствующие параметры (физический адрес, логический адрес, маска, адрес шлюза, метка виртуальной локальной сети) задаются для виртуального интерфейса сети каждого участника взаимодействия. Каждый участник-шлюз имеет два дополнительных виртуальных интерфейса, а обычный участник – только один.

Для того чтобы участники могли взаимодействовать между собой, участники-шлюзы последовательно-случайно соединяются между собой. Происходит это следующим образом. Все участники-шлюзы помещаются в массив, который случайно перемешивается. После чего каждый участник-шлюз вступает в группу участника-шлюза, следующего за ним в массиве. Параметры новой группы создаются участником-сервером для второго дополнительного интерфейса каждого участника-шлюза. Этот новый член группы становится внешним участником-шлюзом по отношению к новой группе.

Для передачи данных участнику-шлюзу необходимо знать, куда перенаправить сетевой пакет для достижения пункта назначения: внешнему участнику-шлюзу группы, в которой он является внутренним, либо внутреннему участнику-шлюзу группы, в которой он является внешним. Для формирования таблицы маршрутизации также используется массив подключения участников-шлюзов между собой. Массив снова проходится последовательно: сначала в одну сторону, а затем – в другую. Для каждого участника-шлюза в таблицу маршрутизации добавляется запись о других виртуальных локальных сетях, в которые входят другие участники-шлюзы и находятся в массиве за следующим участником-шлюзом относительно текущего. В обратную сторону – аналогичным образом.

Для каждого сетевого устройства формируются параметры по следующему принципу: для порта, к которому подключен участник взаимодействия, задается разрешение на прием-передачу трафика, помеченного метками тех групп, к которым подключен данный участник взаимодействия. Потенциально опасные узлы, для которых принято решение об исключении, в соответствии с правилами исключения могут быть изолированы от взаимодействия соответствующей конфигурацией.

Предполагается, что взаимодействие между узлами осуществляется при помощи доменных имен, поэтому участник-сервер при создании новой конфигурации сети также формирует список доменных имен и соответствующих им логических адресов в сети. Этот список также включается в конфигурацию для каждого участника взаимодействия. После чего участник-сервер рассылает конфигурации для сетевого оборудования через их интерфейсы управления. В случае неудачи участник-сервер оповещает об ошибке и прекращает свою работу. В случае успеха инициируется передача конфигурации для каждого участника взаимодействия. В случае неудачи узлы-участники, которые не сообщили об успешном приеме конфигурации, исключаются из перечня подключенных, а топология сетевого уровня реконфигурируется заново. Предполагается, что для управления узлами выделена отдельная виртуальная локальная сеть. Для управления сетевым оборудованием выделена отдельная сеть с ограниченным доступом.

На рисунке 3 приведен алгоритм процесса формирования перечня участников взаимодействия. Сначала выбирается узел из физического описания сети, который должен быть включен во взаимодействие. Далее он проверяется на наличие в перечне потенциально опасных узлов. В случае если узел является потенциально опасным, он может быть исключен из взаимодействия в соответствии с правилами исключения. Узел может быть исключен из взаимодействия следующими способами: 1) отключен от локальной сети; 2) перемещен в отдельную изолированную подсеть, например, для дальнейшего изучения его поведения. Кроме того, узел может быть исключен: 1) на определенное время, т. е. узел может быть включен во взаимодействие через истечение заданного временного интервала; 2) навсегда. Правила исключения определяются администратором.

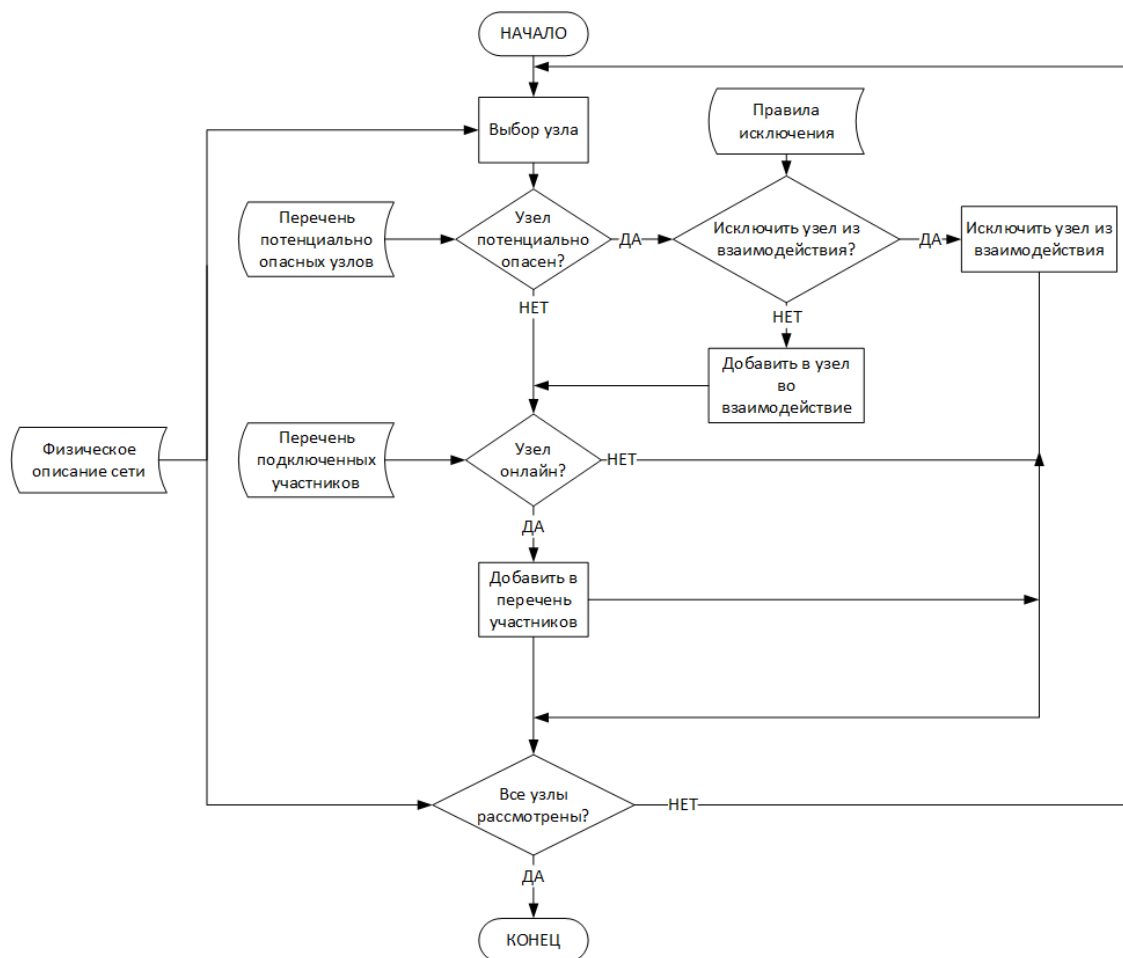


Рисунок 3 – Алгоритм формирования перечня участников взаимодействия

Если узел в сети, т. е. он достижим для участника-сервера и готов принимать конфигурации, то узел добавляется в перечень участников взаимодействия. Затем рассматривается следующий узел из физического описания и т. д.

Такие постоянные реконфигурации топологии сетевого уровня должны привести к тому, что злоумышленник не может обладать долгосрочной полезной информацией об организации локальной вычислительной сети, так как узлы при каждой новой итерации метода имеют новые адреса, как физические, так и логические. Кроме того, узлы каждую новую итерацию распределены по новым виртуальным локальным сетям, отличным от предыдущих.

Далее необходимо оценить вероятность идентификации узла при каждой итерации алгоритма, а также убедиться, что граф при перестроении остается связным при отсутствии воздействия злоумышленника на каналы связи.

### МАТЕМАТИЧЕСКАЯ МОДЕЛЬ

Пусть у нас есть некоторая локальная вычислительная сеть, объединяющая внутри себя  $N$  устройств. Тогда логический уровень данной сети можно представить в виде множества вершин  $V$  мощностью  $N$ . С момента начала эксперимента в начале каждой итерации реконфигурации будут выполняться следующие действия:

1. Случайным равновероятным образом выбирается число групп разбиения  $M$ , принадлежащее промежутку от 2 до  $N/2$  включительно. Следует также отметить, что на протяжении всего эксперимента для выбора случайного элемента используется равномерное распределение.

2. Все вершины из множества  $V$  разделяются на  $M$  множеств  $S_i$ .

3. Каждое множество  $S_i$  ( $i = 1, \dots, M$ ) дополняется ребрами  $E_{S_i}$  таким образом, чтобы получить граф-звезду  $G_{S_i}(S_i, E_{S_i})$ . Обозначим те вершины, степень которой больше 1, центральными (или центрами), остальные вершины – крайними. Отметим, что центральная вершина в каждом множестве  $S_i$  выбирается случайным образом.

4. Составляется граф  $G(V, E)$ , получаемый при помощи объединения графов  $G_{S_i}$  путем соединения их между собой при помощи множества дополнительных ребер  $e'_i$  таким образом, что ребро  $e'_i$  связывает центральную вершину графа  $G_{S_i}$  с центральной вершиной графа  $G_{S_{i+1}}$ .

Для того чтобы отразить процессы передачи информации между узлами графа  $G$ , введем множество  $W = \{w_i\}$  направленных каналов передачи информации между вершинами, каждый элемент которого представляет из себя последовательное множество вершин, через которые проходит кратчайший маршрут от одного заранее заданного узла к другому (включая начальный и конечный узел). Следует также отметить, что в силу особенностей строения графа  $G$  каждой паре вершин  $(v_i, v_j)$  будет соответствовать единственный маршрут во множестве  $W$ .

Данная модель подвергается атаке со стороны злоумышленника, включающей в себя две возможные глобальные стратегии [7].

Первая, наиболее простая, включает в себя активное сканирование всех возможных адресов в сети без непосредственного внедрения в узлы. В таком случае, если искомой целью злоумышленника является один конкретный узел  $V_i$ , а узлы, в свою очередь, распределены случайно по всем адресам сети, то ему придется придерживаться следующей стратегии.

Во-первых, прежде чем проводить попытку идентификации, злоумышленнику будет необходимо узнать, за какими адресами в сети находятся функционирующие устройства, для чего он будет проводить активное сканирование.

Во-вторых, сам процесс активного сканирования сети будет занимать (1):

$$t_{\text{ск.}} = N_{\text{адр}} \cdot t_{\text{ска.}} \quad (1)$$

где  $t_{\text{ска.}}$  – время одного сканирования одного адреса;

$N_{\text{адр}}$  – количество узлов в сети.

В-третьих, после завершения процесса сканирования злоумышленник получит список проверенных адресов сети мощностью  $N$ , после чего начнет процесс идентификации узлов. В процессе идентификации вероятность успешного опознания нужного узла  $v_i$  (что в нашем случае будет означать успешную компрометацию сети) будет составлять (2):

$$N_{\text{поп.}}/N, \quad (2)$$

где  $N_{\text{поп.}}$  – количество попыток идентификации со стороны злоумышленника.

Таким образом, наилучшей стратегией защиты в данном случае является определение критического количества попыток идентификации  $N_{\text{поп.кр.}}$  через верхнюю границу вероятности обнаружения искомого узла  $p_{\text{кр.}}$  при помощи соотношения (3):

$$N_{\text{поп.кр.}} = N \cdot p_{\text{кр.}} \quad (3)$$

А задание периода автоматической реконфигурации топологии сетевого уровня равно (4):

$$\tau = N_{\text{поп.кр.}} \cdot t_{\text{поп.}} + t_{\text{ск.}} = p_{\text{кр.}} \cdot N \cdot t_{\text{поп.}} + t_{\text{ск.}} \quad (4)$$

где  $t_{\text{поп.}}$  – время, требуемое злоумышленнику на реализацию одной попытки опознания узла.

При отсутствии введения такого среднего времени на опознание нужного узла будет занимать в среднем (5):

$$t_{\text{оп.}} = (N/2) \cdot t_{\text{поп.}} + N_{\text{адр}} \cdot t_{\text{ска.}} \quad (5)$$

В-четвертых, заметим, что злоумышленник может оптимизировать свою стратегию поиска с точки зрения скорости обнаружения необходимого узла, если будет производить попытку идентификации сразу после обнаружения очередного адреса, соответствующего устройству в сети. В таком случае среднее время на компрометацию сети составит (6):

$$t_{\text{оп.опт.}} = 0.5(N \cdot t_{\text{поп.}} + N_{\text{адр}} \cdot t_{\text{ска.}}) \quad (6)$$

А период автоматической реконфигурации будет задаваться как (7):

$$\tau = p_{\text{кр.}}(N \cdot t_{\text{поп.}} + N_{\text{адр}} \cdot t_{\text{ска.}}) \quad (7)$$

Вторая стратегия злоумышленника исходит из предположения, что активное сканирование сети вызовет срабатывание системы обнаружения сканирования и приведет к реконфигурации модели, вследствие чего злоумышленнику приходится использовать тактику внедрения в сеть и пассивного анализа трафика с целью выявления адресов, по которым потенциально может располагаться целевой узел. В результате предполагаемая стратегия правонарушителя будет выглядеть следующим образом.

С точки зрения злоумышленника, исследуемая локальная вычислительная сеть представляет собой множество вершин  $V_{\text{зл.}}$ , соответствующее множеству адресов, известных ему и по которым распределены узлы – вершины графа  $G$ . Отметим, что поскольку модель реконфигурации каждый раз случайно распределяет узлы по адресам сети без учета информации о предыдущем распределении, то в начале каждой итерации реконфигурации  $V_{\text{зл.}}$  будет соответствовать пустому множеству  $\emptyset$ . В процессе анализа сети для того, чтобы добавить адрес некоторой вершины  $v_i$  в множество  $V_{\text{зл.}}$ ,

злоумышленнику необходимо получить информацию о том, по какому адресу в сети располагается данный узел. Для этого злоумышленник в начале каждой итерации способен «внедриться» в некоторую случайную вершину  $v_{зр.}$ , после чего он будет способен перехватывать все сетевые пакеты, включенные в маршруты, проходящие через «зараженную» вершину. После сбора необходимой информации об адресах узлов в сети у злоумышленника будет лишь одна попытка идентификации необходимой ему вершины среди известных ему адресов, так как после этого события система обнаружения сканирования выявит злоумышленника, и он будет изолирован от сети.

Само множество вершин  $V_{зл.}$  в конечном итоге представляет собой те вершины, для которых  $v_{зр.}$  является элементом маршрута передачи информации (при этом неважно, конечным или промежуточным пунктом) и соответственно выражается через множество маршрутов  $W$  (8):

$$V_{зл.} = \{x \in w_i: v_{зр.} \in w_i, w_i \in W\}. \quad (8)$$

Обозначим множество  $V_{зл.}$  при выборе определенной вершины  $v_i$  в качестве зараженной вершины как  $V_{зл.}(v_i)$ . В таком случае среднее количество известных злоумышленнику вершин  $N_i$  будет равно (9):

$$N_i = \frac{1}{N} \sum_{v_i \in V} |V_{зл.}(v_i)|. \quad (9)$$

Общая вероятность компрометации сети в модели с пассивным наблюдением складывается последовательно из двух событий:

1. Злоумышленник выбрал зараженную вершину таким образом, что через нее проходит канал общения целевой вершины с другими.

2. Злоумышленник смог из всего множества известных ему вершин выбрать целевую.

Поскольку зараженная и целевая вершина выбираются случайным образом, то вероятность первого события (10):

$$P_1 = N_i/N, \quad (10)$$

где  $N_i$  – среднее число известных злоумышленнику вершин;

$N$  – общее количество вершин в сети.

Поскольку злоумышленник лишь знает адреса сети, которым соответствуют узлы сети, но не распределение самих узлов, то выбирать из известных ему придется опять случайным образом и шанс второго события (11):

$$P_2 = 1/N_i. \quad (11)$$

Таким образом, общая вероятность компрометации сети при модели пассивного наблюдения равна (12):

$$P_k = P_1 \cdot P_2 = \frac{N_i}{N} \cdot \frac{1}{N_i} = \frac{1}{N}. \quad (12)$$

Основной характеристикой, влияющей на защищенность сети как при пассивном наблюдении, так и при активном сканировании, является мощность множества  $V$ . Проведем более подробный анализ на примере стратегии активного сканирования сети и оценим оптимальный период реконфигурации топологии сетевого уровня  $\tau$ .

Для того чтобы оценить период реконфигурации, необходимо определить время  $t_{зл.}$ , за которое злоумышленник сможет собрать всю информацию о локальной вычислительной сети. Модель анализа локальной вычислительной сети злоумышленником состоит из следующих шагов:

Начало отсчета времени совпадает с моментом заражения узла злоумышленником.

Каждый узел  $v_i$  раз в определенный период времени  $\tau$  отправляет пакет данных равновероятно по одному из заранее определенных каналов связи.

При попадании пакета данных, содержащего информацию о незараженной вершине, к злоумышленнику он будет вынужден потратить время обработки  $t_{обр.}$  на каждую новую вершину, о которой он узнает из пакета данных.

Пока злоумышленник обрабатывает один пакет данных, он не может начать обрабатывать следующий, но может сохранять их для последующей обработки. Также злоумышленник не тратит время для обработки уже выявленного адреса.

Введем функцию  $\rho(v_i)$ , которая означает вероятность того, что пакет данных, содержащий информацию о вершине  $v_i$  попадет в зараженную вершину при рассылке пакетов. В таком случае среднее время, за которое информация о вершине  $v_i$  попадет в зараженную вершину, равно  $1/\rho(v_i)$ , при этом  $\rho(v_i)$  составляет (13):

$$\rho(v_i) = 1 - \prod_{v_j \in V} \left(1 - \frac{\alpha(v_i, v_j)}{\beta(v_j)}\right), \quad (13)$$

где  $\beta(v_j)$  – количество маршрутов из вершины  $v_j$ ;

$\alpha(v_j, v_i)$  – количество маршрутов из вершины  $v_j$ , включающих в себя вершину  $v_i$ .

Пусть  $t_1, t_2, \dots, t_n$  – неубывающая последовательность, где  $t_i$  – среднее время ожидания, за которое злоумышленник будет узнавать об определенной вершине. Тогда можно задать рекуррентно последовательность моментов времени, во время которых злоумышленник будет узнавать о новых узлах.

Пусть  $F_i$  – момент времени, в который злоумышленник получит информацию об  $i$ -том узле,  $ReLU(x)$  – Rectified Linear Unit (выпрямленная линейная функция) [8]. Тогда получим следующий набор (14).

Таким образом, при заданном графе  $G(V, E)$  и множестве маршрутов передачи сообщений  $W$  можно построить график получения злоумышленником информации о существующих вершинах и в соответствии с ним (принимая  $F_{N_i} = \tau$ ) произвести оценку критического периода реконфигурации топологии сетевого уровня.

$$\begin{aligned} F_1 &= t_1 + t_{\text{обр.}}, \\ F_2 &= F_1 + Relu(t_2 - F_1) + t_{\text{обр.}}, \\ &\vdots \\ F_i &= F_{i-1} + Relu(t_i - F_{i-1}) + t_{\text{обр.}} \end{aligned} \quad (14)$$

## РЕЗУЛЬТАТЫ

Результаты математического моделирования показывают, что вероятность идентификации целевого узла злоумышленником как в случае активного сканирования, так и в случае пассивного наблюдения за трафиком сети, составляет  $1/N$ , где  $N$  – это количество узлов сети, на каждой итерации работы предложенного алгоритма. Длительность актуальности текущей конфигурации топологии сетевого уровня, при которой вероятность идентификации целевого узла злоумышленником остается равной  $1/N$ , вычисляется для каждой инфраструктуры индивидуально и зависит она от нескольких факторов: количество узлов сети и характер распределения трафика между узлами. Соответственно, данную величину можно корректировать путем добавления дополнительных ложных узлов в сеть, которые будут генерировать дополнительный ложный трафик.

В работе [9] приведены некоторые практические оценки функционирования предложенного алгоритма. Апробация осуществлялась с использованием технологии VXLAN и может быть применена в системах защиты виртуальных машин.

## ЗАКЛЮЧЕНИЕ

Авторами предложен метод защиты от исследования локальной вычислительной сети на основе реконфигурации топологии сетевого уровня. Действие данного метода направлено на ограничение длительности времени актуальности информации о топологии сетевого уровня локальной вычислительной сети, при котором злоумышленник не будет обладать полезной информацией при наступлении этапа осуществления атаки.

Независимо от того, какой стратегии (активное сканирование, пассивное наблюдение, гибридная) будет придерживаться злоумышленник в процессе исследования сети, вероятность обнаружения целевого узла будет составлять  $1/N$ , где  $N$  – это количество узлов сети, что подтверждено при помощи математической модели метода. То есть злоумышленник вынужден выбирать целевой узел случайным образом на каждой итерации алгоритма, который лежит в основе предложенного метода.

Длительность актуальности текущей конфигурации топологии сетевого уровня зависит от количества узлов сети и характера распределения трафика между узлами. Поэтому длительность необходимо оценивать для каждой инфраструктуры индивидуально. Шаги и формулы расчета данной величины приведены в разделе, посвященном математической модели данной статьи.

## Список источников

1. Sushil, Jajodia. Moving target defense: creating asymmetric uncertainty for cyber threats / Sushil Jajodia, Anup K. Ghosh, Vipin Swarup, Cliff Wang, X. Sean Wang. – Springer. – 2011. – Vol. 54. – 184 p.
2. Cheng, Lei. Moving target defense techniques: A survey / Cheng Lei, Hong-Qi Zhang, Jing-Lei Tan, Yu-Chen Zhang, Xiao-Hu Liu // Security and Communication Networks. – 2018. – Vol. 2018.
3. Sailik, Sengupta. A survey of moving target defenses for network security / Sailik Sengupta, Ankur Chowdhary, Abdulhakim Sabur, Adel Alshamrani, Dijiang Huang, Subbarao Kambhampati // IEEE Communications Surveys & Tutorials. – 2020. – Vol. 22, № 3. – P. 1909–1941.
4. Марков, А. С. Организационно-технические проблемы защиты от целевых вредоносных программ типа Stuxnet / А. С. Марков, А. А. Фадин // Вопросы кибербезопасности. – 2013. – № 1 (1). – С. 28–36.
5. Путято, М. М. Исследование применения технологии deception для предотвращения угроз кибербезопасности / М. М. Путято, А. С. Макарян, Ш. М. Чич, В. К. Маркова // Прикаспийский журнал: управление и высокие технологии. – 2020. – № 4 (52). – С. 85–98.
6. Devika, Jay. Deception technology based intrusion protection and detection mechanism for digital substations: A Game theoretical approach / Devika Jay // IEEE Access. – 2023. – Vol. 11. – P. 53301–53314.
7. Tundis, A. A review of network vulnerabilities scanning tools: Types, capabilities and functioning / A. Tundis, W. Mazurczyk, M. Mühlhäuser // Proceedings of the 13th international conference on availability, reliability and security. – 2018. – P. 1–10.

8. Hara, K. Analysis of function of rectified linear unit used in deep learning / K. Hara, D. Saito, H. Shouno // 2015 international joint conference on neural networks (IJCNN). – 2015. – P. 1–8.

9. Кушко, Е. А. Организация защищенного обмена внутри программно-управляемой локальной сети / Е. А. Кушко, Н. Ю. Пароткин, В. В. Золотарев // Вестник СибГУТИ. – 2023. – Т. 17, № 4. – С. 62–73.

#### References

1. Sushil, Jajodia, Anup, K. Ghosh, Vipin, Swarup, Cliff, Wang, X. Sean, Wang. *Moving target defense: creating asymmetric uncertainty for cyber threats*. Springer, 2011, vol. 54. 184 p.

2. Cheng, Lei, Hong-Qi, Zhang, Jing-Lei, Tan, Yu-Chen, Zhang, Xiao-Hu, Liu. Moving target defense techniques: A survey. *Security and Communication Networks*, 2018, vol. 2018.

3. Sailik, Sengupta, Ankur, Chowdhary, Abdulhakim, Sabur, Adel, Alshamrani, Dijiang, Huang, Subbarao, Kambhampati. A survey of moving target defenses for network security. *IEEE Communications Surveys & Tutorials*, 2020, vol. 22, no. 3, pp. 1909–1941.

4. Markov, A. S., Fadin, A. A. Organizational and technical problems of protection against targeted malware such as Stuxnet. *Cybersecurity Issues*, 2013, no. 1 (1), pp. 28–36 (In Russ.).

5. Putyato, M. M., Makaryan, A. S., Chich, Sh. M., Markova, V. K. Research on the use of deception technology to prevent cybersecurity threats. *Caspian Journal: Control and High Technologies*, 2020, no. 4 (52), pp. 85–98 (In Russ.).

6. Devika, Jay. Deception technology based intrusion protection and detection mechanism for digital substations: A Game theoretical approach. *IEEE Access*, 2023, vol. 11, pp. 53301–53314.

7. Tundis, A., Mazurczyk, W., Mühlhäuser, M. A review of network vulnerabilities scanning tools: Types, capabilities and functioning. *Proceedings of the 13th international conference on availability, reliability and security*, 2018, pp. 1–10.

8. Hara, K., Saito, D., Shouno, H. Analysis of function of rectified linear unit used in deep learning. *2015 international joint conference on neural networks (IJCNN)*, 2015, pp. 1–8.

9. Kushko, E. A., Parotkin, N. Yu., Zolotarev, V. V. Ensuring secure data exchange in software-defined local network. *Bulletin of the Siberian State University of Telecommunications and Information Science*, 2023, vol. 17, no. 4, pp. 62–73 (In Russ.).

Статья поступила в редакцию 02.05.2024; одобрена после рецензирования 27.05.2024; принята к публикации 13.06.2024.

The article was submitted 02.05.2024; approved after reviewing 27.05.2024; accepted for publication 13.06.2024.

УДК 004.056.53

**МЕТОДИКА ОЦЕНКИ ЭФФЕКТИВНОСТИ  
СРЕДСТВ ЗАЩИТЫ ОТ DOS И DDOS-АТАК**

*Демин Николай Олегович*, Астраханский государственный университет имени В. Н. Татищева, 414056, Российская Федерация, г. Астрахань, ул. Татищева, 20а,  
аспирант, ORCID: 0009-0000-7404-3597, dnicolas1988@gmail.com

*Демина Раиса Юрьевна*, Астраханский государственный университет имени В. Н. Татищева, 414056, Российская Федерация, г. Астрахань, ул. Татищева, 20а,  
кандидат технических наук, доцент, ORCID: 0009-0009-1615-5641, e-mail: raisa.demina.91@mail.ru

*Ажмухамедов Искандар Маратович*, Астраханский государственный университет имени В. Н. Татищева, 414056, Российская Федерация, г. Астрахань, ул. Татищева, 20а,  
доктор технических наук, профессор, ORCID: 0000-0001-9058-123X, e-mail: aim\_agtu@mail.ru

*Меркулова Аlesia Михайловна*, Астраханский государственный университет имени В. Н. Татищева, 414056, Российская Федерация, г. Астрахань, ул. Татищева, 20а,  
старший преподаватель, ORCID: 0009-0004-6258-6067, e-mail: alesya\_m-d@mail.ru

В рамках статьи была освещена проблема DDoS-атак в современном мире. Участвовавшее количество атак существенно повысило спрос на средства защиты. Для повышения качества средств защиты от DDoS-атак необходимо проверять их не на удобных и простых сценариях, а на максимально сложных, учитывающих многие факторы, сигнатуры которых не содержатся в базе средства защиты. В рамках статьи предлагается математическая модель оценки средства защиты от DDoS-атак, которая может использоваться не только для непосредственной оценки средств защиты, но и для оценки успешности атаки, которая, в свою очередь, будет использоваться для создания новых сценариев с использованием генетических алгоритмов.

**Ключевые слова:** информационная безопасность, оценка эффективности, межсетевой экран, средство защиты, DDoS-атаки, отказ в обслуживании

**METHODOLOGY FOR EVALUATING THE EFFECTIVENESS  
OF PROTECTION AGAINST DOS AND DDOS-ATTACKS**

*Demin Nikolay O.*, Astrakhan Tatishchev State University, 20a Tatishchev St., Astrakhan, 114056, Russian Federation,  
postgraduate student, ORCID: 0009-0000-7404-3597, e-mail: dnicolas1988@gmail.com.

*Demina Raisa Yu.*, Astrakhan Tatishchev State University, 20a Tatishchev St., Astrakhan, 114056, Russian Federation,  
Cand. Sci. (Engineering), Associate Professor, ORCID: 0009-0009-1615-5641, e-mail: raisa.demina.91@mail.ru.

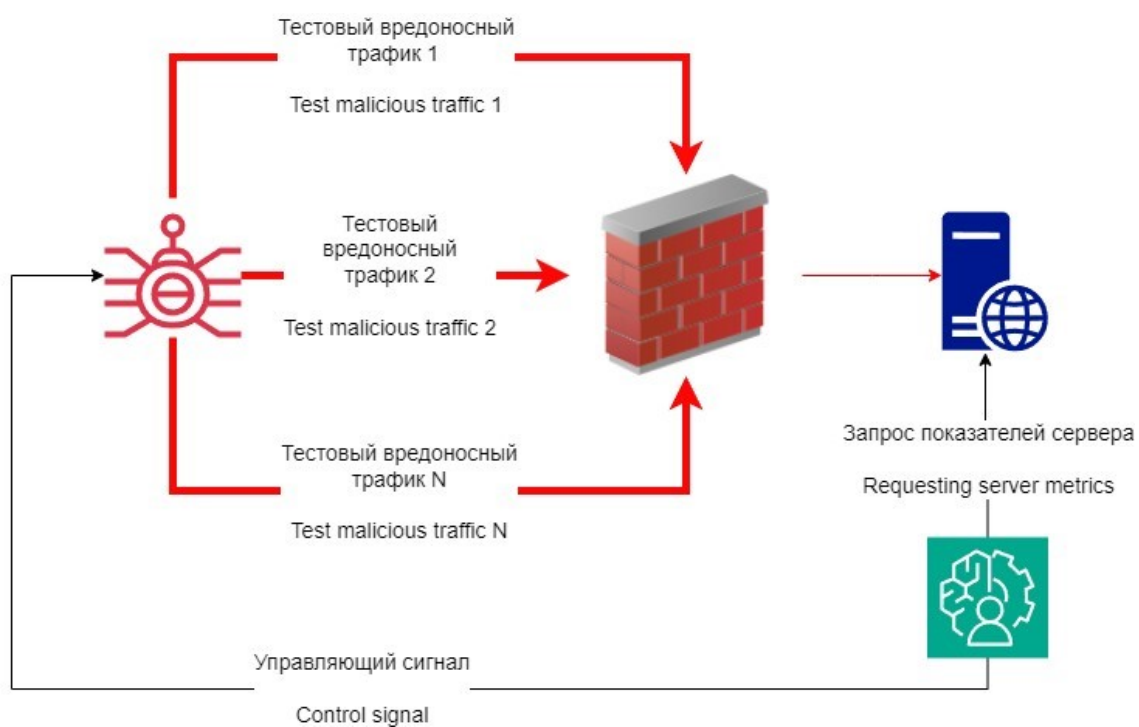
*Azhmuhamedov Iskandar M.*, Astrakhan Tatishchev State University, 20a Tatishchev St., Astrakhan, 114056, Russian Federation,  
Doct. Sci. (Engineering), Professor, ORCID: 0000-0001-9058-123X, e-mail: aim\_agtu@mail.ru.

*Merkulova Alesya M.*, Astrakhan Tatishchev State University, 20a Tatishchev St., Astrakhan, 114056, Russian Federation,  
Senior Lecturer, ORCID: 0009-0004-6258-6067, e-mail: alesya\_m-d@mail.ru

The article highlighted the problem of DDoS-attacks in the modern world. The increased number of attacks has significantly increased the demand for protective equipment. To improve the quality of DDoS protection tools, it is necessary to test them not on convenient and simple scenarios, but on the most complex ones, taking into account many factors whose signatures are not contained in the database of the protection tool. Within the framework of the article, a mathematical model for evaluating a means of protection against DDoS-attacks is proposed, which can be used not only for direct evaluation of means of protection, but also for evaluating the success of an attack, which, in turn, will be used to create new scenarios using genetic algorithms.

**Keywords:** information security, efficiency assessment, firewall, security tool, DDoS-attacks, denial of service

## Graphical annotation (Графическая аннотация)

**ВВЕДЕНИЕ**

DoS и DDoS-атаки являются одним из наиболее распространённых механизмов нарушения работоспособности информационных систем организаций и предприятий, особенно расположенных на территории Российской Федерации.

В первом полугодии 2022 г., согласно отчету Ростелеком Солар [1], был зафиксирован аномальный всплеск числа DDoS-атак (рис. 1). В марте 2022 г. на 30 % возрос спрос на межсетевые экраны, в чей функционал входит отражение межсетевых атак.



Рисунок 1 – Динамика числа DDoS-атак в 2022 г.

Снижение интенсивности DDoS-атак в конце 2022 г. связано с переключением внимания злоумышленников на атаки иного вида, а не с тем фактом, что было выработано универсальное, безусловно, эффективное решение обеспечения доступности различных информационных систем.

Согласно упомянутому выше отчету, 30 % атак пришлось на госсектор. Можно сделать вывод, что DDoS-атаки стали не просто способом получить преимущество в недобросовестной конкурентной борьбе, но и инструментом демонстрации политической позиции, используемым для демонстрации силы, поскольку факт отказа доступности информационного ресурса скрыть невозможно, в отличие от некоторых других успешных атак (компрометации конфиденциальных данных и т. п.).

DDoS-атаки в XXI в. стали спутниками политических и военных конфликтов. Аналитики CloudFlare сообщают о DDoS-атаках, нацеленных на информационные системы оповещения населения 7 октября 2023 г. в Израиле (рис. 2). «Первоначальная атака достигла пика в 100 тысяч запросов в секунду и продолжалась десять минут. Сорок пять минут спустя произошла вторая, гораздо более мощная атака, достигшая пика в 1 миллион запросов/с. Это длилось шесть минут. Дополнительные более мелкие DDoS-атаки продолжали поражать веб-сайты в последующие часы» [2].

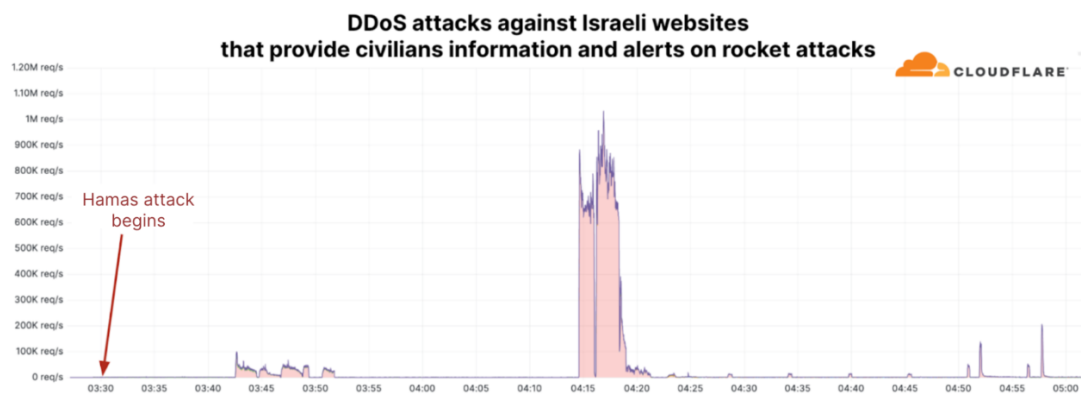


Рисунок 2 – Динамика DDoS-атак на информационные системы Израиля

В 2023 г. проблема противодействия DDoS-атакам в Российской Федерации также продолжает быть актуальной. Согласно опубликованным выводам компании «Гарда», «Россия оказалась на 9-м месте в мире по общему количеству зарегистрированных DDoS-атак. Во II квартале 2023 г. на российскую ИТ-инфраструктуру пришлось 2,3 % таких атак от общего их числа в мире, что вдвое больше, чем было в I квартале этого года» [3].

Вышесказанное позволяет сделать вывод о том, что DDoS-атаки стали «классическим» инструментом оказания влияния на наиболее критически значимые сферы любого государства. Кроме того, необходимо учесть, что DDoS-атаки могут быть организованы довольно простым способом (хотя, конечно, злоумышленники используют и сложные сценарии атак) и могут быть «развернуты» против любого объекта инфраструктуры за минимальное время после получения соответствующей команды. Для лиц, ответственных за обеспечение информационной безопасности на предприятии, актуален вопрос подбора наиболее подходящего для них средства межсетевое экранирования. Российский рынок средств защиты за последние годы получил мощный толчок и на нем конкурируют многие производители, и бесспорного абсолютного лидера, который мог бы подойти каждому, нет.

В связи с вышесказанным было принято решение предложить методику оценки эффективности средств межсетевое экранирования (СМЭ) с учетом результатов имитирования различного рода DDoS-атак.

#### АНАЛИЗ СУЩЕСТВУЮЩИХ ПОДХОДОВ К ОЦЕНКЕ ЭФФЕКТИВНОСТИ СРЕДСТВ МЕЖСЕТЕВОЙ ЗАЩИТЫ

Анализ литературы [4–6] показал, что в вопросе оценки эффективности средств межсетевой защиты исследователи выделяют несколько основных подходов:

- 1) методы, основанные на формализации процесса обеспечения ИБ и анализе построенных моделей (вероятностные, графовые модели и пр.);
- 2) методы, основанные на анализе результатов тестирования;
- 3) комбинированные методы.

При более близком рассмотрении методов 2 и 3 типов было выявлено, что какая бы математическая модель ни использовалась для оценки эффективности, в конечном итоге все равно всё сводится к результатам нагрузочного тестирования, которые, в свою очередь, зависят от разработанных

скриптов с нагрузкой. Если нагрузочные тесты будут слишком простыми для современных средств защиты или вычислительные мощности, потраченные на тестирование, будут недостаточными, то полученная оценка средств защиты нерелевантна, поскольку все средства защиты успешно пройдут тестирование. Нагрузочные тесты часто строятся по определенным, широко известным шаблонам. И разработчикам средств защиты эти шаблоны хорошо известны. Разработчики средств защиты тщательно отслеживают тенденции в вопросе проведения DDoS-атак и совершенствуют механизмы их обнаружения и отражения. Таким образом, встает вопрос об автоматической генерации сценариев для нагрузочного тестирования с использованием генетических алгоритмов.

Такой подход рассматривается в диссертационном исследовании Н. И. Синадского [7] в контексте проблемы разработки стендов по информационной безопасности для студентов. В статье [8] продемонстрированы экспериментальные данные, подтверждающие перспективность данного направления исследования. У авторов получилось с помощью генеративных нейронных сетей создать новые эффективные сценарии для нагрузочного тестирования.

Рассмотрим в комплексе вопросы оценки эффективности средств защиты и генерации сценариев нагрузочного тестирования.

### МЕТОДИКА ОЦЕНКИ ЭФФЕКТИВНОСТИ СРЕДСТВ ЗАЩИТЫ ОТ DDoS-АТАК

Для получения интегральной оценки эффективности средств защиты от DDoS-атак предлагается следующая математическая модель (рис. 3).

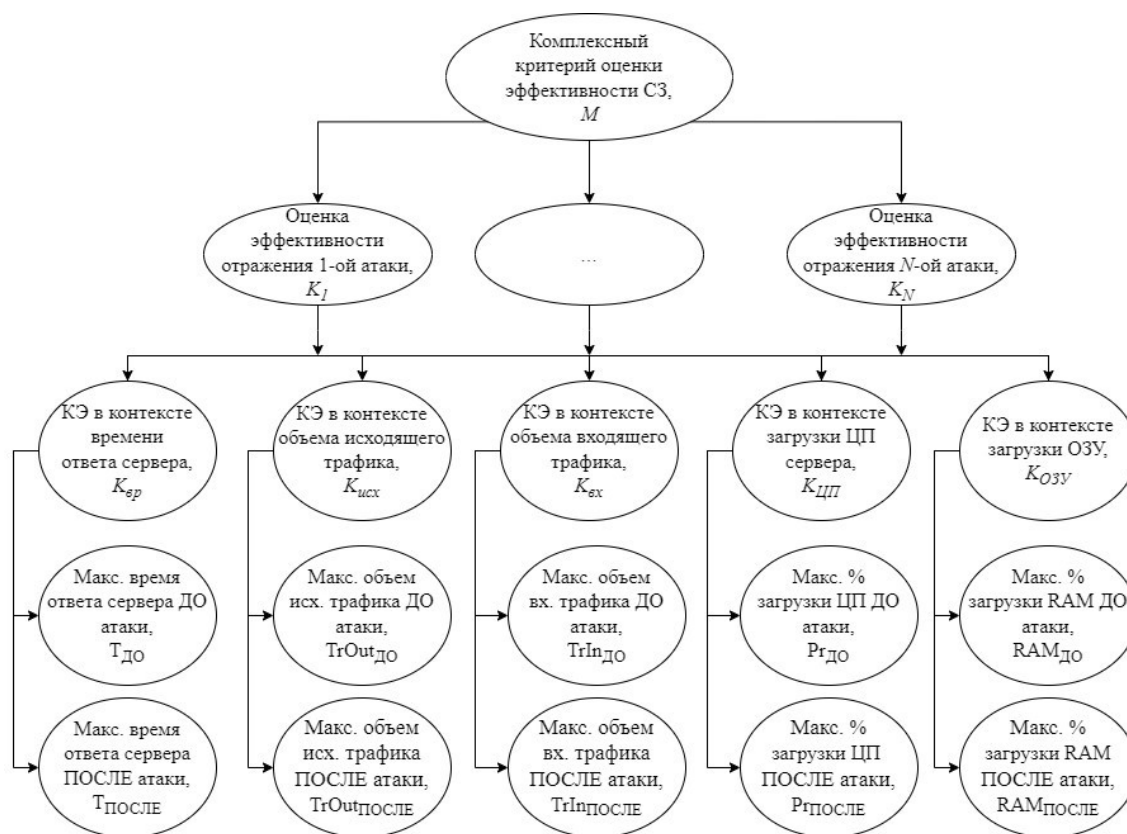


Рисунок 3 – Математическая модель оценки эффективности средств защиты от DDoS-атак

Комплексный критерий оценки эффективности средств защиты от DDoS-атак  $M$  определяется как минимальная оценка, выставленная за какую-то из тестовых атак (1):

$$M = \min(K_i), \quad (1)$$

где  $K_i$  – оценка, выставленная СЗ от DDoS-атак по результатам  $i$ -й тестовой атаки;

$i$  – номер тестовой атаки от 0 до  $N$ ;

$N$  – количество атак.

Каждая  $i$ -я атака проводится на ресурс при прочих равных условиях и отличие состоит только в сценарии атаки. До начала атаки фиксируются максимальные значения показателей состояния ресурса: время ответа сервера, объем входящего и исходящего трафика, загруженность процессора и оперативной памяти. Далее проводится тестовая атака и фиксируются аналогичные показатели. По результатам замеров средству защиты от DDoS-атак по формуле 2 выставляется оценка  $K$ :

$$K = S \left( 1 - \frac{k_{вр} + k_{вх} + k_{исх} + k_{ЦП} + k_{ОЗУ}}{5} \right), \quad (2)$$

где  $S \in \{0; 1\}$  (0 – средство защиты не смогло отразить текущую атаку; 1 – средство защиты отразило текущую атаку);

$k_{вр}$  – коэффициент эффективности в контексте времени ответа сервера;

$k_{вх}$  – коэффициент эффективности в контексте объема входящего трафика;

$k_{исх}$  – коэффициент эффективности в контексте объема исходящего трафика;

$k_{ЦП}$  – коэффициент эффективности в контексте загрузки процессора;

$k_{ОЗУ}$  – коэффициент эффективности в контексте занятой оперативной памяти.

Каждый из вышеперечисленных коэффициентов в контексте времени ответа сервера, объема входящего трафика, объема исходящего трафика, загрузки процессора и оперативной памяти определяется по формуле, зависящей от соответствующих показателей до и после атаки. Так, например, коэффициент эффективности в контексте времени ответа сервера  $k_{вр}$  определяется по формуле 3:

$$k_{вр} = \frac{T_{после} - T_{до}}{T_{после}}, \quad (3)$$

где  $T_{после}$  – максимальное время ответа сервера после начала атаки;

$T_{до}$  – максимальное время ответа сервера до начала атаки.

Предложенная математическая модель может стать ориентиром не только в оценке средств защиты от DDoS-атак, но и в оценке тестовых сценариев атак. Если было замечено, что каким-либо тестовым сценариям средству защиты от DDoS-атак противостоять сложнее, то имеет смысл в большей степени проверять именно на тестах такого типа с целью выявления недостатков средства защиты и его дальнейшего совершенствования.

Рассмотрим модель системы, при которой оценка средства защиты после проведения той или иной атаки влияет на последующие тесты.

#### СТЕНД ДЛЯ ТЕСТИРОВАНИЯ СРЕДСТВ ЗАЩИТ ОТ DDoS-АТАК

Для тестирования средств защиты от DDoS-атак был разработан стенд, схема которого представлена на рисунке 6.

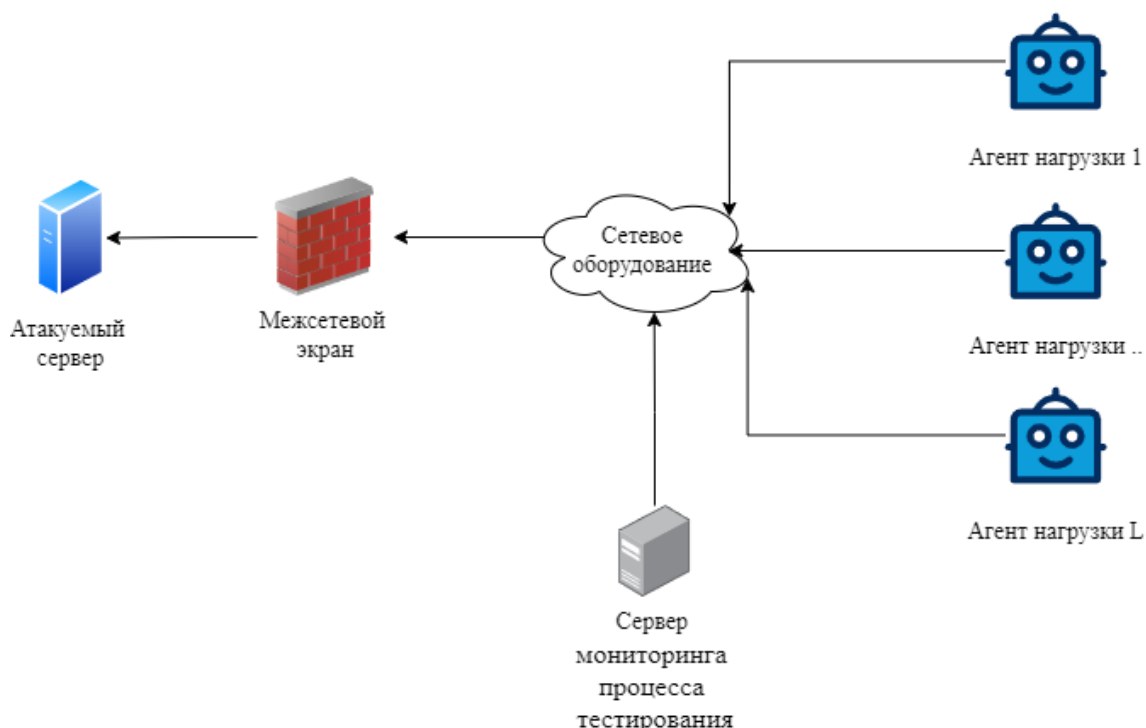


Рисунок 4 – Схема стенда для тестирования средств защиты от DDoS-атак

Стенд состоит из следующих основных частей:

- 1) атакуемый сервер – сервер, с которого будут считываться показатели работы, учитываемые в математической модели (рис. 4);
- 2) межсетевой экран – тестируемое средство защиты от DDoS-атак;
- 3) сервер мониторинга процесса тестирования – сервер, предназначенный для фиксации показателей работоспособности атакуемого сервера;
- 4) сетевое оборудование – комплекс коммутаторов и/или маршрутизаторов, предназначенных для передачи данных и маршрутизации трафика.
- 5) агенты нагрузки (1...L) – комплекс из реальных и/или виртуальных машин, чье назначение состоит в генерации трафика.

На базе стенда, собранного по представленной на рисунке 6 схеме, можно проводить тестирование различных средств защиты от DDoS-атак, которые могут установлены в качестве межсетевого экрана. На агенты нагрузки при этом могут подаваться разнообразные, заранее подготовленные сценарии атак.

#### Расчетный пример

Была проведена серия экспериментов по оценке эффективности некоего межсетевого экрана на использовании предложенного стенда. Были разработаны 5 сценариев атак:

- ICMP flood;
- TCP SYN flood;
- UDP flood;
- HTTP flood (PPS);
- HTTP flood.

В таблицу сведены результаты по расчету всех коэффициентов эффективности по каждому из реализованных сценариев атак.

Таблица – Результаты тестирования

	Время ответа сервера, мс			Объем входящего трафика, Мб/с			Объем исходящего трафика, Мб/с			Загруженность ЦП, %			Загруженность ОЗУ, %			К	М
	ДО атаки	ПОСЛЕ атаки	Квр	ДО атаки	ПОСЛЕ атаки	Квх	ДО атаки	ПОСЛЕ атаки	Кисх	ДО атаки	ПОСЛЕ атаки	Кцп	ДО атаки	ПОСЛЕ атаки	Козу		
ICMP flood		1,89	0,41799		8,87	0,76325		8,38	0,98926			2	0,5		12	0,16667	0,43257
TCP SYN flood		2,27	0,51542		2,49	0,15663		0,6	0,85			1	0		12	0,16667	0,66226
UDP flood	1,1	1,3	0,15385	2,1	117	0,98205	0,09	0,6	0,85	1		10	0,9	10	12,3	0,18699	0,38542
HTTP flood (PPS)		1,56	0,29487		127	0,98346		0,3	0,7			5	0,8		20	0,5	0,34433
HTTP flood		10	0,89		30	0,93		0,1	0,1			10	0,9		12,3	0,18699	0,3986

На основе данных, представленных в таблице, был рассчитан комплексный критерий оценки тестируемого средства защиты  $M = 0,66$ . Это показывает, что средство защиты справилось с серией тестовых атак.

Кроме того, анализ полученных экспериментальных данных показывает, что наибольшие сложности для тестируемого средства защиты создали сценарии ICMP-flood и TCP SYN-flood. Таким образом, для дальнейшего изучения и совершенствования средства защиты целесообразно использовать в ходе тестирования именно данный вектор атак.

#### Схема генерации тестовой нагрузки с обратной связью

Традиционно перед тестированием средств защиты от DDoS-атак разрабатывается ряд тестовых сценариев атак, которым последовательно подвергается защищаемый тестовый ресурс (рис. 5).

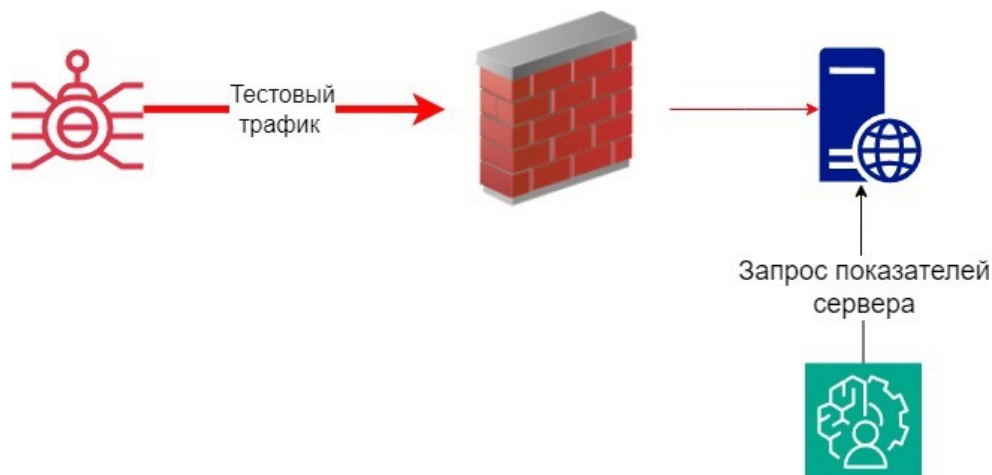


Рисунок 5 – Традиционный подход к тестированию средств защиты от DDoS-атак

Более перспективным представляется подход, учитывающий динамики изменения показателей защищаемого ресурса с целью воздействия на процесс генерации тестового трафика (рис. 6).

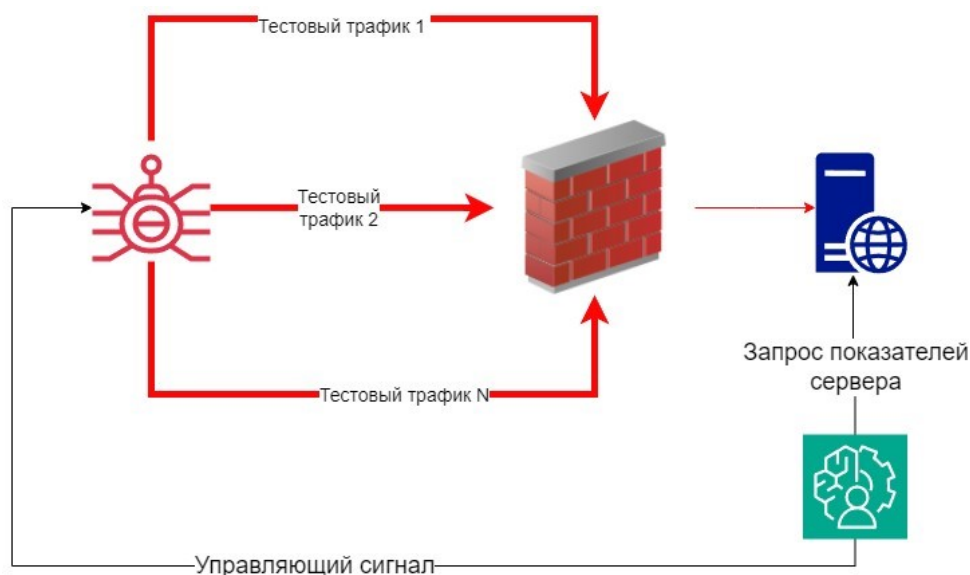


Рисунок 6 – Модифицированный подход к тестированию средств защиты от DDoS-атак

В рамках данной схемы встает вопрос о способе генерации тестового трафика. Если наблюдается ситуация, при которой в ходе тестовой атаки существенно ухудшаются показатели ресурса, то имеет смысл расширить пространство тестов за счет похожих сценариев или за счет сценариев, которые объединяют в себе свойства тех сценариев атак, с которым средство защиты справлялось хуже. Ручное переписывание сценариев атак может занять существенное время для специалиста по тестированию. Данную задачу целесообразно было бы автоматизировать, но полный перебор всех возможных модифицированных усовершенствованных сценариев атак может занять очень много времени и среди них, с большой долей вероятности, будут те сценарии, которые не являются принципиально лучшими по сравнению с изначальными. Для генерации тестового трафика, на котором более целесообразно проверять средства защиты от DDoS-атак, в [8] предлагалось использование генетических алгоритмов. Для выбора наиболее успешных атак, на основе которых будут генерироваться новые атаки, будет использоваться изображенная на рисунке 3 математическая модель.

### ЗАКЛЮЧЕНИЕ

В рамках статьи была предложена модель оценки эффективности средств защиты от DDoS-атак, которая опирается на оценки средства защиты после каждой из заранее подготовленных по определенному сценарию атак. Представленная математическая модель может быть использована как для оценки эффективности средства защиты, так и для оценки успешности атаки. Оценка успешности атаки может применяться для создания более эффективных сценариев атак, которые

в дальнейшем позволят выявить недостатки средства защиты от DDoS-атак. Для генерации сценариев предлагается использование генетических алгоритмов. Детальному рассмотрению вопроса использования генетических алгоритмов для генерации сценариев атак планируется посвятить внимание в дальнейших работах.

#### Список источников

1. Отчет об атаках на онлайн-ресурсы российских компаний за 2022 год // SOLAR. – URL: [https://rt-solar.ru/upload/iblock/34a/5w4h9o57axovdbv3ng7givr271ykir3/Ataki-na-onlayn\\_resursy-rossiyskikh-kompaniy-v-2022-godu.pdf](https://rt-solar.ru/upload/iblock/34a/5w4h9o57axovdbv3ng7givr271ykir3/Ataki-na-onlayn_resursy-rossiyskikh-kompaniy-v-2022-godu.pdf) (дата обращения: 12.05.2024).
2. Cyber attacks in the Israel-Hamas war // The Cloudflare Blog. – URL: <https://blog.cloudflare.com/cyber-attacks-in-the-israel-hamas-war/> (дата обращения: 12.05.2024).
3. Россия вошла в ТОП-10 стран по количеству DDoS-атак // Ведомости. – URL: <https://www.vedomosti.ru/technology/articles/2023/08/25/991814-rossiya-voshla-v-top-10-stran-po-kolichestvu-ddos-atak> (дата обращения: 12.05.2024).
4. Бобрышева, Г. В. Методы оценки информационной безопасности / Г. В. Бобрышева, А. О. Звозникова // Проблемы информатики в образовании, управлении, экономике и технике : сборник статей XVIII Международной научно-технической конференции, посвященной 75-летию Пензенского государственного университета., Пенза, 25–26 октября 2018 года. – Пенза : Автономная некоммерческая научно-образовательная организация «Приволжский Дом знаний», 2018. – С. 123–127. – EDN MGETBR
5. Алькаев, В. А. Средства анализа защищенности, применяемые для оценки эффективности функционирования средств защиты информации / В. А. Алькаев, А. Г. Фатеев // Инжиниринг и технологии. – 2018. – Т. 3, № 2. – С. 25–28.
6. Яковенко, Я. И. Особенности прогнозирования и оценки эффективности средств антивирусной защиты и межсетевое экранирование в АСУ ТП на предприятии нефтегазовой отрасли / Я. И. Яковенко, Е. В. Каменева // Инновационные научные исследования: теория, методология, практика : сборник статей XV Международной научно-практической конференции. – Пенза : Наука и Просвещение, 2018. – С. 73–82.
7. Синадский, Н. И. Методология синтеза интерактивной сетевой среды для компьютерных полигонов в сфере информационной безопасности : дис. ... д-ра техн. наук: 2.3.6. / Н. И. Синадский. – Екатеринбург, 2022. – 304 с.
8. Синадский, Н. И. Автоматизация тестирования сетевых средств защиты информации на основе применения эволюционно-генетического подхода / Н. И. Синадский, А. В. Агафонов // Математические структуры и моделирование. – 2018. – № 2 (46). – С. 125–134.

#### References

1. Report on attacks on online resources of Russian companies for 2022. URL: [https://rt-solar.ru/upload/iblock/34a/5w4h9o57axovdbv3ng7givr271ykir3/Ataki-na-onlayn\\_resursy-rossiyskikh-kompaniy-v-2022-godu.pdf](https://rt-solar.ru/upload/iblock/34a/5w4h9o57axovdbv3ng7givr271ykir3/Ataki-na-onlayn_resursy-rossiyskikh-kompaniy-v-2022-godu.pdf) (In Russ.).
2. Cyber attacks in the Israel-Hamas war. *The Cloudflare Blog*. URL: <https://blog.cloudflare.com/cyber-attacks-in-the-israel-hamas-war/> (In Russ.).
3. Russia has entered the TOP 10 countries in terms of the number of DDoS attacks. *Vedomosti*. URL: <https://www.vedomosti.ru/technology/articles/2023/08/25/991814-rossiya-voshla-v-top-10-stran-po-kolichestvu-ddos-atak> (In Russ.).
4. Bobrysheva, G. V., Zvoznikova, A. O. Methods for assessing information security. *Problems of computer science in education, management, economics and technology : collection of articles of the XVIII International Scientific and Technical Conference dedicated to the 75th anniversary Penza State University, Penza, October 25–26, 2018*. Penza, Autonomous non-profit scientific and educational organization “Privolzhsky House of Knowledge”, 2018, pp. 123–127. EDN MGETBR (In Russ.).
5. Alkaev, V. A., Fateev, A. G. Security analysis tools used to evaluate the effectiveness of information security tools. *Engineering and Technology*, 2018, vol. 3, no. 2, pp. 25–28 (In Russ.).
6. Yakovenko, Ya. I., Kameneva, E. V. Features of forecasting and assessing the effectiveness of anti-virus protection and firewalling in automated process control systems at an oil and gas industry enterprise. *Innovative scientific research: theory, methodology, practice : collection of articles of the XV International Scientific and Practical Conference*. Penza: Nauka i Prsveshchenie Publ., 2018, pp. 73–82 (In Russ.).
7. Sinadsky, N. I. *Methodology for the synthesis of an interactive network environment for computer testing grounds in the field of information security : dis. ... Dr. Tech. Sciences: 2.3.6*. Ekaterinburg, 2022. 304 p. (In Russ.).
8. Sinadsky, N. I., Agafonov, A. V. Automation of testing of network information security tools based on the application of an evolutionary genetic approach. *Mathematical Structures and Modeling*, 2018, no. 2 (46), pp. 125–134 (In Russ.).

Статья поступила в редакцию 06.06.2024; одобрена после рецензирования 24.06.2024; принята к публикации 04.07.2024.

The article was submitted 06.06.2024; approved after reviewing 24.06.2024; accepted for publication 04.07.2024.

УДК 004.032.26

## РАЗРАБОТКА МОДЕЛИ ПОВЕДЕНЧЕСКОЙ АНАЛИТИКИ ПОЛЬЗОВАТЕЛЕЙ И СУЩНОСТЕЙ НА ОСНОВЕ ЧАСТНОГО ПРОФИЛИРОВАНИЯ

**Частикова Вера Аркадьевна**, Кубанский государственный технологический университет, 350072, Российская Федерация, г. Краснодар, ул. Московская, 2, кандидат технических наук, доцент, ORCID: 0000-0003-2372-8275, e-mail: chastikova\_va@mail.ru  
**Шоничев Егор Владимирович**, Кубанский государственный технологический университет, 350072, Российская Федерация, г. Краснодар, ул. Московская, 2, студент, ORCID: 0009-0000-6940-8096, e-mail: shonichev-egor@yandex.ru

Проблема инсайдерских угроз всегда являлась актуальной в рамках процессов информационной безопасности. Несовершенство методов идентификации злонамеренных действий во внутренней сети организации ставило под угрозу защищенность информационных активов. В рамках данной статьи приведено описание наиболее перспективного направления развития систем поведенческой аналитики – нейросетевого анализа, представлена методика выявления девиантного поведения на основе частного профилирования субъектов интранета, и разработана, проанализирована модель нейронной сети на конкретном наборе данных.

**Ключевые слова:** поведенческая аналитика, инсайдерская угроза, нейронная сеть, архитектура LSTM, профилирование, информационная безопасность, UEBA

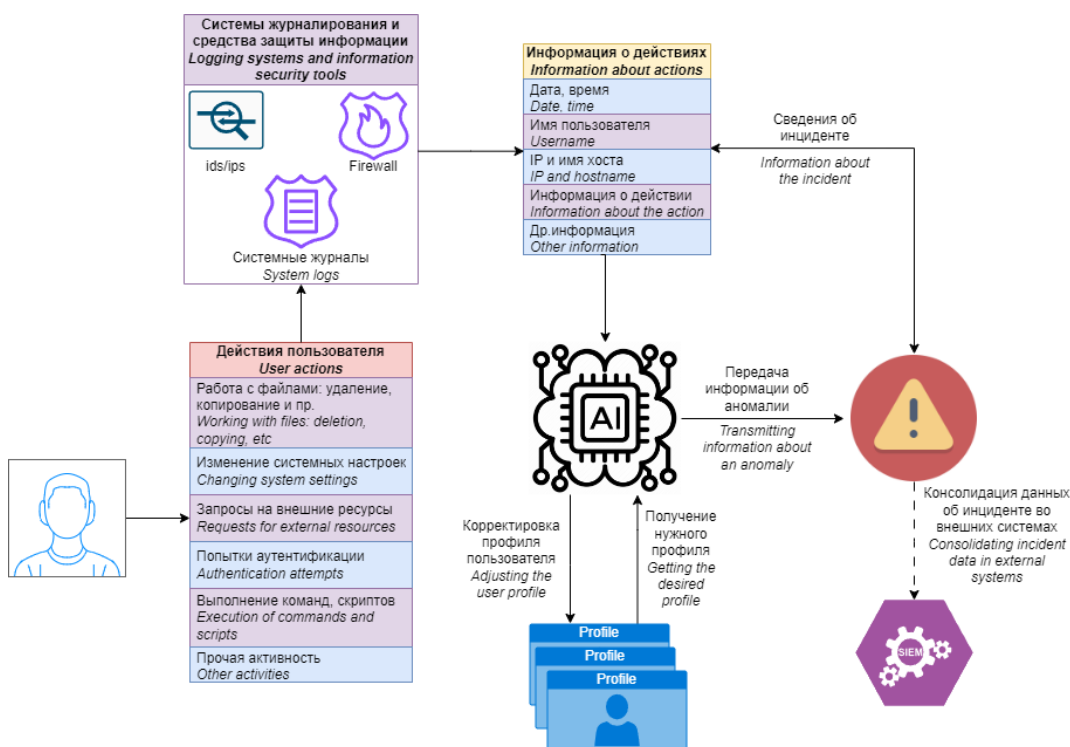
## DEVELOPMENT OF A BEHAVIORAL ANALYTICS MODEL FOR USERS AND ENTITIES BASED ON PERSONAL PROFILING

**Chastikova Vera A.**, Kuban State Technological University, 2 Moskovskaya St., Krasnodar, 350072, Russian Federation, Cand. Sci. (Engineering), Associate Professor, ORCID: 0000-0003-2372-8275, e-mail: chastikova\_va@mail.ru  
**Shonichev Egor V.**, Kuban State Technological University, 2 Moskovskaya St., Krasnodar, 350072, Russian Federation, student, ORCID: 0009-0000-6940-8096, e-mail: shonichev-egor@yandex.ru

The problem of insider threats has always been relevant within the framework of information security processes. The imperfection of methods for identifying malicious actions in the internal network of the organization jeopardized the security of information assets. Within the framework of this article, a description of the most promising direction for the development of behavioral analytics systems – neural network analysis will be provided, its own methodology for detecting deviant behavior based on personal profiling of intranet subjects will be presented, and a neural network model on a specific dataset will be developed and analyzed.

**Keywords:** behavioral analytics, insider threat, neural network, LSTM architecture, profiling, information security, UEBA

### Graphic annotation (Графическая аннотация)



## ВВЕДЕНИЕ

В рамках процессов обеспечения информационной безопасности часто не уделяется должное внимание проблеме инсайдерских атак и утечек данных. Большинство средств защиты информации (СЗИ), сфокусированные на выявлении злонамеренного воздействия на инфраструктуру организаций извне, не учитывают существование довольно неочевидной угрозы, исходящей от авторизованных субъектов (сотрудников, сетевого оборудования и пр.).

UEBA-система (англ. User and Entity Behavior Analytics – аналитика поведения пользователей и сущностей) – класс решений, ориентированный на выявление внутренних угроз безопасности, опирающийся на детектирование несвойственного поведения пользователей и сущностей.

Стоит уточнить, что под сущностями понимается любой субъект локальной сети предприятия, не являющийся человеком: это может быть сетевое устройство, такое, как маршрутизатор или управляемый коммутатор, IoT (IP-камера, датчики и др.), хосты (рабочая станция пользователя, сервер, виртуальная машина), а также программное обеспечение.

Почти три четверти (74 %) всех нарушений связаны с человеческим фактором. Люди злоупотребляют привилегиями, используют украденные учетные данные, допускают ошибки и занимаются мошенничеством с помощью социальной инженерии [1].

Злоумышленнику не требуется иметь специальные навыки или особое оборудование для социальной инженерии, а соответственно, такой способ даже если не принесет выгоды, то и серьезных расходов не потребует. Благодаря этому мошенники для реализации своих целей все чаще обращаются именно к такому способу атаки [2].

Количество утечек в 2024 г. все также остается на достаточно высоком уровне, несмотря на видимое уменьшение активности некоторых группировок, преимущественно публикующих материалы в Telegram-каналах. В Q1'24 61 % утечек пользовательских данных публикуется в свободном доступе в течение месяца после выгрузки из базы данных атакованной компании, в Q1'23 данное значение было 49 %. Подобное обстоятельство связано с увеличением количества публикуемых баз данных, относящихся к предприятиям малого и среднего бизнеса, которые, в свою очередь, представляют меньший интерес, чем базы крупных компаний, и используются преимущественно для поднятия репутации злоумышленника или проекта [3].

На рисунке 1 представлен график публикации утечек данных согласно информации аналитического отчета от «Лаборатории Касперского».

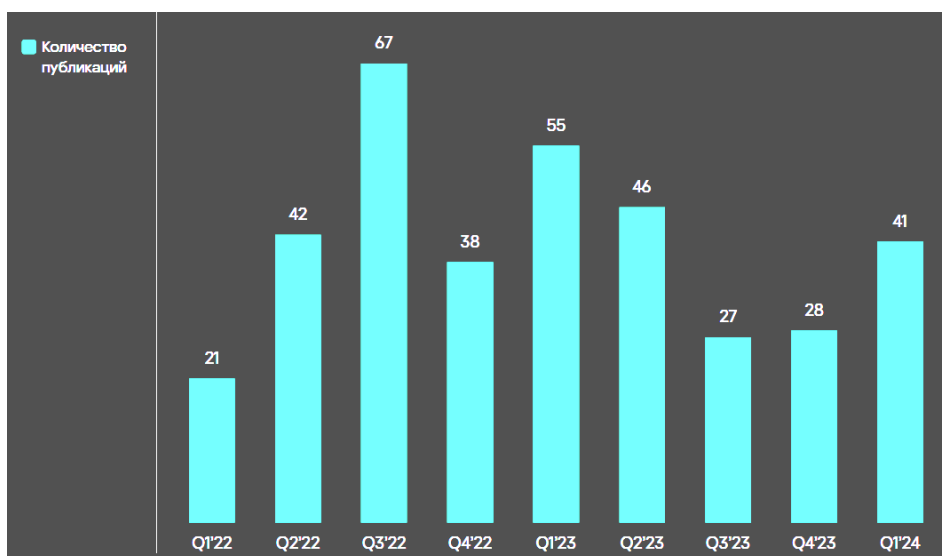


Рисунок 1 – Распределение публикаций утечек данных [3]

## НЕЙРОСЕТЕВОЙ ПОДХОД В ПОВЕДЕНЧЕСКОМ АНАЛИЗЕ

Принцип работы UEBA на самом деле очень прост. Имя пользователя и пароль сотрудника можно легко украсть, но значительно сложнее имитировать типичное поведение сотрудника, когда он находится в сети [4].

Решения поведенческой аналитики с самого своего «зарождения» используют следующие методы выявления неестественного поведения:

- правила: заранее прописанные паттерны поведения, которые имеют четко определенные маркеры, однако такой подход требует уже ранее выявленные инциденты, проанализированные экспертами;
- статистика: статистическое моделирование предлагает менее точное прогнозирование в отличие от правил, однако оно позволяет выявлять ранее неизвестные инциденты и значительные масштабные отклонения от нормы. Однако в виду обобщенности данных очень часто возникают

ложные срабатывания, в особенности это происходит в пики отчетных периодов, когда вырастает объем трафика от пользователей и, как следствие, число событий.

Решения, основанные на поведенческой аналитике, ведут базу данных объектов системы, а также сохраняют данные о нормальном и ненормальном поведении системы, что является обнаружением аномалий. Обнаружение аномалий позволяет определить, не соответствует ли шаблон данных ожидаемому нормальному поведению [5].

На сегодняшний день классические методы в определении аномального поведения начинают значительно уступать нейросетевому анализу ввиду его гибкости и приспособляемости. Бурный рост количества разнообразных архитектур нейронных сетей также способствуют развитию указанной технологии. Однако стоит заметить, что не каждая архитектура является эффективной для конкретной задачи из-за особенности их работы с входными данными и форматом возвращаемых данных [6]. Поэтому при разработке подобных моделей стоит учитывать контекст: тип данных, желаемые результаты, особые требования к обработке и пр.

Стоит заметить, что применение методов глубокого обучения значительно повышает универсальность аналитического модуля, поскольку выделение необходимых признаков и всего набора данных происходит автоматически без необходимости вмешательства со стороны человека. В итоге работа с нейронной сетью сводится к модели черного ящика: человеку не важно, как работает система, ему важен результат. В итоге он способен влиять на последующие результаты благодаря методу обратного распространения ошибки.

В рамках данного исследования для решения задачи выявления девиантного поведения на основе частного профилирования была разработана модель нейронной сети, приведённая на рисунке 2. Она состоит из двух последовательно соединенных LSTM-слоев (англ. Long Short-Term Memory – долгая краткосрочная память) [7–8] и Dense-слоя (полносвязный). Ввиду специфики работы LSTM размерность выходных данных, полученных с этого слоя, соответствует входным, при этом задача классификации заключается в сведении всего числа этих признаков к одному емкому ответу (классу): 0, если указанный набор действий является нормой, и 1 – несвойственное поведение. Для задачи изменения числа выходных признаков обычно применяют полносвязные слои.



Рисунок 2 – Модель нейронной сети

LSTM-архитектура была выбрана из-за ее способности обрабатывать обширные объемы контекстуальных данных. Рекуррентные нейронные сети, такие как RNN и LSTM, могут учитывать последовательности событий или временные ряды, что делает их чрезвычайно полезными в поведенческой аналитике, где целостная картина формируется последовательностью действий некоторого субъекта [9]. На рисунке 3 приведена схема работы LSTM-слоя.

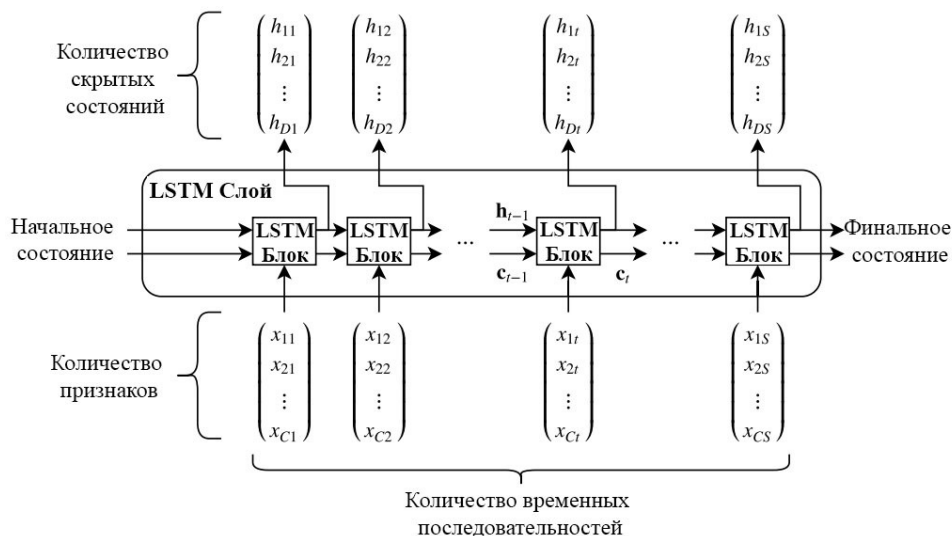


Рисунок 3 – Схема работы LSTM-слоя нейронной сети

### ОПИСАНИЕ МЕТОДИКИ ЧАСТНОГО ПРОФИЛИРОВАНИЯ

Несмотря на значительные преимущества нейросетевых подходов в аналитике поведения, не исключена агрегация данных, как и в случае со статистическим анализом, что влечет за собой обобщение поведенческого профиля [10]. Иными словами, нейронная сеть в процессе своего обучения может либо переучиться, что повлечет за собой неспособность к выявлению новых аномалий, отличных от тех, которые были в исходном наборе данных, либо станет неспособной выявлять частные, несвойственные конкретному субъекту действия (например, приняв за норму действия администраторов при настройке локальной сети, такое поведение у обычных пользователей не будет учитываться). Риск переобучения нивелируется тщательно подобранным датасетом и корректировкой гиперпараметров модели, таких как:

- скорость обучения;
- количество эпох;
- количество скрытых состояний LSTM-слоев;
- количество временных последовательностей;
- размер батча (пакета) и пр.

Тем не менее обобщение модели представляет собой наиболее актуальную проблему, поскольку оно требует контроля за поступающими данными, их нормализации и грамотной корректировки весов модели. Предлагаемая методика должна решить подобные проблемы в анализе больших данных при работе с разнообразными пользователями и сущностями.

В основах методики частного профилирования лежит агрегация большого числа данных с разнообразных систем-сборщиков (систем журналирования, анализа трафика, межсетевых экранов и пр.), на их основе генерируется профиль под каждого конкретного субъекта сети (пользователь или сущность). Профили будут храниться в упорядоченном виде и при необходимости подгружаться для прогнозирования действий каждого пользователя. Затем данные будут сопоставляться и в случае выявления аномалий передаваться со всем контекстом (возможно, во внешние системы обработки и консолидации сведений об инцидентах) оператору системы для проведения расследования. Структурная схема разрабатываемого модуля представлена на рисунке 4.

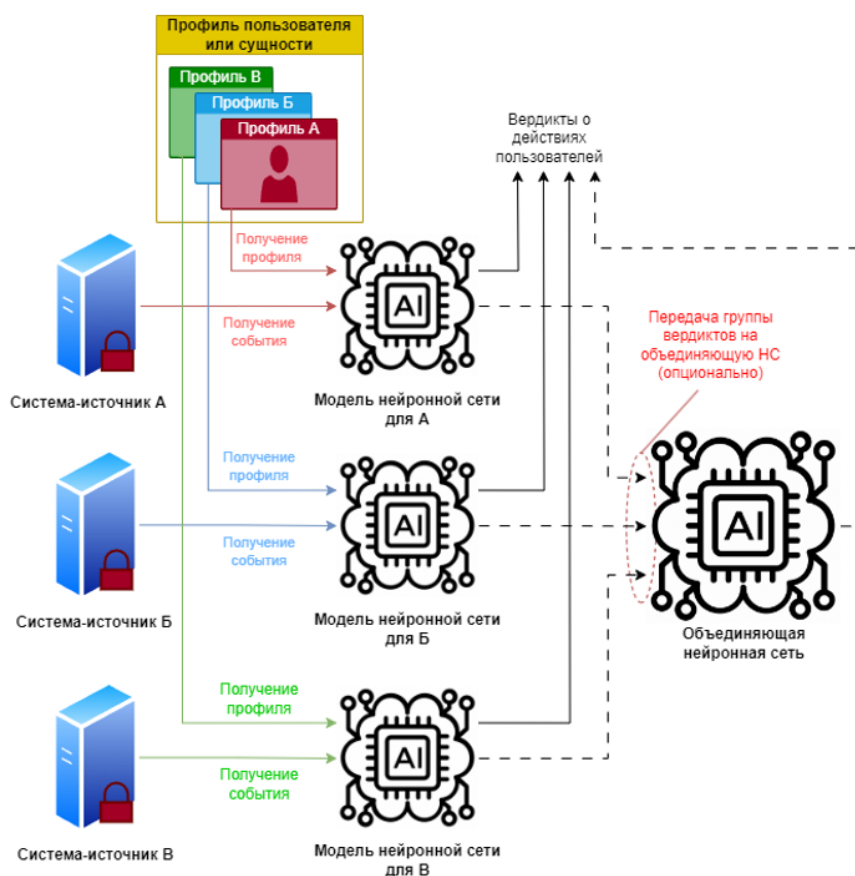


Рисунок 4 – Структурная схема работы UEBA-модуля

### ОЦЕНКА ЭФФЕКТИВНОСТИ ИДЕНТИФИКАЦИИ ИНСАЙДЕРСКИХ АТАК ПО ПРЕДЛОЖЕННОЙ МЕТОДИКЕ

В рамках данного исследования был выполнен комплексный анализ эффективности предложенной методики частного моделирования совместно с описанной ранее нейросетевой архитектурой, включающей отбор и исследование набора данных. Для обучения модели был использован датасет «UNSW\_NB15» [10], включающий в себя более 2,5 миллионов событий, представленных в форме необработанного трафика. Среди всех событий встречается девять типов атак:

- 1) фаззинг;
- 2) разведка;
- 3) анализ инфраструктуры;
- 4) бекдоры;
- 5) отказ в обслуживании (Dos);
- 6) эксплойты;
- 7) черви;
- 8) код оболочки (shell code);
- 9) генерация трафика (generic).

Учитывая критерии релевантности, разнообразия и обширности данных, выбранный датасет позволяет достичь максимальной точности при оценке эффективности предлагаемой архитектуры нейронной сети. Таким образом, отбор тестового (оценочного) датасета является неотъемлемой частью методологии данного исследования и обеспечивает объективную и всестороннюю оценку применимости и результативности предложенной архитектуры нейронной сети в контексте поставленных задач исследования.

Обучение модели производилось со специально подобранными гиперпараметрами, отображенными в таблице.

Таблица 1 – Гиперпараметры модели

Наименование	Значение
Скорость обучения	0,01
Размер батча (пакета)	64
Количество эпох обучения	5
Число входных признаков	42
Количество скрытых состояний	78
Количество временных последовательностей	8

Оценка эффективности разработанной модели производилась по следующим метрикам [11]:

- Accuracy – доля правильных ответов алгоритма;
- Precision – точность (предсказуемость), определяющая долю релевантных классов, которые были извлечены;
- Recall – полнота (чувствительность), являющаяся долей объектов положительного класса из всех объектов этого класса, найденных алгоритмом;
- F1-score – среднее гармоническое для precision и recall.

Для оценки эффективности датасет был разбит на две выборки: тренировочную (training), составляющую 80 % от всех записей, и валидационную (validation) – 20 %. В итоге наличие в обоих выборках уникальных записей позволяет произвести адекватную оценку эффективности модели. На рисунках 5 и 6 представлены графики обучения и проверки (валидации) работы алгоритма по метрикам, приведенным выше. В качестве функции потерь при обучении была выбрана бинарная кросс-энтропия, позволяющая выявить незначительные отклонения от целевых результатов.

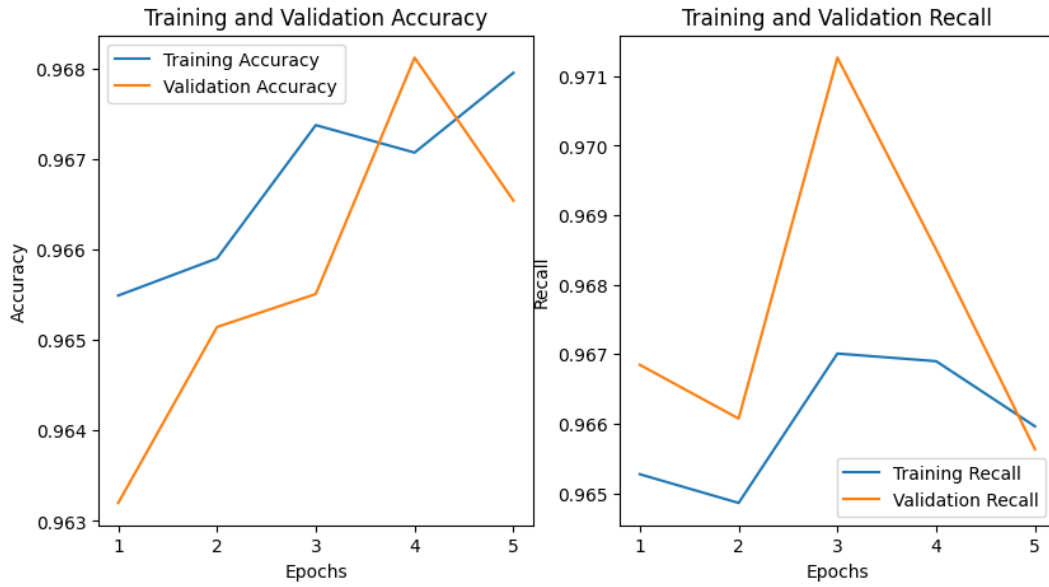


Рисунок 5 – Графики эффективности по Accuracy и Recall

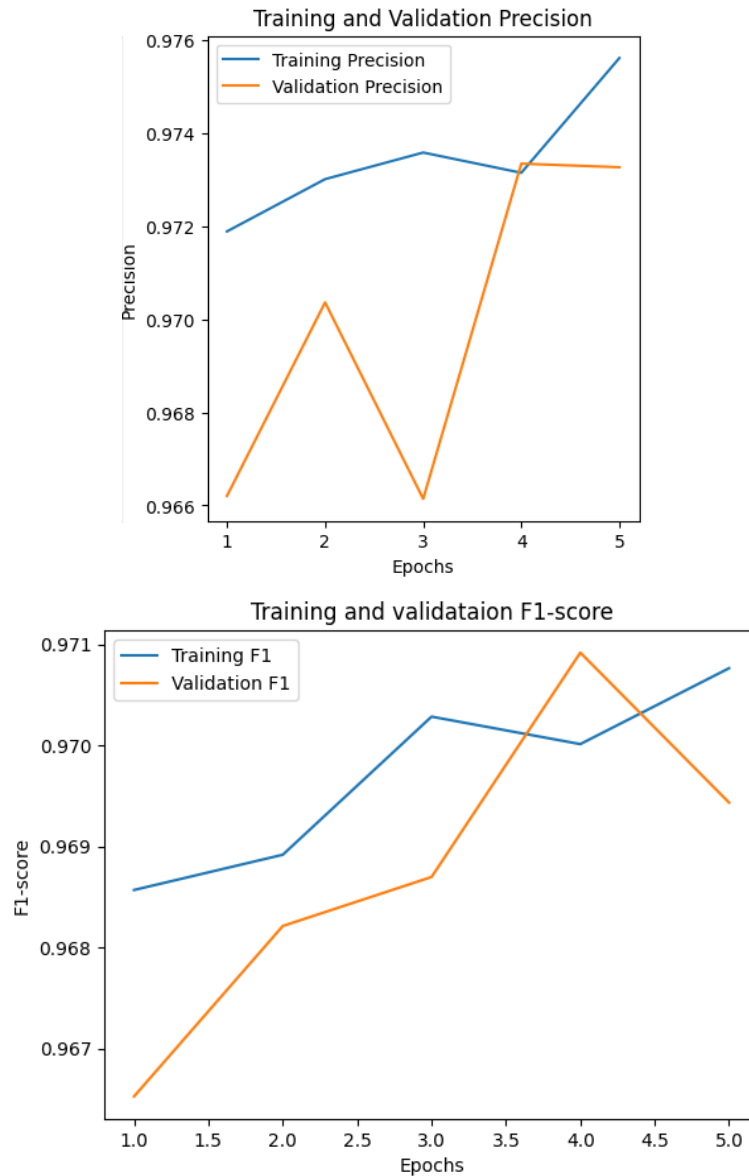


Рисунок 6 – Графики эффективности Precision и F1-score

## ЗАКЛЮЧЕНИЕ

Таким образом, опираясь на результаты исследования, приведенные ранее, можно сказать, что точность алгоритма имеет высокие результаты, что может способствовать снижению числа ложных срабатываний (ошибки первого рода), а также уменьшению ошибок второго рода. Также, исходя из проведенных испытаний, можно сделать вывод, что на 5 эпохе существенно возрастает риск переобучения, что видно по линии оранжевого цвета на графиках, резко идущей на снижение по всем метрикам.

Разработанная методика частного профилирования способна не просто снизить риски инсайдерских атак, эксплуатации уязвимостей нулевого дня (zero-day), утечки данных, но также обеспечить реактивное реагирование на зарождающийся инцидент ввиду быстрого анализа методами нейросетевого анализа. Данная методика допускает возможность использовать отдельно разработанные архитектуры нейронных сетей для каждой системы-источника, что может сказаться на повышении точности и снижении вычислительных затрат.

В итоге в рамках работы реализованного модуля будет динамически подгружаться профиль субъекта информационной инфраструктуры, представленный полноценным экземпляром обученной нейронной сети под конкретного пользователя или сущность, при этом этот экземпляр будет храниться в себе не просто набор скорректированных весов и отклонений, а полноценные слои нейронной сети с их связями между собой. При этом скорость загрузки модели может считаться практически мгновенной и зависит лишь от количества хранимых слоев (размера файла) и аппаратных характеристик накопителя и процессора (центрального или графического).

## Список источников

1. Verizon 2023 Data Breach Report: A Bulleted Summary. – URL: <https://rublon.com/blog/verizon-2023-data-breach-report-summary/#:~:text=Verizon's%202023%20Data%20Breach%20Investigation,6%20continents%20and%20%20industries> (дата обращения: 15.04.2024).
2. Частикова, В. А. Методика обнаружения атак социальной инженерии на основе алгоритмов анализа естественного языка / В. А. Частикова, В. Г. Гуляй // Прикаспийский журнал: управление и высокие технологии. – 2022. – № 3 (59). – С. 61–71.
3. Ландшафт угроз для России и СНГ 2024. Аналитический отчет от лаборатории Касперского. – URL: [https://go.kaspersky.com/rs/802-IJN-240/images/Report\\_Threat\\_Landscape\\_RU.pdf](https://go.kaspersky.com/rs/802-IJN-240/images/Report_Threat_Landscape_RU.pdf) (дата обращения: 16.04.2024).
4. Raguvir, S. Detecting Anomalies in Users – An UEBA Approach / S. Raguvir, Prof. Shekar Babu // IEOM Society International : Proceedings of the International Conference on Industrial Engineering and Operations Management, Dubai, UAE, March 10–12, 2020.
5. Manya, Ali Salitin. The role of User Entity Behavior Analytics to detect network attacks in real time / Manya Ali Salitin, Ali Hussein Zolait // 2018 International Conference on Innovation and Intelligence for Informatics, Computing, and Technologies (3ICT). – IEEE, 2018. – 978-1-5386-9207-3/18/\$31.00.
6. Pratik, Dhaygudea. A Review Paper on Different Deep Learning Methodologies for User and Entity Behavior Analytics (UEBA) / Pratik Dhaygudea, Nilesh Dhulshettea, Omkar Ganjalea, Umesh Sawanta, Prof. Pranali Navghare // International Journal of Research Publication and Reviews. – May 2023. – Vol. 4, № 5. – P. 3757–3762.
7. Zachary, C. Lipton. A Critical Review of Recurrent Neural Networks for Sequence Learning / Zachary C. Lipton, John Berkowitz, Charles Elkan // arXiv:1506.00019v4 [cs.LG]. – 17 Oct 2015.
8. Francesco, Nocera. A User Behavior Analytics (UBA)-based solution using LSTM Neural Network to mitigate DDoS Attack in Fog and Cloud Environment / Francesco Nocera, Simone Demilito, Piergiorgio Ladisa, Marina Mongiello, Awais Aziz Shah, Jawad Ahmad, Eugenio Di Sciascio // 2022 2nd International Conference of Smart Systems and Emerging Technologies (SMARTTECH). – 978-1-6654-0973-5/22/\$31.00. – IEEE, 2022. – DOI: 10.1109/SMARTTECH54121.2022.00029.
9. Zhihong, Tian. User and Entity Behavior Analysis under Urban Big Data / Zhihong Tian, Chaochao Luo, Hui Lu, Shen Su, Yanbin Sun, and Man Zhang // ACM/IMS Trans. Data Sci. – September 2020. – Vol. 1, № 3. – Article 16. – 19 p. – <https://doi.org/10.1145/3374749>.
10. Rabie, Osama. Security model for smart grid SCADA systems using stochastic neural network / Rabie Osama & Shitharth, & Alhazzawi, Daniyal & Kumar, Abhishek & Hamid, Syed & Zubair, Mohd // IET Generation, Transmission & Distribution. – 2023. – A 17. – n/a-n/a. 10.1049/gtd2.12943.
11. Santosh Nepal. User Behavior Analytics for Insider Threat Detection using Deep Learning / Santosh Nepal, Basanta Joshi // Proceedings of 10th IOE Graduate Conference. – October 2021. – Vol. 10. – P. 232–238. – ISSN 2350-8914 (Online), 2350-8906 (Print).

## References

1. Verizon 2023 Data Breach Report: A Bulleted Summary. URL: <https://rublon.com/blog/verizon-2023-data-breach-report-summary/#:~:text=Verizon's%202023%20Data%20Breach%20Investigation,6%20continents%20and%20%20industries> (accessed 15.04.2024).
2. Chastikova, V. A., Gulyai, V. G. Methods for detecting social engineering attacks based on natural language analysis algorithms. *Caspian Journal: Management and High Technologies*, 2022, no. 3 (59), pp. 61–71 (In Russ.).
3. *The threat landscape for Russia and the CIS 2024. Analytical report from Kaspersky Lab*. URL: [https://go.kaspersky.com/rs/802-IJN-240/images/Report\\_Threat\\_Landscape\\_RU.pdf](https://go.kaspersky.com/rs/802-IJN-240/images/Report_Threat_Landscape_RU.pdf) (In Russ.).
4. Raguvir, S, Prof. Shekar, Babu. Detecting Anomalies in Users – An UEBA Approach. IEOM Society International. *Proceedings of the International Conference on Industrial Engineering and Operations Management, Dubai, UAE, March 10–12, 2020*.

5. Many, Ali Salitin, Ali, Hussein Zolait. The role of User Entity Behavior Analytics to detect network attacks in real time. *2018 International Conference on Innovation and Intelligence for Informatics, Computing, and Technologies (3ICT)*. IEEE, 2018. 978-1-5386-9207-3/18/\$31.00.
6. Pratik, Dhaygudea, Nilesh, Dhulshettea, Omkar, Ganjalea, Umesh, Sawanta, Prof. Pranali, Navghare. A Review Paper on Different Deep Learning Methodologies for User and Entity Behavior Analytics (UEBA). *International Journal of Research Publication and Reviews*, May 2023, vol. 4, no. 5, pp. 3757–3762.
7. Zachary, C. Lipton, John, Berkowitz, Charles, Elkan. A Critical Review of Recurrent Neural Networks for Sequence Learning. *arXiv:1506.00019v4 [cs.LG]*. 17 Oct 2015.
8. Francesco, Nocera, Simone, Demilito, Piergiorgio, Ladisa, Marina, Mongiello, Awais, Aziz Shah, Jawad, Ahmad, Eugenio, Di Sciascio. A User Behavior Analytics (UBA)-based solution using LSTM Neural Network to mitigate DDoS Attack in Fog and Cloud Environment. *2022 2nd International Conference of Smart Systems and Emerging Technologies (SMARTTECH)*. IEEE, 2022. 978-1-6654-0973-5/22/\$31.00. DOI: 10.1109/SMARTTECH54121.2022.00029.
9. Zhihong, Tian, Chaochao, Luo, Hui, Lu, Shen, Su, Yanbin, Sun, and Man, Zhang. User and Entity Behavior Analysis under Urban Big Data. *ACM/IMS Trans. Data Sci.* September 2020, vol. 1, no. 3, article 16. 19 p. <https://doi.org/10.1145/3374749>.
10. Rabie, Osama & Shitharth, & Alghazzawi, Daniyal & Kumar, Abhishek & Hamid, Syed & Zubair, Mohd. A security model for smart grid SCADA systems using stochastic neural network. *IET Generation, Transmission & Distribution*, 2023, A 17. n/a-n/a. 10.1049/gtd2.12943.
11. Santosh, Nepal, Basanta, Joshi. User Behavior Analytics for Insider Threat Detection using Deep Learning. *Proceedings of 10th IOE Graduate Conference*, October 2021, vol. 10, pp. 232–238. ISSN 2350-8914 (Online), 2350-8906 (Print).

Статья поступила в редакцию 26.06.2024; одобрена после рецензирования 03.07.2024; принята к публикации 10.07.2024.

The article was submitted 26.06.2024; approved after reviewing 03.07.2024; accepted for publication 10.07.2024.

УДК 004.052.3

## РАЗРАБОТКА МАТЕМАТИЧЕСКОЙ МОДЕЛИ ОТКАЗОУСТОЙЧИВОГО SPN-ШИФРАТОРА AES

**Проворнов Игорь Александрович**, Институт цифрового развития ФГАОУ ВО «Северо-Кавказский федеральный университет, 355017, Российская Федерация, г. Ставрополь, ул. Пушкина, 1, соискатель, ORCID: 0000-0001-8902-3759, e-mail: igorprovornov@yandex.ru

**Калмыков Игорь Анатольевич**, Институт цифрового развития ФГАОУ ВО «Северо-Кавказский федеральный университет», 355017, Российская Федерация, г. Ставрополь, ул. Пушкина, 1, доктор технических наук, профессор, ORCID: 0000-0002-9854-5310, e-mail: kia762@yandex.ru

Статья посвящена вопросу повышения отказоустойчивости SPN-шифратора, реализованного в полиномиальной системе классов вычетов. Представлены результаты анализа существующих вариантов повышения отказоустойчивости устройств криптографической защиты информации, рассмотрены их недостатки, обоснована актуальность дальнейшего совершенствования SPN-шифраторов. Обоснован выбор оснований для используемой полиномиальной системы классов вычетов. Описана математическая модель отказоустойчивой реализации SPN-шифратора, включающая процедуры обнаружения и исправления ошибок после процедур SubBytes и MixColumns, с использованием модифицированного варианта корректирующего кода на основе вычисления синдромов ошибок. Целью исследования является повышение устойчивости к отказам SPN-шифратора, возникающим во время его работы, за счет использования разработанной математической модели SPN-шифратора, функционирующего в полиномиальной системе классов вычетов. Рассмотрены примеры применения предложенной математической модели, доказывающие её эффективность. Определены направления дальнейших исследований в области повышения отказоустойчивости устройств криптографической защиты информации.

**Ключевые слова:** криптоалгоритм AES, преобразование MixColumns, преобразование SubBytes, полиномиальная система классов вычетов, отказоустойчивость

## DEVELOPMENT OF A MATHEMATICAL MODEL OF A FAULT-TOLERANT IMPLEMENTATION OF THE SPN ENCODER

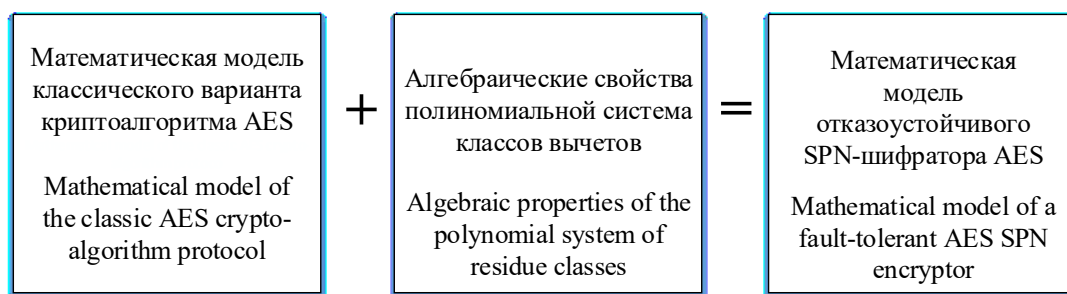
**Provornov Igor A.**, Institute for Digital Development of the North Caucasian Federal University, 1 Pushkin St., Stavropol, 355017, Russian Federation, applicant, ORCID: 0000-0001-8902-3759, e-mail: igorprovornov@yandex.ru

**Kalmykov Igor A.**, the Institute of Digital Development of the North Caucasus Federal University, 1 Pushkin St., Stavropol, 355017, Russian Federation, Doct. Sci. (Engineering), Professor, ORCID: 0000-0002-9854-5310, e-mail: kia762@yandex.ru

The article discusses the issue of enhancing the fault tolerance of an SPN (Substitution-Permutation Network) encoder implemented in a system of polynomial residue classes. The goal of the study is to develop a mathematical model for a fault-tolerant version of the SPN encoder that outperforms existing methods for ensuring the operational security of cryptographic systems. The article presents the results of an analysis of existing approaches for enhancing the fault tolerance of crypto devices, discusses their shortcomings, and justifies the need for further improvements to the SPN encoding technique. It also justifies the choice of base polynomials used in the polynomial residue class system. A mathematical model for implementing a fault-tolerant SPN encoder in a secure manner is described, incorporating procedures for detecting and correcting errors following the SubBytes and MixColumns steps, utilizing a modified version of the error correction code based on calculating error syndromes. Examples of the application of the proposed mathematical model are discussed, demonstrating its effectiveness. Directions for further research in the area of increasing the resilience of cryptographic information protection devices are identified.

**Keywords:** AES cryptographic algorithm, MixColumns transformation, SubBytes transformation, polynomial residue class system, fault tolerance, corrective coding

### Graphical annotation (Графическая аннотация)



## ВВЕДЕНИЕ

В настоящее время одним из ключевых факторов, поддерживающих стабильное развитие любой сферы жизни общества, является обеспечение защищенности информации и информационной среды. Особенно сохранность обрабатываемой информации важна для набирающей популярность технологии «Интернет вещей» (IoT), представляющей собой сеть технологических устройств, способных взаимодействовать между собой [1]. Наиболее чувствительными к утечке обрабатываемых данных являются системы «промышленного интернета вещей», обеспечивающие функционирование критически важных объектов инфраструктуры (атомные электростанции, промышленные производства и т. п.).

Для обеспечения защищенности конфиденциальной информации в таких системах широко применяются криптографические средства, которые могут быть реализованы программным, программно-аппаратным или аппаратным способом. AES является одним из наиболее распространенных алгоритмов шифрования за счет следующих преимуществ: высокая криптостойкость, высокая скорость шифрования, относительно низкая стоимость используемых вычислительных платформ.

Вместе с тем одновременно с развитием средств защиты информации идет параллельное развитие способов дешифрования, в том числе существует класс атак по побочным каналам, использующих уязвимости криптографических систем как физических объектов. В частности, известны атаки по ошибкам вычисления, в ходе которых шифратор подвергается воздействию различных внешних факторов (высокое напряжение, значительное изменение температуры, электромагнитное излучение и т. п.), в результате чего в его работе возникают сбои или отказы. Анализ возникших по этим причинам ошибок позволяет злоумышленнику определить фрагменты защищаемых данных или другие элементы используемого алгоритма, которые можно использовать для дальнейшего криптоанализа.

В связи с этим одним из принципов разработки криптографических систем является обеспечение их высокой отказоустойчивости, заключающейся в способности сохранять работоспособное состояние при возникновении различных сбоев.

Одним из способов повышения отказоустойчивости вычислительных систем, получившим наибольшее распространение, является их аппаратное резервирование по схеме «2 из 3», согласно которой одновременно работают 3 одинаковых устройства, выходные сигналы которых поступают на вход устройства мажоритарного выбора, осуществляющего сравнение сигналов и формирование выходного сигнала, идентичного как минимум двум входным сигналам [2].

Однако такой подход обладает следующими недостатками:

- значительная аппаратная избыточность, снижающая общую надежность устройства;
- невозможность исправить ошибки, возникшие одновременно в двух или трех устройствах в одинаковых разрядах обрабатываемого числа.

В связи с этим перспективным направлением исследований является разработка математических моделей отказоустойчивых SPN-систем на основе таких алгебраических систем, которые позволяют снизить аппаратную избыточность.

В качестве такой алгебраической системы можно использовать полиномиальную систему классов вычетов (ПСКВ), которая, согласно известным исследованиям, обладает всеми предпосылками для построения на её базе отказоустойчивых вычислительных систем [2]. Поэтому разработка математической модели отказоустойчивого SPN-шифратора, функционирующего в ПСКВ, является актуальной задачей.

## ПРИНЦИПЫ РЕАЛИЗАЦИИ КРИПТОАЛГОРИТМА AES

AES – итерационный блочный шифр, длина блоков которого согласно соответствующему стандарту составляет 128 бит. Начальное и промежуточные состояния обрабатываемых данных принято называть «состояниями» («State») и представлять в виде прямоугольного массива размером 4x4 байта (рис. 1) [3].

$S_{00}$	$S_{01}$	$S_{02}$	$S_{03}$
$S_{10}$	$S_{11}$	$S_{12}$	$S_{13}$
$S_{20}$	$S_{21}$	$S_{22}$	$S_{23}$
$S_{30}$	$S_{31}$	$S_{32}$	$S_{33}$

Рисунок 1 – Графическое представление обрабатываемого массива данных криптоалгоритма AES

В 10 раундах AES-шифрования выполняются преобразования SubBytes (подстановка по таблице замены), ShiftRows (побайтовый сдвиг строк), MixColumns (умножение столбцов на константы (не выполняется в последнем раунде), AddRoundKey (сложение с раундовым ключом).

Для реализации процедуры SubBytes используется таблица замены, представленная на рисунке 2 [4].

	0	1	2	3	4	5	6	7	8	9	a	b	c	d	e	f
0	63	7C	77	7B	F2	6B	6F	C5	30	01	67	2B	FE	D7	AB	76
1	CA	82	C9	7D	FA	59	47	F0	AD	D4	A2	AF	9C	A4	72	C0
2	B7	FD	93	26	36	3F	F7	CC	34	A5	E5	F1	71	D8	31	15
3	04	C7	23	C3	18	96	05	9A	07	12	80	E2	EB	27	B2	75
4	09	83	2C	1A	1B	6E	5A	A0	52	3B	D6	B3	29	E3	2F	84
5	53	D1	00	ED	20	FC	B1	5B	6A	CB	BE	39	4A	4C	58	CF
6	D0	EF	AA	FB	43	4D	33	85	45	F9	02	7F	50	3C	9F	A8
7	51	A3	40	8F	92	9D	38	F5	BC	B6	DA	21	10	FF	F3	D2
8	CD	0C	13	EC	5F	97	44	17	C4	A7	7E	3D	64	5D	19	73
9	60	81	4F	DC	22	2A	90	88	46	EE	B8	14	DE	5E	0B	DB
a	E0	32	3A	0A	49	06	24	5C	C2	D3	AC	62	91	95	E4	79
b	E7	C8	37	6D	8D	D5	4E	A9	6C	56	F4	EA	65	7A	AE	08
c	BA	78	25	2E	1C	A6	B4	C6	E8	DD	74	1F	4B	BD	8B	8A
d	70	3E	B5	66	48	03	F6	0E	61	35	57	B9	86	C1	1D	9E
e	E1	F8	98	11	69	D9	8E	94	9B	1E	87	E9	CE	55	28	DF
f	8C	A1	89	0D	BF	E6	42	68	41	99	2D	0F	B0	54	BB	16

Рисунок 2 – Классическая таблица SubBytes

Линейная операция ShiftRows заключается в сдвиге каждой строки обрабатываемого массива на разное количество позиций согласно выражению:

$$SR_{vw} = \begin{cases} S_{vw}, & v = 0, \\ S_{v(w+1) \bmod 4}, & v = 1, \\ S_{v(w+2) \bmod 4}, & v = 2, \\ S_{v(w+3) \bmod 4}, & v = 3, \end{cases} \quad (1)$$

где  $SR_{vw}$  – байт  $v$ -й строки  $w$ -го столбца массива после преобразования ShiftRows;

$S_{vw}$  – байт  $v$ -й  $w$ -го столбца массива до преобразования ShiftRows.

Преобразование MixColumns заключается в умножении столбцов входной последовательности по модулю  $x^4 \oplus 1$  на полином  $a(x) = 3x^3 \oplus x^2 \oplus x^2 \oplus 2$  [3], что эквивалентно матричному умножению:

$$\begin{bmatrix} M_{0w} \\ M_{1w} \\ M_{2w} \\ M_{3w} \end{bmatrix} = \begin{bmatrix} 02 & 03 & 01 & 01 \\ 01 & 02 & 03 & 01 \\ 01 & 01 & 02 & 03 \\ 03 & 01 & 01 & 02 \end{bmatrix} \cdot \begin{bmatrix} SR_{0w} \\ SR_{1w} \\ SR_{2w} \\ SR_{3w} \end{bmatrix}, \quad (2)$$

где  $M_{vw}$  – байт  $v$ -й строки  $w$ -го столбца после преобразования MixColumns.

Результатом такого умножения является столбец с номером  $w$ , значения байтов которого определяются выражениями:

$$M_{0w} = (2 \cdot SR_{0w}) \oplus (3 \cdot SR_{1w}) \oplus (1 \cdot SR_{2w}) \oplus (1 \cdot SR_{3w}), \quad (3)$$

$$M_{1w} = (1 \cdot SR_{0w}) \oplus (2 \cdot SR_{1w}) \oplus (3 \cdot SR_{2w}) \oplus (1 \cdot SR_{3w}), \quad (4)$$

$$M_{2w} = (1 \cdot SR_{0w}) \oplus (1 \cdot SR_{1w}) \oplus (2 \cdot SR_{2w}) \oplus (3 \cdot SR_{3w}), \quad (5)$$

$$M_{3w} = (3 \cdot SR_{0w}) \oplus (1 \cdot SR_{1w}) \oplus (1 \cdot SR_{2w}) \oplus (2 \cdot SR_{3w}). \quad (6)$$

Преобразование AddRoundKey заключается в сложении байтов обрабатываемого массива и байтов ключевой информации.

Рассмотрим пример выполнения одного раунда шифрования в соответствии с алгоритмом AES. Пусть на вход SPN-шифратора поступает массив данных  $A_{ex. ПСС}$ , элементами которого являются следующие байты:

$$A_{ex. ПСС} = \begin{bmatrix} (19_{16}) & (A0_{16}) & (9A_{16}) & (E9_{16}) \\ (3D_{16}) & (F4_{16}) & (C6_{16}) & (F8_{16}) \\ (E3_{16}) & (E2_{16}) & (8D_{16}) & (48_{16}) \\ (BE_{16}) & (2B_{16}) & (2A_{16}) & (08_{16}) \end{bmatrix}. \quad (7)$$

В результате замены каждого байта этого массива в соответствии с таблицей, представленной на рисунке, будет получен массив  $S$ :

$$S = \begin{bmatrix} D4_{16} & E0_{16} & B8_{16} & 1E_{16} \\ 27_{16} & BF_{26} & B4_{16} & 41_{16} \\ 11_{16} & 98_{16} & 5D_{16} & 52_{16} \\ AE_{16} & F1_{16} & E5_{16} & 30_{16} \end{bmatrix}.$$

Следующим преобразованием над массивом  $S$  будет побайтовый сдвиг строк. Таким образом, в ходе преобразования ShiftRows будет получен массив  $SR$ :

$$SR = \begin{bmatrix} D4_{16} & E0_{16} & B8_{16} & 1E_{16} \\ BF_{16} & B4_{16} & 41_{16} & 27_{16} \\ 5D_{16} & 52_{16} & 11_{16} & 98_{16} \\ 30_{16} & AE_{16} & F1_{16} & E5_{16} \end{bmatrix}.$$

Третьим этапом раунда является процедура MixColumns, в ходе которой будет осуществлено перемещение столбцов массива  $SR$ , в результате чего будет получен массив  $M$ :

$$M = \begin{bmatrix} 04_{16} & E0_{16} & 48_{16} & 28_{16} \\ 66_{16} & CB_{16} & F8_{16} & 06_{16} \\ 81_{16} & 19_{16} & D3_{16} & 26_{16} \\ E5_{16} & 9A_{16} & 7A_{16} & 4C_{16} \end{bmatrix}.$$

Для реализации последнего раундового преобразования AddRoundKey необходима ключевая информация, объем которой совпадает с объемом обрабатываемого массива.

Допустим, массив ключевой информации  $K$  имеет вид:

$$K = \begin{bmatrix} A0_{16} & 88_{16} & 23_{16} & 2A_{16} \\ FA_{16} & 54_{16} & A3_{16} & 6C_{16} \\ FE_{16} & 2C_{16} & 39_{16} & 76_{16} \\ 17_{16} & B1_{16} & 39_{16} & 05_{16} \end{bmatrix}. \quad (8)$$

Тогда преобразование AddRoundKey, в ходе которого формируется массив  $ARK$ , будет описываться выражением:

$$ARK = \begin{bmatrix} 04_{16} & E0_{16} & 48_{16} & 28_{16} \\ 66_{16} & CB_{16} & F8_{16} & 06_{16} \\ 81_{16} & 19_{16} & D3_{16} & 26_{16} \\ E5_{16} & 9A_{16} & 7A_{16} & 4C_{16} \end{bmatrix} \oplus \begin{bmatrix} A0_{16} & 88_{16} & 23_{16} & 2A_{16} \\ FA_{16} & 54_{16} & A3_{16} & 6C_{16} \\ FE_{16} & 2C_{16} & 39_{16} & 76_{16} \\ 17_{16} & B1_{16} & 39_{16} & 05_{16} \end{bmatrix} = \begin{bmatrix} A4_{16} & 68_{16} & 6B_{16} & 02_{16} \\ 9C_{16} & 9F_{16} & 5B_{16} & 6A_{16} \\ 7F_{16} & 35_{16} & EA_{16} & 50_{16} \\ F2_{16} & 2B_{16} & 43_{16} & 49_{16} \end{bmatrix}.$$

Проведенный анализ криптоалгоритма AES показал, что наиболее вероятны сбои в блоках выполнения процедур SubBytes и MixColumns, так как именно в этих блоках реализуется процедура умножения многочленов как элементов конечного поля  $GF(2^8)$ . Таким образом, можно сделать вывод, что в разрабатываемой модели достаточно обеспечить отказоустойчивую реализацию данных процедур.

### ПРИНЦИПЫ ПОСТРОЕНИЯ КОРРЕКТИРУЮЩИХ КОДОВ, РЕАЛИЗОВАННЫХ В ПОЛИНОМИАЛЬНОЙ СИСТЕМЕ КЛАССОВ ВЫЧЕТОВ

ПСКВ является такой непозиционной системой счисления, в которой в качестве оснований используются многочлены расширенного поля Галуа  $p_1(x), p_2(x), \dots, p_k(x)$ , а число  $A$  представляется в виде набора остатков от деления его полиномиального представления на эти многочлены [5]:

$$A = (a_1(x), a_2(x), \dots, a_i(x), \dots, a_k(x)), \quad (9)$$

где  $\alpha_i(x) = A \bmod p_i(x)$ ,  $i = 1, \dots, k$ .

При этом в качестве  $p_i(x)$  используются неприводимые полиномы, удовлетворяющие условию [6]:

$$\deg p_1(x) \leq \deg p_2(x) \leq \dots \leq \deg p_k(x). \quad (10)$$

Представление чисел в ПСКВ обладает следующими преимуществами: высокая скорость вычисления за счет сокращения разрядности обрабатываемых чисел и возможности параллелизации, отсутствие перехода ошибок между вычислительными трактами.

На данных алгебраических свойствах ПСКВ основаны корректирующие коды, которые предназначены для обнаружения и исправления ошибок в работе вычислительных устройств. В соответствии с теорией кодирования [7], для реализации таких возможностей используются избыточные данные, которые позволяют оценить степень искажения информации в процессе её обработки. Для внесения такой избыточности увеличивают число оснований ПСКВ с  $k$  до  $n$  и добавленные контрольные основания в количестве  $r = n - k$  используют для обнаружения и устранения ошибок. Такая ПСКВ описывается выражениями:

$$A = (\alpha_1(x), \alpha_2(x), \dots, \alpha_k(x), \alpha_{k+1}(x), \alpha_{k+2}(x), \dots, \alpha_n(x)), \quad (11)$$

$$\deg p_1 \leq \deg p_k \leq \deg p_{k+1} \leq \deg p_n. \quad (12)$$

В настоящее время известны следующие алгоритмы локализации и устранения ошибок, реализованные в ПСКВ [8]: нулевизация, вычисление интервального номера, алгоритм на основе коэффициентов обобщенной полиадической системы (ОПС), алгоритм на основе вычисления синдромов ошибок.

Сравнительный анализ классических реализаций этих методов показал необходимость использования двух контрольных оснований для коррекции однократных ошибок. Однако в связи с увеличением количества контрольных оснований увеличиваются и их степени, что, в свою очередь, приводит к повышению разрядности обрабатываемых остатков. Следствием этого является необходимость увеличения числа вычислительных блоков в процессоре, т. е. его аппаратного усложнения.

С этой точки зрения, перспективным методом является модификация алгоритма на основе определения синдромов ошибки [9], позволяющая вычислять два контрольных остатка с использованием одного контрольного полинома согласно:

$$\alpha_{k+1}(x) = \sum_{i=1}^k \alpha_i(x), \quad (13)$$

$$\alpha_{k+2}(x) = \sum_{i=1}^k (i(x)\alpha_i(x)) \bmod p_{k+1}(x),$$

где  $\Sigma$  – суммирование по модулю два;

$i(x)$  – полиномиальная форма  $i$ -го порядкового номера основания,  $i = 1, \dots, k$ .

Синдромы ошибки вычисляются в соответствии с выражением:

$$\delta_i(x) = (\alpha_{i+k}(x) \oplus \alpha_{i+k}^*(x)) \bmod p_{k+1}(x), \quad (14)$$

где  $\alpha_{i+k}^*(x)$  – проверочные контрольные остатки, вычисляемые на основе информационных остатков, полученных табличным способом.

Совокупность синдромов ошибок  $\delta_1(x)$  и  $\delta_2(x)$  позволяет сделать вывод о местоположении и глубине ошибки, что является достаточной информацией для определения вектора ошибок  $\Delta e(x)$ , используемого для коррекции обрабатываемого значения согласно:

$$\dot{A}(x) = \tilde{A}(x) \oplus \Delta e(x), \quad (15)$$

где  $\dot{A}(x)$  – исправленное значение обрабатываемого числа;

$\sim$  – идентификатор искаженных значений.

Дополнительные исследования [9] показали, что в случае возникновения ошибки в первом остатке  $\Delta\alpha_1(x)$ :

$$\Delta\alpha_1(x) = \delta_1(x) = \delta_2(x) \neq 0. \quad (16)$$

Это свойство позволяет осуществлять сравнение синдромов ошибки и вычислять вектор ошибки  $\Delta e(x)$ , который будет состоять из нулей за исключением ошибочного остатка.

Соответственно, такой подход к вычислению вектора ошибки позволяет уменьшить объем используемого блока памяти на  $2^{\deg p_1(x)} - 1$ .

Таким образом, разработку математической модели отказоустойчивой реализации раунда SPN-шифра целесообразно осуществлять с использованием модифицированного алгоритма коррекции на основе вычисления синдромов ошибки.

#### РАЗРАБОТКА МАТЕМАТИЧЕСКОЙ МОДЕЛИ ОТКАЗОУСТОЙЧИВОГО SPN-ШИФРА

К настоящему моменту проведены ряд исследований [10, 11] направленных на повышение отказоустойчивости реализации отдельных процедур криптоалгоритма AES, в ходе которых определен оптимальный набор полиномов-оснований для обработки 8-разрядных чисел:

$$\begin{aligned} p_1(x) &= x^4 \oplus x \oplus 1, \\ p_2(x) &= x^4 \oplus x^3 \oplus 1, \\ p_3(x) &= x^4 \oplus x^3 \oplus x^2 \oplus x \oplus 1. \end{aligned}$$

Тогда полное описание используемой ПСКВ будет иметь вид:

$$A = (\alpha_1(x), \alpha_2(x), \alpha_3(x), \alpha_4(x)), \quad (17)$$

где  $\alpha_1(x) = A \bmod p_1(x)$ ;

$$\alpha_2(x) = A \bmod p_2(x);$$

$$\alpha_3(x) = \alpha_1(x) \oplus \alpha_2(x);$$

$$\alpha_4(x) = (\alpha_1(x) \oplus x \cdot \alpha_2(x)) \bmod p_3(x).$$

Формирование четырех остатков в ПСКВ с тремя основаниями становится возможным за счет применения модифицированного алгоритма локализации и устранения ошибок [9]. Информационные остатки получаются в результате деления обрабатываемого числа на полиномы  $p_1(x)$  и  $p_2(x)$ , а контрольные – в результате вычислений согласно (13) с использованием полинома  $p_3(x)$  в качестве контрольного полинома  $p_{k+1}(x)$ .

Для описания разрабатываемой математической модели введем обозначения:  $A_{\text{вх.ПСС}}$  – входной массив данных в позиционной системе счисления (ПСС);  $A_{\text{вх.ПСКВ}}$  – входной массив данных в ПСКВ;  $S$  – результат выполнения процедуры SubBytes;  $SR$  – результат выполнения процедуры ShiftRows;  $M$  – результат выполнения процедуры MixColumns;  $ARK$  – результат выполнения процедуры AddRoundKey.

В соответствии со стандартом AES входное состояние обрабатываемого массива можно описать выражением:

$$A_{\text{вх.ПСС}} = \begin{bmatrix} A_{00} & A_{01} & A_{02} & A_{03} \\ A_{10} & A_{11} & A_{12} & A_{13} \\ A_{20} & A_{21} & A_{22} & A_{23} \\ A_{30} & A_{31} & A_{32} & A_{33} \end{bmatrix}, \quad (18)$$

где  $A_{vw}$  – представление байта в ПСС.

Так как в разработанной модели используется представление обрабатываемых чисел в ПСКВ, то перед началом раунда алгоритма необходимо выполнить преобразование входного массива из ПСС в ПСКВ.

Контрольные остатки в ПСКВ используются для обнаружения ошибок в ходе преобразований SubBytes и MixColumns, для остальных преобразований достаточно только двух информационных остатков.

После перевода каждого байта входного массива в безыбыточный вариант используемой ПСКВ массив входных данных раунда криптоалгоритма AES будет иметь вид:

$$A_{\text{вх. ПСКВ}} = \begin{bmatrix} (a_1(A_{00}), a_2(A_{00})) & (a_1(A_{01}), a_2(A_{01})) & (a_1(A_{02}), a_2(A_{02})) & (a_1(A_{03}), a_2(A_{03})) \\ (a_1(A_{10}), a_2(A_{10})) & (a_1(A_{11}), a_2(A_{11})) & (a_1(A_{12}), a_2(A_{12})) & (a_1(A_{13}), a_2(A_{13})) \\ (a_1(A_{20}), a_2(A_{20})) & (a_1(A_{21}), a_2(A_{21})) & (a_1(A_{22}), a_2(A_{22})) & (a_1(A_{23}), a_2(A_{23})) \\ (a_1(A_{30}), a_2(A_{30})) & (a_1(A_{31}), a_2(A_{31})) & (a_1(A_{32}), a_2(A_{32})) & (a_1(A_{33}), a_2(A_{33})) \end{bmatrix}, \quad (19)$$

где  $a_1(A_{vw}), a_2(A_{vw})$  – информационные остатки представления байта  $A_{vw}$ , рассчитанные в соответствии (17).

Тогда преобразование SubBytes в соответствии с [10] будет описываться выражением:

$$\begin{aligned} S_{vw} &= (f_{1sub}(A_{vw}), f_{2sub}(A_{vw}), f_{3sub}(A_{vw}), f_{4sub}(A_{vw})) = \\ &= (s_1^{vw}(x), s_2^{vw}(x), s_3^{vw}(x), s_4^{vw}(x)), \end{aligned} \quad (20)$$

где  $S_{vw}$  – результат реализации SubBytes;

$f_{1sub}, f_{2sub}, f_{3sub}, f_{4sub}$  – функции нахождения остатков  $S_{vw}$  по таблицам, представленным в [10];

$s_1^{vw}(x), s_2^{vw}(x), s_3^{vw}(x), s_4^{vw}(x)$  – остатки представления  $S_{vw}$  в ПСКВ.

В соответствии с [10] четыре остатка представления результата преобразования SubBytes хранятся в четырех таблицах.

Рассмотрим формирование таблицы хранения первого информационного остатка результата преобразования SubBytes на примере входного байта  $39_{16}$ . Известно, что входному байту  $39_{16}$  в классической реализации SubBytes соответствует байт  $12_{16}$ .

Входной байт  $39_{16}$  в безыбыточном варианте ПСКВ будет иметь вид:

$$\begin{aligned} 39_{16} &= 111001_2 = x^5 \oplus x^4 \oplus x^3 \oplus 1 = \\ &= (x^5 \oplus x^4 \oplus x^3 \oplus 1 \bmod p_1(x) | x^5 \oplus x^4 \oplus x^3 \oplus 1 \bmod p_2(x)) = (C_{16}, B_{16}). \end{aligned}$$

Соответствующий входному байту  $39_{16}$  выходной байт  $12_{16}$  в избыточном варианте ПСКВ будет иметь вид:

$$\begin{aligned} 12_{16} &= 10010_2 = x^5 \oplus x = \\ &= (x^5 \oplus x \bmod p_1(x) | x^5 \oplus x \bmod p_2(x) | x^5 \oplus x \bmod p_1(x) \oplus x^5 \oplus x \bmod p_2(x) | \\ &| (x^5 \oplus x \bmod p_1(x) \oplus x \cdot (x^5 \oplus x) \bmod p_2(x)) \bmod p_3(x)) = (1_{16}, B_{16}, A_{16}, 8_{16}). \end{aligned}$$

Тогда на пересечении строки  $C_{16}$  и столбца  $B_{16}$  будет расположен первый информационный остаток ( $1_{16}$ ) выходного байта, как показано на рисунке 3.

	0	...	B	...	F
0	9	...	3	...	1
1	9	...	e	...	6
...	...	...	...	...	...
C	4	...	1	...	9
...	...	...	...	...	...
F	9	...	d	...	f

Рисунок 3 – Фрагмент таблицы первого информационного остатка выходного значения SubBytes-преобразования

Таблицы хранения второго информационного остатка и двух контрольных остатков результата преобразования SubBytes формируются аналогичным образом.

В результате применения выражения (20) к каждому байту массива  $A_{\text{ex.ПСКВ}}$  будет получен массив  $S$ , элементами которого являются байты  $S_{vw}$ . Для устранения ошибок, возникших в байтах  $S_{vw}$ , необходимо выполнить алгоритм поиска и коррекции ошибок, согласно (14) и (15):

$$\delta_1(x) = s_3(x) \oplus s_3^*(x), \quad (21)$$

$$\delta_2(x) = s_4(x) \oplus s_4^*(x), \quad (22)$$

$$\dot{S}_{vw}(x) = \tilde{S}_{vw} \oplus f(\delta_1(x), \delta_2(x)) = \tilde{S}_{vw}(x) \oplus \Delta e_{vw}(x). \quad (23)$$

Так как корректируемый байт  $\tilde{S}_{vw}(x)$  представляется в виде двух остатков ( $s_1^{vw}(x)$  и  $s_2^{vw}(x)$ ), то и вектор ошибки  $\Delta e_{vw}(x)$  соответственно состоит из двух остатков ( $\Delta e_1^{vw}(x)$  и  $\Delta e_2^{vw}(x)$ ). Согласно правилам выполнения операции сложения в ПСКВ, выражение (23) можно записать в виде:

$$\begin{aligned} \dot{S}_{vw}(x) &= \tilde{S}_{vw}(x) \oplus \Delta e_{vw}(x) = (s_1^{vw}(x), s_2^{vw}(x)) \oplus (\Delta e_1^{vw}(x), \Delta e_2^{vw}(x)) = \\ &= (s_1^{vw}(x) \oplus \Delta e_1^{vw}(x), s_2^{vw}(x) \oplus \Delta e_2^{vw}(x)) \end{aligned} \quad (24)$$

Таким образом, ошибки, возникшие при табличном определении информационных остатков байтов массива  $S$ , будут устранены за счет избыточной информации контрольных остатков.

В соответствии с порядком реализации процедур криптоалгоритма AES массив  $\dot{S}$  поступает на ShifRows-преобразователь, который преобразовывает массив  $\dot{S}$  в массив  $SR$ .

В свою очередь, байты массива  $SR$  являются входными данными для реализации преобразования MixColumns, в ходе которого формируется массив  $M$  путем перемешивания и умножения на константы байтов  $SR_{vw}$ .

В ранее проведенных исследованиях [11] доказано, что результаты преобразования MixColumns в рассматриваемой ПСКВ можно представить в виде:

$$M_{vw} = (m_1^{vw}(x), m_2^{vw}(x), m_3^{vw}(x), m_4^{vw}(x)). \quad (25)$$

При этом остатки байта  $M^{vw}$  вычисляются согласно:

$$m_1^{vw}(x) = sr_1^{(w+2 \bmod 4)w} \oplus sr_1^{(w+3 \bmod 4)w} \oplus f_{1mix}(SR_{vw}) \oplus (f_{1mix}(SR_{(w+1 \bmod 4)w}) \oplus sr_1^{(w+1 \bmod 4)w}), \quad (26)$$

$$m_2^{vw}(x) = sr_2^{(w+2 \bmod 4)w} \oplus sr_2^{(w+3 \bmod 4)w} \oplus f_{2mix}(SR_{vw}) \oplus (f_{2mix}(SR_{(w+1 \bmod 4)w}) \oplus sr_2^{(w+1 \bmod 4)w}), \quad (27)$$

$$\begin{aligned} m_3^{vw}(x) &= (sr_1^{(w+2 \bmod 4)w} \oplus sr_2^{(w+2 \bmod 4)w}) \oplus (sr_1^{(w+3 \bmod 4)w} \oplus sr_2^{(w+3 \bmod 4)w}) \oplus f_{3mix}(SR_{vw}) \oplus \\ &\oplus (f_{3mix}(SR_{(w+1 \bmod 4)w}) \oplus (sr_1^{(w+1 \bmod 4)w} \oplus sr_2^{(w+1 \bmod 4)w})) \end{aligned} \quad (28)$$

$$\begin{aligned} m_4^{vw}(x) &= (sr_1^{(w+2 \bmod 4)w} \oplus x \cdot sr_2^{(w+2 \bmod 4)w}) \bmod p3(x) \oplus (sr_1^{(w+3 \bmod 4)w} \oplus x \cdot sr_2^{(w+3 \bmod 4)w}) \bmod p3(x) \oplus \\ &\oplus f_{4mix}(SR_{vw}) \oplus (f_{4mix}(SR_{(w+1 \bmod 4)w}) \oplus (sr_1^{(w+1 \bmod 4)w} \oplus x \cdot sr_2^{(w+1 \bmod 4)w}) \bmod p3) \end{aligned} \quad (29)$$

где  $f_{1mix} - f_{4mix}$  – функции нахождения остатков результатов умножения байтов  $SR_{vw}$  на константу  $2_{16}$  по таблицам, описанным в [11].

Таким образом, для реализации преобразования MixColumns в ПСКВ необходимы 4 таблицы. Контрольные остатки результатов умножения байтов на константу  $1_{16}$  вычисляются согласно (13). Операция умножения байта на константу  $3_{16}$  выполняется за счет дополнительного умножения этого байта на результат его умножения на константу  $2_{16}$ .

Рассмотрим пример реализации преобразования MixColumns для столбца  $S = [CA_{16}, D1_{16}, E2_{16}, 4F_{16}]$ . В безизбыточном варианте ПСКВ данный столбец будет иметь вид  $S = [(D_{16}, 2_{16}), (5_{16}, 0_{16}), (3_{16}, 1_{16}), (3_{16}, 0_{16})]$ . Допустим, данный столбец является нулевым столбцом обрабатываемого массива, тогда  $(D_{16}, 2_{16})$  необходимо умножить на  $2_{16}$ ,  $(5_{16}, 0_{16})$  – на  $3_{16}$ ,  $(3_{16}, 1_{16})$  и  $(3_{16}, 0_{16})$  – на  $1_{16}$ .

Получение первого остатка результата умножения на  $2_{16}$  будет осуществляться с помощью таблицы, представленной на рисунке 4. В математической записи такая процедура обозначена  $f_{1mix}$ . Как видно на рисунке 4, из таблицы хранения первого информационного остатка будет извлечено значение  $4_{16}$ .

	0	1	...	F
0	0	d	...	d
1	f	2	...	2
...	...	...	...	...
D	9	4	...	4
...	...	...	...	...
F	0	d	...	d

Рисунок 4 – Таблица хранения первых информационных остатков результатов умножения на константу  $2_{16}$

По аналогичному принципу из второй таблицы будет получено значение  $8_{16}$ , из третьей –  $C_{16}$ , из четвертой –  $B_{16}$ .

Для реализации умножения  $(5_{16}, 0_{16})$  на  $3_{16}$  необходимо:

1) табличным способом определить результат умножения  $(5_{16}, 0_{16})$  на  $2_{16}$ . Для рассматриваемого примера:

$$\begin{aligned} f_{1mix}(5_{16}, 0_{16}) &= 7_{16}; \\ f_{2mix}(5_{16}, 0_{16}) &= C_{16}; \\ f_{3mix}(5_{16}, 0_{16}) &= B_{16}; \\ f_{4mix}(5_{16}, 0_{16}) &= 0_{16}. \end{aligned}$$

2) к информационной части результата умножения на  $2_{16}$  прибавить информационные остатки  $(5_{16}, 0_{16})$ :

$$\begin{aligned} a_1(x) &= 7_{16} \oplus 5_{16} = x^2 \oplus x \oplus 1 \oplus x^2 \oplus 1 = x = 2_{16}; \\ a_2(x) &= C_{16} \oplus 0_{16} = x^3 \oplus x^2 = C_{16}. \end{aligned}$$

3) вычислить контрольные остатки для обрабатываемого значения согласно (13):

$$\begin{aligned} a_3(x) &= 5_{16} \oplus 0_{16} = x^2 \oplus 1 = 5_{16}; \\ a_4(x) &= 5_{16} \oplus (2_{16} \cdot 0_{16}) = (x^2 \oplus 1 \oplus x \cdot 0) \bmod p_3 = x^2 \oplus 1 = 5_{16}. \end{aligned}$$

4) сложить вычисленные контрольные остатки с контрольными остатками результата умножения на  $2_{16}$ :

$$\begin{aligned} a_3(x) &= B_{16} \oplus 5_{16} = x^3 \oplus x \oplus 1 \oplus x^2 \oplus 1 = x^3 \oplus x^2 \oplus x = E_{16}, \\ a_4(x) &= 0_{16} \oplus 5_{16} = x^2 \oplus 1 = 5_{16}. \end{aligned}$$

Таким образом, в результате умножения  $(5_{16}, 0_{16})$  на константу  $3_{16}$  будет получено значение  $(2_{16}, C_{16}, E_{16}, 5_{16})$ .

Для реализации умножения байтов на константу  $1_{16}$  необходимо информационные остатки оставить без изменения, а контрольные – вычислить согласно (13). Таким образом, значение  $(3_{16}, 1_{16})$  будет преобразовано в  $(3_{16}, 1_{16}, 2_{16}, 1_{16})$ , значение  $(3_{16}, 0_{16})$  – в  $(3_{16}, 0_{16}, 3_{16}, 3_{16})$ .

Теперь при наличии всех результатов умножения входных байтов на константы можно вычислить выходное значение 0-й строки 0-го столбца, проведя поэлементное сложение остатков согласно (26)–(29):

$$\begin{aligned} m_1^{vw}(x) &= x^2 \oplus x \oplus x \oplus 1 \oplus x \oplus 1 = x^2 \oplus x = 6_{16}; \\ m_2^{vw}(x) &= x^3 \oplus x^3 \oplus x^2 \oplus 1 \oplus 0 = x^2 \oplus 1 = 5_{16}; \\ m_3^{vw}(x) &= x^3 \oplus x^2 \oplus x^3 \oplus x^2 \oplus x \oplus x \oplus x \oplus 1 = x \oplus 1 = 3_{16}; \\ m_4^{vw}(x) &= x^3 \oplus x \oplus 1 \oplus x^2 \oplus 1 \oplus 1 \oplus x \oplus 1 = x^3 \oplus x^2 = C_{16}. \end{aligned}$$

Таким образом, в рассматриваемом примере для нулевого входного столбца  $S = [CA_{16}, D_{16}, E_{2_{16}}, 4F_{16}]$  значением нулевой строки нулевого столбца будет  $S_{\text{вых}} = (6_{16}, 5_{16}, 3_{16}, C_{16})$ .

В результате преобразования MixColumns формируется массив  $M$ , элементы которого состоят из 4 остатков (два информационных и два контрольных). Аналогично преобразованию SubBytes контрольные остатки используются для реализации процедуры обнаружения и коррекции ошибок с помощью модифицированного алгоритма на основе определения синдромов ошибки:

$$\delta_1(x) = m_3(x) \oplus m_3^*(x), \quad (30)$$

$$\delta_2(x) = m_4(x) \oplus m_4^*(x), \quad (31)$$

$$\dot{M}_{vw}(x) = \tilde{M}_{vw}(x) \oplus \Delta e^{vw}(x). \quad (32)$$

Таким образом, в шестнадцати байтах  $\dot{M}_{vw}$  массива  $\dot{M}$  будут отсутствовать ошибки, вызванные сбоями и отказами блоков замены, реализующих преобразование (25).

Последним раундовым преобразованием криптоалгоритма AES является преобразование AddRoundKey, заключающееся в поразрядном сложении каждого байта шифруемого блока с раундовой ключевой информацией  $K^j$  ( $j=1, 2, \dots, 10$ ), которая должна быть представлена в безызбыточной полиномиальной форме:

$$K^j = \{K_{vw}^j\} = \{(k_2^{vw}(x), k_2^{vw}(x))\}. \quad (33)$$

Тогда реализация преобразования AddRoundKey будет сведена к выполнению сложения байтов  $\dot{M}_{vw}$  и  $K_{vw}^j$  в соответствии с принципами выполнения суммирования по модулю два в системе остаточных классов:

$$ARK_{vw} = \dot{M}_{vw} \oplus K_{vw}^j = (\dot{m}_1^{vw}(x) \oplus k_1^{vw}(x), \dot{m}_2^{vw}(x) \oplus k_2^{vw}(x)) = (ark_1^{vw}(x), ark_2^{vw}(x)). \quad (34)$$

Массив  $ARK$ , полученный в результате выполнения процедуры AddRoundKey, будет являться выходным значением раунда криптоалгоритма.

Таким образом, совокупность выражений (18)–(34) будет являться математической моделью отказоустойчивого SPN-шифратора, реализующего один раунд криптоалгоритма AES.

SPN-шифратор, реализованный с использованием разработанной модели, будет способен сохранять работоспособность в случае возникновения сбоев в аппаратных средствах, обеспечивающих выполнение преобразований SubBytes и MixColumns, за счет наличия процедур выявления и устранения ошибок ( $f_i$  и  $f_{\dot{M}}$ ).

Рассмотрим операцию шифрования данных с использованием разработанной математической модели. Пусть входным массивом раунда SPN-шифра является массив (18), ранее используемый для описания классического варианта криптоалгоритма.

После преобразования из ПСС в безызбыточную ПСКВ данный массив будет иметь вид:

$$A_{\text{ск.ПСКВ}} = \begin{bmatrix} (A_{16}, 0_{16}) & (D_{16}, C_{16}) & (2_{16}, 4_{16}) & (8_{16}, A_{16}) \\ (8_{16}, F_{16}) & (6_{16}, E_{16}) & (1_{16}, 4_{16}) & (F_{16}, E_{16}) \\ (2_{16}, 0_{16}) & (3_{16}, 1_{16}) & (6_{16}, A_{16}) & (4_{16}, 7_{16}) \\ (0_{16}, B_{16}) & (D_{16}, 0_{16}) & (C_{16}, 1_{16}) & (8_{16}, 8_{16}) \end{bmatrix} \quad (35)$$

Тогда в результате применения к каждому байту выражения (20) из массива  $A$  должен получиться массив  $S$ :

$$S = \begin{bmatrix} (0_{16}, 5_{16}, 5_{16}, A_{16}) & (1_{16}, 3_{16}, 2_{16}, 7_{16}) & (6_{16}, D_{16}, B_{16}, F_{16}) & (D_{16}, 7_{16}, A_{16}, 3_{16}) \\ (1_{16}, C_{16}, D_{16}, A_{16}) & (1_{16}, A_{16}, B_{16}, 6_{26}) & (A_{16}, 1_{16}, B_{16}, 8_{16}) & (D_{16}, E_{16}, 3_{16}, 2_{16}) \\ (2_{16}, 8_{16}, A_{16}, 1_{16}) & (0_{16}, 6_{16}, 6_{16}, C_{16}) & (2_{16}, B_{16}, 9_{16}, 7_{16}) & (D_{16}, 4_{16}, 9_{16}, 5_{16}) \\ (3_{16}, 2_{16}, 1_{16}, 7_{16}) & (3_{16}, B_{16}, 8_{16}, 6_{16}) & (4_{16}, 6_{16}, 2_{16}, 8_{16}) & (5_{16}, 2_{16}, 7_{16}, 1_{16}) \end{bmatrix} \quad (36)$$

В случае отсутствия ошибок в ходе SubBytes массив  $\dot{S}$  будет состоять из информационных значений массива  $S$ :

$$\dot{S} = \begin{bmatrix} (0_{16}, 5_{16}) & (1_{16}, 3_{16}) & (6_{16}, D_{16}) & (D_{16}, 7_{16}) \\ (1_{16}, C_{16}) & (1_{16}, A_{16}) & (A_{16}, 1_{16}) & (D_{16}, E_{16}) \\ (2_{16}, 8_{16}) & (0_{16}, 6_{16}) & (2_{16}, B_{16}) & (D_{16}, 4_{16}) \\ (3_{16}, 2_{16}) & (3_{16}, B_{16}) & (4_{16}, 6_{16}) & (5_{16}, 2_{16}) \end{bmatrix}.$$

Массив, получаемый в результате процедуры ShiftRows, будет состоять из элементов массива  $\dot{S}$ , при этом положение элементов массива  $SR$  определяется в зависимости от его номера строки в массиве  $\dot{S}$ :

$$SR = \begin{bmatrix} (0_{16}, 5_{16}) & (1_{16}, 3_{16}) & (6_{16}, D_{16}) & (D_{16}, 7_{16}) \\ (1_{16}, A_{16}) & (A_{16}, 1_{16}) & (D_{16}, E_{16}) & (1_{16}, C_{16}) \\ (2_{16}, B_{16}) & (D_{16}, 4_{16}) & (2_{16}, 8_{16}) & (0_{16}, 6_{16}) \\ (5_{16}, 2_{16}) & (3_{16}, 2_{16}) & (3_{16}, B_{16}) & (4_{16}, 6_{16}) \end{bmatrix}. \quad (37)$$

Тогда результатом выполнения преобразования MixColumns над массивом  $SR$  в соответствии с (25) будет являться массив  $M$ , элементами которого являются значения из множества  $GF(2^8)$ , представленные в избыточном варианте используемой ПСКВ:

$$M = \begin{bmatrix} (4_{16}, 4_{16}, 0_{16}, C_{16}) & (1_{16}, 3_{16}, 2_{16}, 7_{16}) & (4_{16}, 7_{16}, 3_{16}, A_{16}) & (E_{16}, 3_{16}, D_{16}, 8_{16}) \\ (C_{16}, 2_{16}, E_{16}, 8_{16}) & (3_{16}, 3_{16}, F_{16}, A_{16}) & (A_{16}, 2_{16}, 8_{16}, E_{16}) & (6_{16}, 6_{16}, 0_{16}, A_{16}) \\ (A_{16}, 6_{16}, C_{16}, 6_{16}) & (A_{16}, 0_{16}, A_{16}, A_{16}) & (7_{16}, 5_{16}, 2_{16}, 3_{16}) & (0_{16}, D_{16}, D_{16}, 9_{16}) \\ (4_{16}, 6_{16}, 2_{16}, 8_{16}) & (2_{16}, 4_{16}, 6_{16}, A_{16}) & (3_{16}, 7_{16}, 4_{16}, D_{16}) & (0_{16}, 3_{16}, 3_{16}, 6_{16}) \end{bmatrix}. \quad (38)$$

Контрольные остатки массива  $M$  используются для обнаружения и исправления ошибок в информационных остатках. Если ошибки в информационных остатках не обнаружены, формируется массив  $\dot{M}$ , отличающийся от массива  $M$  отсутствием контрольных остатков:

$$\dot{M} = \begin{bmatrix} (4_{16}, 4_{16}) & (1_{16}, 3_{16}) & (4_{16}, 7_{16}) & (E_{16}, 3_{16}) \\ (C_{16}, 2_{16}) & (3_{16}, 3_{16}) & (A_{16}, 2_{16}) & (6_{16}, 6_{16}) \\ (A_{16}, 6_{16}) & (A_{16}, 0_{16}) & (7_{16}, 5_{16}) & (0_{16}, D_{16}) \\ (4_{16}, 6_{16}) & (2_{16}, 4_{16}) & (3_{16}, 7_{16}) & (0_{16}, 3_{16}) \end{bmatrix}.$$

Элементы  $\dot{M}$ , в свою очередь, складываются по модулю два с элементами ключевого массива  $K$  при реализации преобразования AddRoundKey. Допустим, применяется ключевой массив  $K$  (8), ранее используемый для описания классического варианта криптоалгоритма. Его представление в ПСКВ будет иметь вид:

$$K = \begin{bmatrix} (D_{16}, C_{16}) & (3_{16}, F_{16}) & (5_{16}, 8_{16}) & (C_{16}, 1_{16}) \\ (8_{16}, 0_{16}) & (B_{16}, 2_{16}) & (E_{16}, F_{16}) & (6_{16}, 8_{16}) \\ (C_{16}, 4_{16}) & (A_{16}, 7_{16}) & (C_{16}, B_{16}) & (F_{16}, B_{16}) \\ (4_{16}, E_{16}) & (B_{16}, 5_{16}) & (C_{16}, B_{16}) & (5_{16}, 5_{16}) \end{bmatrix}. \quad (39)$$

Тогда в соответствии с (34):

$$\begin{aligned} ARK = \dot{M} \oplus K &= \begin{bmatrix} (4_{16}, 4_{16}) & (1_{16}, 3_{16}) & (4_{16}, 7_{16}) & (E_{16}, 3_{16}) \\ (C_{16}, 2_{16}) & (3_{16}, 3_{16}) & (A_{16}, 2_{16}) & (6_{16}, 6_{16}) \\ (A_{16}, 6_{16}) & (A_{16}, 0_{16}) & (7_{16}, 5_{16}) & (0_{16}, D_{16}) \\ (4_{16}, 6_{16}) & (2_{16}, 4_{16}) & (3_{16}, 7_{16}) & (0_{16}, 3_{16}) \end{bmatrix} + \\ &+ \begin{bmatrix} (D_{16}, C_{16}) & (3_{16}, F_{16}) & (5_{16}, 8_{16}) & (C_{16}, 1_{16}) \\ (8_{16}, 0_{16}) & (B_{16}, 2_{16}) & (E_{16}, F_{16}) & (6_{16}, 8_{16}) \\ (C_{16}, 4_{16}) & (A_{16}, 7_{16}) & (C_{16}, B_{16}) & (F_{16}, B_{16}) \\ (4_{16}, E_{16}) & (B_{16}, 5_{16}) & (C_{16}, B_{16}) & (5_{16}, 5_{16}) \end{bmatrix} = \\ &= \begin{bmatrix} (9_{16}, 8_{16}) & (2_{16}, C_{16}) & (1_{16}, F_{16}) & (2_{16}, 2_{16}) \\ (4_{16}, 2_{16}) & (7_{16}, 1_{16}) & (4_{16}, D_{16}) & (0_{16}, E_{16}) \\ (6_{16}, 2_{16}) & (0_{16}, 7_{16}) & (B_{16}, 9_{16}) & (F_{16}, 6_{16}) \\ (0_{16}, 8_{16}) & (D_{16}, 0_{16}) & (F_{16}, C_{16}) & (5_{16}, 6_{16}) \end{bmatrix}. \quad (40) \end{aligned}$$

Массив  $ARK$  является результатом выполнения процедуры AddRoundKey и, соответственно, выходным значением раунда шифрования криптоалгоритма AES.

Массив, полученный в результате отказоустойчивой реализации в ПСКВ алгоритма AES, полностью совпадает с массивом, полученным в результате классической реализации алгоритма шифрования:

$$\begin{bmatrix} A_{16} & 68_{16} & 6B_{16} & 02_{16} \\ 9C_{16} & 9F_{16} & 5B_{16} & 6A_{16} \\ 7F_{16} & 35_{16} & EA_{16} & 50_{16} \\ F2_{16} & 2B_{16} & 43_{16} & 49_{16} \end{bmatrix} = \begin{bmatrix} (9_{16}, 8_{16}) & (2_{16}, C_{16}) & (1_{16}, F_{16}) & (2_{16}, 2_{16}) \\ (4_{16}, 2_{16}) & (7_{16}, 1_{16}) & (4_{16}, D_{16}) & (0_{16}, E_{16}) \\ (6_{16}, 2_{16}) & (0_{16}, 7_{16}) & (B_{16}, 9_{16}) & (F_{16}, 6_{16}) \\ (0_{16}, 8_{16}) & (D_{16}, 0_{16}) & (F_{16}, C_{16}) & (5_{16}, 6_{16}) \end{bmatrix}.$$

Таким образом, рассмотренный пример подтверждает адекватность разработанной математической модели и её непротиворечивость известным работам в данной предметной области.

Допустим, в ходе реализации процедуры SubBytes произошел сбой, результатом которого стало искажение первого информационного остатка элемента  $S_{03}$  на величину  $\Delta S_1^{03}(x) = x^3$ , т. е. искаженное значение  $\tilde{S}_{03}$  будет иметь вид:  $\tilde{S}_{03} = (\tilde{5}_{16}, 7_{16}, A_{16}, 3_{16})$ .

Тогда при реализации процедуры обнаружения и коррекции ошибок для элемента  $\tilde{S}_{03}$  из информационных остатков  $(\tilde{5}_{16}, 7_{16})$  будут получены следующие проверочные контрольные остатки:

$$s_3^*(x) = x^2 \oplus x \oplus 1 \oplus x^2 \oplus 1 = x,$$

$$s_4^*(x) = (x^2 \oplus 1 \oplus x \cdot (x^2 \oplus x \oplus 1)) \bmod x^4 \oplus x^3 \oplus x^2 \oplus 1 = x^3 \oplus x \oplus 1.$$

На основе этих остатков и контрольных остатков, полученных табличным способом, рассчитываются синдромы ошибки согласно (21), (22):

$$\delta_1(x) = A_{16} \oplus 2_{16} = x^3 \oplus x \oplus x = x^3,$$

$$\delta_2(x) = 3_{16} \oplus B_{16} = x \oplus 1 \oplus x^3 \oplus x \oplus 1 = x^3.$$

С учетом выражения (16) вектор ошибки будет иметь вид:

$$\Delta e(x) = (x^3, 0).$$

Коррекция возникшей ошибки выполняется в ходе сложения информационных остатков искаженного элемента и вектора ошибки в соответствии с (24):

$$\begin{aligned} \dot{S}_{03}(x) &= (x^2 \oplus 1, x^2 \oplus x \oplus 1) + (x^3, 0) = (x^2 \oplus 1 \oplus x^3, x^2 \oplus x \oplus 1 \oplus 0) = \\ &= (x^3 \oplus x^2 \oplus 1, x^2 \oplus x \oplus 1) = (D_{16}, 7_{16}). \end{aligned}$$

Таким образом, в ходе выполнения процедуры обнаружения и коррекции ошибки была исправлена ошибка, возникшая в ходе преобразования SubBytes.

Аналогичным образом могут быть исправлены однократные ошибки, возникающие в ходе реализации преобразования MixColumns.

### ЗАКЛЮЧЕНИЕ

Разработанная математическая модель SPN-шифратора обеспечивает выполнение криптоалгоритма AES, несмотря на возникающие в процессе однократные сбои в работе шифратора. Такие свойства были получены за счет применения математического аппарата полиномиальной системы классов вычетов, обеспечивающего обнаружение и исправление ошибок. Данную математическую модель целесообразно использовать при построении шифратора американского стандарта шифрования, способного сохранять работоспособное состояние в условиях сбоев SPN-преобразователя. Перспективным направлением исследования является разработка математических моделей отказоустойчивой реализации дешифратора американского стандарта шифрования и системы передачи данных с реализацией криптографической защиты информации на основе американского стандарта шифрования.

### Список источников

1. Папуловская, Н. В. Основы интернета вещей : учебно-методическое пособие / Н. В. Папуловская. – Екатеринбург : Урал. ун-та, 2022. – 104 с.
2. Калмыков, И. А. Математические модели нейросетевых отказоустойчивых вычислительных средств, функционирующих в полиномиальной системе классов вычетов / И. А. Калмыков ; под ред. Н. И. Червякова. – Москва : ФИЗМАТЛИТ, 2005. – 276 с.
3. Зензин, О. С. Стандарт криптографической защиты – AES. Конечные поля / О. С. Зензин, М. А. Иванова ; под ред. М. А. Иванова. – Москва : КУДИЦ-ОБРАЗ, 2002. – 176 с.
4. Рацеев, С. М. Реализации некоторых криптосистем и корректирующих кодов / С. М. Рацеев. – Санкт-Петербург : Лань, 2024. – 288 с.
5. Червяков, Н. И. Элементы компьютерной математики и нейроинформатики / Н. И. Червяков, И. А. Калмыков, В. А. Галкина и др. – Москва : Физико-математическая литература, 2003. – 216 с.
6. Червяков, Н. И. Модулярная арифметика и ее приложения в инфокоммуникационных технологиях / Н. И. Червяков и др. – Москва : ФИЗМАТЛИТ, 2017 – 400 с.
7. Кларк, Дж. Кодирование с исправлением ошибок в системах цифровой связи / Дж. Кларк. – Москва : Радио и связь, 1987. – 392 с.

8. Калмыков, И. А. Применение корректирующих кодов полиномиальной системы классов вычетов для построения спецпроцессоров цифровой обработки сигналов / И. А. Калмыков, В. Р. Гахов и др. // Труды Международного форума по проблемам науки, техники и образования / под ред. В. П. Савиных, В. В. Вишневого. – Москва : Академия наук, 2004. – Т. 1 – С. 133–135.

9. Калмыков, И. А. Алгоритм коррекции ошибок в модулярном коде классов вычетов, обеспечивающий повышение отказоустойчивости систем OFDM / И. А. Калмыков, В. С. Сляднев, М. И. Калмыков и др. // Инженерный вестник Дона. – 2024. – № 2 (110). – С. 224–237.

10. Проворнов, И. А. Исследование корректирующей способности модулярных кодов, применяемых в AES-системах / И. А. Проворнов // Проблемы разработки перспективных микро- и наноэлектронных систем (МЭС). – 2022. – № 4. – С. 136–141.

11. Проворнов, И. А. Разработка математической модели отказоустойчивого преобразователя MixColumns SPN-шифра в полиномиальной системе классов вычетов. / И. А. Проворнов, И. А. Калмыков // Приборы. – 2024. – № 1 (283). – С. 25–34.

#### References

1. Papulovskaya, N. V. *Fundamentals of the Internet of Things: Educational manual*. Ekaterinburg, 2022. 104 p. (In Russ.).

2. Kalmykov, I. A. Chervyakov, N. I. (ed.). *Mathematical models of neural network fault-tolerant computing tools operating in a polynomial system of deduction classes*. Moscow, 2005. 276 p. (In Russ.).

3. Zenzin, O. S., Ivanov, M. A. *The cryptographic security standard is AES. End fields*. Moscow, 2002. 176 p. (In Russ.).

4. Ratseev, S. M., *Implementations of some cryptosystems and corrective codes*. Saint-Petersburg, 2024, 288 p. (In Russ.).

5. Chervyakov, N. I., Kalmykov, I. A., Galkina, V. A. *Elements of computer mathematics and neuroinformatics*. Moscow, 2003. 216 p. (In Russ.).

6. Chervyakov, N. I., Lavrinenko, I. N., Lavrinenko, A. V., Kolyada, A. A., Lyakhov, P. A., Babenko, M. G. *Modular arithmetic and its applications in information and communication technologies*. Moscow, 2017. 400 p. (In Russ.).

7. Clark, Dzh. *Error Correction Coding in Digital Communication Systems*. Moscow, 1987. 392 p. (In Russ.).

8. Kalmykov, I. A., Gakhov, V. R. Application of correcting codes of a polynomial system of residue classes for the construction of special digital signal processing processors. *Proceedings of the International Forum on Science, Technology and Education*. Moscow, Academy of Sciences, 2004, pp. 133–135 (In Russ.).

9. Kalmykov I. A., Slyadnev V. S., Kalmykov M. I. Error correction algorithm in modular code of residue classes, providing increased fault tolerance of OFDM systems. *Engineering Bulletin of Don*, 2024, pp. 224–237 (In Russ.).

10. Provornov, I. A. Investigation of corrective ability of modular codes used in AES systems. *Problems of developing promising micro- and nanoelectronic systems (MES)*, 2022, pp. 136–141 (In Russ.).

11. Provornov I.A., Kalmykov I.A., Development of a mathematical model of a fault-tolerant converter MixColumns SPN cipher in the polynomial system of residue classes. *Devices*, 2024, pp. 25–34 (In Russ.).

Статья поступила в редакцию 10.06.2024; одобрена после рецензирования 01.07.2024; принята к публикации 13.07.2024.

The article was submitted 10.06.2024; approved after reviewing 01.07.2024; accepted for publication 13.07.2024.

## **МАТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ, ЧИСЛЕННЫЕ МЕТОДЫ И КОМПЛЕКСЫ ПРОГРАММ**

УДК 004.001

### **МИНИМИЗАЦИЯ ЧИСЛА ВХОДНЫХ ПРИЗНАКОВ В ЗАДАЧАХ МЕДИЦИНСКОЙ ДИАГНОСТИКИ**

**Чернавин Павел Федорович**, Уральский федеральный университет, 620002, Российская Федерация, г. Екатеринбург, ул. Мира, 19,  
кандидат экономических наук, доцент, ORCID: 0000-0003-3214-3906, e-mail: chernavin.p.f@gmail.com

**Чернавин Николай Павлович**, Уральский федеральный университет, 620002, Российская Федерация, г. Екатеринбург, ул. Мира, 19,  
ассистент кафедры, ORCID: 0000-0002-2093-9715, e-mail: ch\_k@mail.ru

**Чернавин Федор Павлович**, Уральский федеральный университет, 620002, Российская Федерация, г. Екатеринбург, ул. Мира, 19,  
кандидат экономических наук, доцент, ORCID: 0000-0003-4105-231X, e-mail: chernavin\_fedor@mail.ru

**Черняев Игорь Анатольевич**, Уральский государственный медицинский университет Минздрава России, Российская Федерация, 620028, г. Екатеринбург, ул. Репина, 3,  
старший преподаватель кафедры общественного здоровья и здравоохранения, ORCID: 0000-0002-2439-7087, e-mail: smse@mail.ru

Обычно при решении большинства задач классификации требуется, чтобы показатели качества решающего правила (Accuracy, Precision, AUC ROC) были выше заданных границ при минимальном количестве входных признаков. Кроме того, желательно, чтобы решающее правило было хорошо интерпретируемо и понятно заказчику исследования, а получение выбранных входных признаков было доступно и недорого стоило. Наиболее остро все указанные требования предъявляются при решении задач медицинской диагностики на основе различного рода анализов. В данной статье приведены результаты исследования на основе 8000 наблюдений по 240 нозологиям на основе биохимического, микроэлементологического, аминокислотного и витаминного анализа различных субстратов (сыворотка крови и волосы). В результате исследований выявлено, что 30 заболеваний у женщин и 23 заболевания у мужчин с высокой вероятностью могут быть предсказаны различными методами машинного обучения на основании всего двух входных параметров: возраста и индекса массы тела. Данные исследования приведены в соответствующих таблицах. Результаты имеют теоретический и практический интерес.

**Ключевые слова:** классификация, машинное обучение, линейное разделение, медицинская диагностика, индекс массы тела

### **MINIMIZING THE NUMBER OF INPUT FEATURES IN MEDICAL DIAGNOSTICS**

**Chernavin Pavel F.**, Ural Federal University, 19 Mira St., Ekaterinburg, 620002, Russian Federation, Cand. Sci. (Economics), Associate Professor, ORCID: 0000-0003-3214-3906, e-mail: chernavin.p.f@gmail.com

**Chernavin Nikolai P.**, Ural Federal University, 19 Mira St., Ekaterinburg, 620002, Russian Federation, department assistant, ORCID: 0000-0002-2093-9715, e-mail: ch\_k@mail.ru

**Chernavin Fedor P.**, Ural Federal University, 19 Mira St., Ekaterinburg, 620002, Russian Federation, Cand. Sci. (Economics), Associate Professor, ORCID: 0000-0003-4105-231X, e-mail: chernavin\_fedor@mail.ru

**Cherniaev Igor A.**, Ural State Medical University of the Ministry of Health of the Russian Federation, 3 Repina St., Ekaterinburg, 620028, Russian Federation,  
SeniorLecturer, ORCID: 0000-0002-2439-7087, e-mail: smse@mail.ru

Usually, when solving most classification problems, it is required that the quality indicators of the decision rule (Accuracy, Precision, AUC ROC) be above the specified limits with a minimum number of input features. In addition, it is desirable that the decision rule be well interpreted and understood by the customer of the study, and obtaining the selected input features should be accessible and not expensive. All these requirements are most acute when solving problems of medical diagnostics based on various types of analyses. This article presents the results of a study based on 8000 observations in 240 nosologies with the following input parameters: the content of amino acids, vitamins and biochemistry of blood serum, the content of trace elements in the hair. As a result of the research, it was found that 30 diseases in women and diseases in men can be predicted with high probability by various machine learning methods based on only two input parameters: age and body mass index. Research data are given in the corresponding tables. The results are of theoretical and practical interest.

**Keywords:** classification, machine learning, linear separation, medical diagnostics, body mass index

## **ВВЕДЕНИЕ**

Данное исследование является логическим продолжением подхода по управлению качеством решающих правил (РП), приведенного в статье [1] с акцентом на минимизацию числа входных признаков и описанием практических результатов по прогнозированию 240 заболеваний при 106 входных параметрах (биохимический, аминокислотный и витаминный состав сыворотки крови, содержание различных микроэлементов в волосах) на основе 8000 наблюдений (6200 по женщинам и 1800 по мужчинам). Необходимость минимизации числа входных признаков при построении РП в задачах классификации отмечается многими авторами [2, 3]. С точки зрения машинного обучения (МО) минимизация признаков необходима для сокращения времени обучения и уменьшения вероятности переобучения. Кроме того, минимизация признаков в большинстве случаев приводит к увеличению точности прогноза. С точки зрения медицинских работников, минимизация числа признаков способствует лучшей интерпретируемости РП. Для пациента минимум анализов – это реальная экономия времени на их сдачу и денег на их оплату. Последнее с развитием платной медицины становится все более актуальным, так как сдача ряда анализов является достаточно дорогостоящим мероприятием, и их сдача без необходимости ничего кроме раздражения у пациента не вызывает.

Естественно, что минимизация числа признаков не должна ухудшать качество РП. Так как в задачах медицинской диагностики обычно классы сильно не сбалансированы, то в качестве критерия качества РП в большинстве случаев используется площадь под ROC-кривой (AUC ROC). Отметим, что в двухклассовых задачах классификации в случае, если выходной параметр принимает значения 0 (нет заболевания) и 1 (есть заболевание), значение AUC ROC совпадает со средним Recall по двум классам. Средний Recall лучше воспринимается практическими специалистами и гораздо проще формализуется для включения в модель, приведенную в [1].

## **ПОСЛЕДОВАТЕЛЬНОСТЬ РАСЧЕТОВ**

Изначально мы хотели определить минимальное количество признаков при построении РП для прогнозирования гипертонии и гипотензии. Оба заболевания относятся к наиболее распространенным заболеваниям сердечно-сосудистой системы. Гипертония является основной причиной инсультов, а гипотензия приводит к кислородному голоданию мозга и других органов. Первоначально в качестве входных признаков использовались возраст, индекс массы тела и 28 показателей биохимии сыворотки крови. Генеральная выборка содержала наблюдения по 900 женщинам и 600 мужчинам. Расчеты для каждого пола проводились отдельно по модели, приведенной в [1]. Мы начали расчеты с линейного разделения множеств, дающего наиболее интерпретируемые результаты. В качестве критерия качества РП использовалось ограничение  $AUC\ ROC > 0,75$  при минимуме признаков. В результате расчетов мы получили несколько неожиданное для нас решение: для достижения требуемых показателей достаточно всего два признака – возраст и индекс массы тела.

Существенной проблемой при анализе медицинских данных является наличие пустых значений. Заполнение пустых значений называется импутацией данных. Проблема импутации данных может решаться различными способами, но в данной статье нет необходимости их рассматривать, так как возраст, рост и вес имеются в большинстве медицинских наблюдений. Этот факт позволил нам увеличить генеральную выборку до 6200 наблюдений по женщинам и 1800 по мужчинам. После этого гипотеза о том, что можно построить РП с  $AUC\ ROC > 0,75$  о наличии гипертонии и гипотензии на основе линейного разделения множеств, была проверена и подтверждена на этих выборках.

Естественно, что после этого мы решили выяснить, какие еще заболевания можно с критерием качества  $AUC\ ROC > 0,75$  прогнозировать на основе возраста и индекса массы тела. При этом мы решили не ограничиваться только линейным разделением, а использовать и другие методы. Результаты расчетов приведены в таблицах 1 и 2. Для компактности представления данных мы использовали следующие обозначения: МКБ – международный классификатор болезней, RF – случайный лес, NB – наивный байесовский метод, KN – метод ближайших соседей, LD – линейный дискриминант, Асс – Ассигасу, AUC – площадь под ROC-кривой. В таблицах приведены результаты по тестовой выборке. Результаты по обучающей выборке выше. Так как мы располагали достаточно большим количеством данных, то разбиение на обучающую и тестовую производилось случайным образом в соотношении 60/40. Результаты расчетов приведены в таблицах 1 и 2.

Таблица 1 – Результаты расчетов для женщин в %

МКБ-10	Название	RF		NB		KN		LD	
		Acc	AUC	Acc	AUC	Acc	AUC	Acc	AUC
D25	Лейомиома матки	89,1	77,7	88,1	76,9				
E11	Инсулиннезависимый сахарный диабет	98,8	95,6	97,5	94,0	98,5	78,1	98	92
E28.1	Избыток андрогенов	97,6	77,5	98,6	75,7				
E28.2	Синдром поликистозных яичников	98,4	79,2	98,4	75,4				
E28.3	Первичная яичниковая недостаточность	98,5	87,4	98,5	84,7				
E66.9	Ожирение	89,9	94,1	88,9	92,0	89,6	90,1	90	93
F90.0	Синдром дефицита внимания и гиперактивности	99,0	95,7	97,4	95,3	99	90,4	99	96
I10	Гипертензия	89,1	92,3	87,7	91,5	87,9	88,9	89	92
I70	Атеросклероз	94,9	89,5	94	88,6	94,8	81,1	95	89
I95	Гипотензия	79,2	81,7	78,3	79,2	75,5	78,9	78	78
K26	Язва двенадцатиперстной кишки	98,6	77,4	98,6	75,1				
K76.0	Жировая печень	90,0	92,1	90,1	0,912	88,2	87,5	90	92
K80	Желчекаменная болезнь	97,6	84,7	97,1	80,3			98	80
K81	Холестит	92,9	80,6	91,6	80,1			92	80
L70	Угри	83,3	81,7	83,3	80,4				
M16	Артроз тазобедренного сустава	93,3	81,8	91,4	81,8			93	83
M19.9	Артроз неуточненный	91,6	87,7	91,5	88,3	90,8	78,1	91	88
M21	Другие деформации конечностей	91,3	75,5	91,3	75,2				
M41	Сколиоз	83,1	78,0	82,5	75,3			82	75
M81.9	Остеопороз неуточненный	94,0	84,5	94,1	83			94	84
N86	Эрозия и эктропион шейки матки	97,5	78,5	97,5	75,8				
N91.2	Аменорея неуточненная	98,1	79,7						
N94.0	Боли в середине менструального цикла	87,3	78,9	87,3	77,2				
N95.1	Менопауза и климактерическое состояние	89,9	95,1	89,1	94			87	95
N95.3	Состояние с искусственно вызванной менопаузой	99	83,2	99	80,4	88,6	93,9	99	81
N97	Женское бесплодие	97,2	78,2	97,2	75,9				
N97.9	Женское бесплодие неуточненное	98,5	80,7	98,5	81,9				
Q78.9	Остеохондродисплазия неуточненная	98,5	76,0						
R41.3	Другие амнезии	91,6	77,1	90,3	77,2			92	78
R73.0	Повышенное содержание глюкозы в крови	97,8	84,1	96,6	83			97	85

Таблица 2 – Результаты расчетов для мужчин в %

МКБ-10	Название	RF		NB		KN		LD	
		Acc	AUC	Acc	AUC	Acc	AUC	Acc	AUC
E11	Инсулиннезависимый сахарный диабет	96,8	84,5					97,5	81,9
E66.9	Ожирение	87,8	92,5			86,4	88,0	86,7	90,7
F84	Первазивные нарушения развития	98,1	84,0					98,0	90,2
F90.0	Синдром дефицита внимания и гиперактивности	94,7	95,6			96,0	86,6	94,3	93,2
I10	Гипертензия	80,1	85,1			76,3	80,0	80,5	86,5
I70	Атеросклероз	93,9	92,7			93,4	88,4	95,2	92,6

Продолжение таблицы 2

I95	Гипотензия	77,1	76,5					76,6	77,5
K76.0	Язва двенадцатиперстной кишки	86,4	85,2			86,7	78,4	87,6	88,3
K81	Холестит	94,4	81,2					93,9	78,0
L63.0	Алоpecia тотальная	98,7	81,7	98,7	86,9			98,1	76,1
L65	Нерубцующая потеря волос	92,0	79,4						
L70	Угри	91,2	80,8	91,2	78,2	90,2	77,4		
L80	Витилиго-хроническое заболевание			96,5	75,8				
M16	Артроз тазобедренного сустава	96,3	76,6	96,3	81,0			95,7	80,7
M19.9	Артроз неуточненный	95,7	77,2	95,7	82,3			95,9	84,5
M21	Другие деформации конечностей	92,3	78,9	91,5	80,2			90,7	78,0
M41	Сколиоз	90,2	77,1	74,7	77,0				
M42	Остеохондроз позвоночника	74,5	74,9	96,5	89,8				
N20	Камни почки и мочеточника							96,8	75,2
N40	Аденома предстательной железы	96,5	89,8	92,8	74,9			95,9	90,4
N41.1	Хронический простатит	92,6	76,1					93,0	81,4
R42	Головокружение и нарушение устойчивости	99,2	76,7						

### АНАЛИЗ РЕЗУЛЬТАТОВ РАСЧЕТОВ

Из приведенных таблиц следует, что для женщин наибольшее количество нозологий (30) при входных параметрах возраст и индекс массы тела можно прогнозировать на основе случайного леса и наивного байесовского метода (28). Наименьшее (8) на основе метода ближайших соседей. Линейное разделение множеств хотя оно и дает промежуточный результат (17), зато наиболее интерпретируемо и применимо на практике, так как дает результат в аналитической форме в виде уравнения разделяющей линии. Кроме того, по ряду заболеваний для женщин (гипертензия, гипотензия, атеросклероз, артроз неуточненный и т. д.) все приведенные методы МО дают прогноз с высокими качественными характеристиками. Поэтому на их основе можно создать ансамбль с более высокой точностью прогноза.

Для мужчин имеем следующие результаты: случайный лес (20), наивный байесовский метод (9), метод ближайших соседей (6) и линейное разделение (16). Причиной меньшего количества хорошо прогнозируемых нозологий в данном случае является, в первую очередь, меньшее число наблюдений по мужчинам. По этой причине ряд нозологий не рассматривался ввиду недостаточности данных.

Для дальнейшего анализа и интерпретации полученных решений нам было необходимо определить информативность признаков. Так как наилучшие результаты показал метод случайного леса, то именно этим методом была сделана оценка информативности признаков. Данные приведены в таблицах 3 и 4. Суммарная информативность равна единице.

Таблица 3 – Информативность признаков для женщин

МКБ-10	Название	Возраст	ИМТ
D25	Лейомиома матки	0,685012	0,314988
E11	Инсулиннезависимый сахарный диабет	0,434565	0,565435
E28.1	Избыток андрогенов	0,437917	0,562083
E28.2	Синдром поликистозных яичников	0,427107	0,572893
E28.3	Первичная яичниковая недостаточность	0,468950	0,531050
E66.9	Ожирение	0,118132	0,881868
F90.0	Синдром дефицита внимания и гиперактивности	0,467461	0,532539
I10	Гипертензия	0,677146	0,322854
I70	Атеросклероз	0,708116	0,291884
I95	Гипотензия	0,610924	0,389076
K26	Язва двенадцатиперстной кишки	0,453299	0,546701
K76.0	Жировая печень	0,292162	0,707838
K80	Желчекаменная болезнь	0,488211	0,511789

Продолжение таблицы 3

K81	Холестицит	0,471601	0,528399
L70	Угри	0,713785	0,286215
M16	Артроз тазобедренного сустава	0,599001	0,400999
M19.9	Артроз неуточненный	0,774182	0,225818
M21	Другие деформации конечностей	0,589119	0,410881
M41	Сколиоз	0,602274	0,397726
M81.9	Остеопороз неуточненный	0,698014	0,301986
N86	Эрозия и эктропион шейки матки	0,425617	0,574383
N91.2	Аменорея неуточненная	0,402762	0,597238
N94.0	Боли в середине менструального цикла	0,729975	0,270025
N95.1	Менопауза и климактерическое состояние	0,859116	0,140884
N95.3	Состояние с искусственно вызванной менопаузой	0,34607	0,65393
N97	Женское бесплодие	0,483988	0,516012
N97.9	Женское бесплодие неуточненное	0,375865	0,624135
Q78.9	Остеохондродисплазия неуточненная	0,366809	0,633191
R41.3	Другие амнезии	0,605279	0,394721
R73.0	Повышенное содержание глюкозы в крови	0,348328	0,651672

Таблица 4 – Информативность признаков для мужчин

МКБ-10	Название	Возраст	ИМТ
E11	Инсулиннезависимый сахарный диабет	0,45559	0,54441
E66.9	Ожирение	0,167563	0,832437
F84	Первазивные нарушения развития	0,420723	0,579277
F90.0	Синдром дефицита внимания и гиперактивности	0,551843	0,448157
I10	Гипертензия	0,549272	0,450728
I70	Атеросклероз	0,608382	0,391618
I95	Гипотензия	0,458197	0,541803
K76.0	Язва двенадцатиперстной кишки	0,335764	0,664236
K81	Холестицит	0,419857	0,580143
L63.0	Алопеция тотальная	0,365617	0,634383
L65	Нерубцующая потеря волос	0,43969	0,56031
L70	Угри	0,586735	0,413265
L80	Витилиго-хроническое заболевание	0,327271	0,672729
M16	Артроз тазобедренного сустава	0,532794	0,467206
M19.9	Артроз неуточненный	0,467082	0,532918
M21	Другие деформации конечностей	0,43861	0,56139
M41	Сколиоз	0,469399	0,530601
M42	Остеохондроз позвоночника	0,575559	0,424441
N20	Камни почки и мочеточника	0,447362	0,552638
N40	Аденома предстательной железы	0,518954	0,481046
N41.1	Хронический простатит	0,491639	0,508361
R42	Головокружение и нарушение устойчивости	0,353445	0,646555

На основе таблиц 3 и 4 приведенные в них заболевания можно разбить на три группы: доминирует возраст, доминирует ИМТ и влияние признаков примерно одинаково.

### ГЕОМЕТРИЧЕСКАЯ ИНТЕРПРЕТАЦИЯ ПОЛУЧЕННЫХ РЕШЕНИЙ

Факт того, что ряд заболеваний можно с высокой вероятностью прогнозировать на основе двух входных признаков, используя метод линейного разделения, предоставляет широкие возможности по наглядной геометрической интерпретации полученных решений, так как множества наблюдений и РП могут быть изображены графически. Кроме того, при использовании метода линейного разделения мы получаем решение в аналитическом виде, т. е. в виде формулы. Для заболевания E11 эта формула выглядит следующим образом:

$$BMI = 59.73 - 0.551 * AGE, \quad (1)$$

где  $BMI$  – критический индекс массы тела;

$AGE$  – возраст пациента.

Под критическим индексом массы тела понимается значение ИМТ для данного возраста, после которого начинает резко возрастать вероятность наличия заболевания. Графически это выглядит следующим образом (рис. 1 и 2).

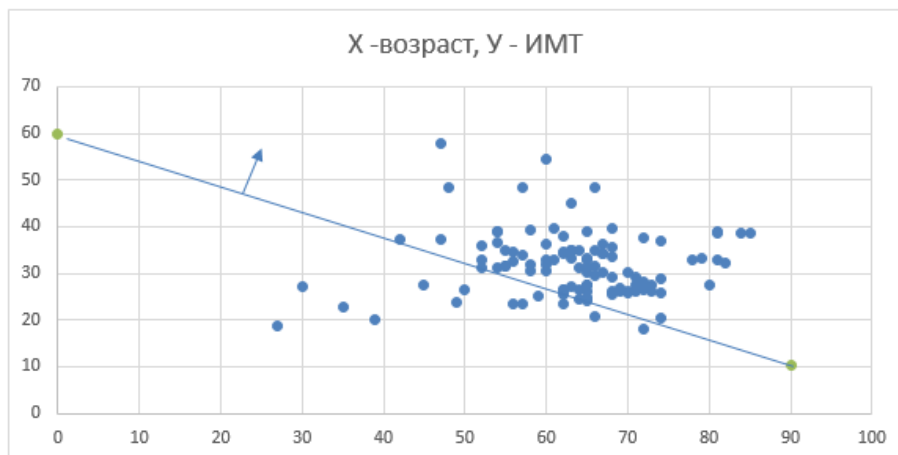


Рисунок 1 – Линейное разделение множества больных для заболевания E11

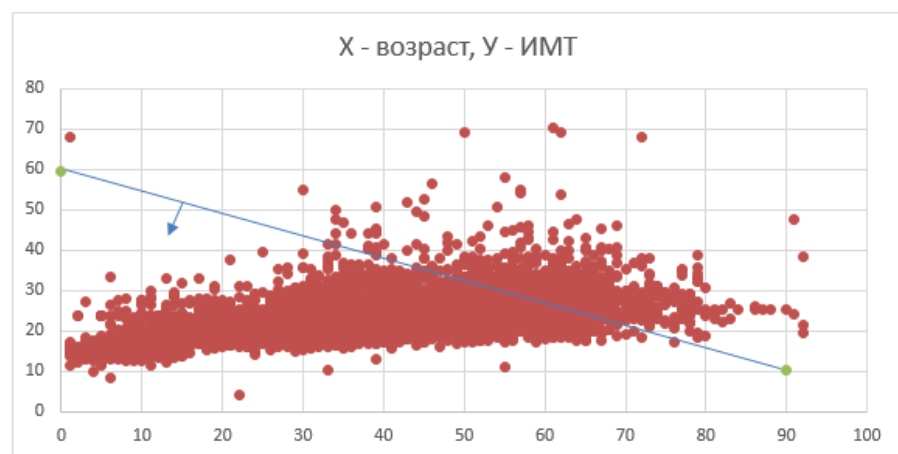


Рисунок 2 – Линейное разделение множества здоровых для заболевания E11

Наглядная геометрическая интерпретация в случае двух признаков очень удобна для практических медицинских работников даже если они не умеют использовать методы машинного обучения в своей работе. Например, на основе приведенных рисунков любой практик быстро сообразит, что все наблюдения, соответствующие больным, находятся выше уровня ИМТ = 18,1. Конечно, такое правило гораздо менее точное, но проще для восприятия и выработки рекомендаций.

#### **СРАВНЕНИЕ ИССЛЕДОВАНИЙ, ПРИВЕДЕННЫХ В СТАТЬЕ С ИССЛЕДОВАНИЯМИ ДРУГИХ АВТОРОВ**

Обобщенные зарубежными авторами данные обсервационных исследований посредством Менделевской рандомизации (Mendelian randomization) указывают на наличие связи между индексом массы тела (ИМТ) и хроническими заболеваниями, такими как сахарный диабет 2 типа (инсулиннезависимый диабет), 14 заболеваний системы кровообращения, астма, хроническая обструктивная болезнь легких, заболевания пищеварительной системы, заболеваний опорно-двигательного аппарата, а также рака пищеварительной системы (шесть вариантов рака), матки, почек и мочевого пузыря [5–7]. В нескольких исследованиях ИМТ критикуется за то, что он не учитывает распределение жира и не отличает жир от мышечной массы и плотности костной ткани (Bell et al., 2018), что делает ИМТ неспособным предсказать долгосрочные последствия для здоровья (Romero-Corral et al., 2008) [8, 9]. Вместе с тем ряд исследователей указывает на то, что траектории изменения значения ИМТ, такие как нелинейные временные тренды и нелинейные изменения ИМТ с возрастом, могут предоставить информацию о лежащих в основе временных закономерностях состояния здоровья [10]. Результаты нашего исследования сходны с аналогичными (Watson, C et al. 2021), основанными на динамических линейных, квадратичных, кубических и естественных сплайновых моделях, которые демонстрируют наличие ассоциаций возраста и о ИМТ с повышенным риском развития ряда заболеваний [11]. В аналогичном нашему исследованию (Ruixuan, 2020), изучающем зависимость между возрастом, составом крови и ИМТ, было установлено, что содержание глюкозы

в крови позволяет отличить пациентов с раком молочной железы от пациентов без рака молочной железы с низкой вероятностью ошибочных оценок и определяет перспективы применения результатов подобных методов в практике [12].

### ЗАКЛЮЧЕНИЕ

В данной статье приведены результаты исследований более чем по 30 нозологиям по методике, предложенной в [1]. Достоинство данной методики состоит в том, что на основе моделей математического программирования можно не только одновременно управлять качеством РП, но и минимизировать число наиболее информативных признаков для достижения желаемого результата. Кроме того, решение получается в аналитическом виде, и возможна его геометрическая интерпретация. По всем нозологиям могут быть приведены формулы, аналогичные формуле (1), и рисунки, аналогичные рисункам 1 и 2.

Отметим, что мы согласны с рядом авторов [8–10], что, рассматривая только эти два признака, можно упустить из вида другие закономерности. Поэтому дальнейшие наши исследования будут направлены на поиск признаков из числа ста шести приведенных в начале статьи, добавление которых будет вести к максимальному улучшению метрик решающих правил по каждой нозологии при минимально возможном числе признаков. Соответствующие математические модели и программное обеспечение под данные исследования уже разработаны.

### Список источников

1. Чернавин, П. Ф. Управление качеством решающего правила и минимизация числа признаков в задачах классификации на основе моделей математического программирования / П. Ф. Чернавин, Н. П. Чернавин, Ф. П. Чернавин // Прикаспийский журнал: управление и высокие технологии. – 2023. – № 1. – С. 112–119.
2. Дятченко, Д. Отбор признаков в машинном обучении / Д. Дятченко. – URL: <https://tproger.ru/translations/feature-engineering-in-machine-learning/> (дата обращения: 01.12.2022).
3. Бурнаков, А. Методические заметки об отборе информативных признаков (feature selection) / А. Бурнаков. – URL: <https://habr.com/en/company/aligntechnology/blog/303750/> (дата обращения: 01.12.2022).
4. Дудченко, П. В. Метрики оценки классификаторов в задачах медицинской диагностики / П. В. Дудченко. – Томск : Изд-во ТПУ, 2019. – С. 164–165.
5. Larsson, S. C. Causal role of high body mass index in multiple chronic diseases: a systematic review and meta-analysis of Mendelian randomization studies / S. C. Larsson, S. Burgess // BMC Med. – 2021. – Vol. 19. – Article 320. – <https://doi.org/10.1186/s12916-021-02188-x>.
6. Alwadeai, Khalid. Relationship Between Heart Disease and Obesity Indicators Among Adults: A Secondary Data Analysis / Alwadeai, Khalid & Almeshari, Mohammed & Alghamdi, Abdulrahman & Alshehri, Abdulrahman & Alsaif, Sulaiman & Al-Heizan, Muhammad & Alwadei, Mesfer & Alahmari, Ayedh & Algarni, Saleh & Alotaibi, Tareq & Alqahtani, Mohammed & Alqahtani, Naji & Alqahtani, Jaber & Aldhahir, Abdulelah & Homoud, Mazen & Alhammad, Saad // Cureus. – 2023. – Vol. 15. – 10.7759/cureus.36738.
7. Canoy, D. Body mass index and incident coronary heart disease in women: a population-based prospective study / D. Canoy, B. J. Cairns, A. Balkwill et al. // BMC Med. – 2013. – Vol. 11. – Article 87. – <https://doi.org/10.1186/1741-7015-11-87>.
8. Bell, J. A. Associations of Body Mass and Fat Indexes With Cardiometabolic Traits / J. A. Bell, D. Carslake, L. M. O'Keefe et al. // J. Am. Coll. Cardiol. – 2018. – Vol. 72 (24). – P. 3142–3154. – DOI: 10.1016/j.jacc.2018.09.066.
9. Romero-Corral, A. Accuracy of body mass index in diagnosing obesity in the adult general population / A. Romero-Corral, V. K. Somers, J. Sierra-Johnson // Int. J. Obes. – 2008. – Vol. 32. – P. 959–966.
10. Tan, L. Association of body mass index trajectory and hypertension risk: A systematic review of cohort studies and network meta-analysis of 89,094 participants / L. Tan, L. Z. Long, X. C. Ma et al. // Front. Cardiovasc. Med. – 2023. – Vol. 9. – P. 941341. – DOI: 10.3389/fcvm.2022.941341.
11. Watson, C. Associations of specific-age and decade recall body mass index trajectories with obesity-related cancer / C. Watson, A.G. Renehan & N. Geifman // BMC Cancer. – 2021. – Vol. 21. – Article 502. – <https://doi.org/10.1186/s12885-021-08226-4>.
12. Ruixuan, Dong. Explore the Characteristics of Age, BMI and Blood Composition of Breast Cancer Patients Based on Multivariate Statistical Analysis / Ruixuan Dong // Applied and Computational Mathematics. – August 2020. – Vol. 9, issue 4. – P. 130–145. – DOI: 10.11648/j.acm.20200904.15.

### References

1. Chernavin, P. F., Chernavin, N. P., Chernavin, F. P. Control of the quality of the decision rule and minimization of the number of features in classification problems based on mathematical programming models. *Caspian Journal: Control and High Technologies*, 2023, no. 1, pp. 112–119 (In Russ.).
2. Dyatchenko, D. *Feature selection in machine learning*. URL: <https://tproger.ru/translations/feature-engineering-in-machine-learning/> (accessed 12.01.2022) (In Russ.).
3. Burnakov, A. *Methodological notes on the selection of informative features (feature selection)*. URL: <https://habr.com/en/company/aligntechnology/blog/303750/> (accessed 12.01.2022) (In Russ.).
4. Dudchenko, P. V. *Metrics for evaluating classifiers in medical diagnostic problems*. Tomsk, TPU Publishing House, 2019, pp. 164–165 (In Russ.).

5. Larsson, S. C., Burgess, S. Causal role of high body mass index in multiple chronic diseases: a systematic review and meta-analysis of Mendelian randomization studies. *BMC Med.*, 2021, vol. 19, article 320. <https://doi.org/10.1186/s12916-021-02188-x>.

6. Alwadeai, Khalid & Almeshari, Mohammed & Alghamdi, Abdulrahman & Alshehri, Abdulrahman & Alsaif, Sulaiman & Al-Heizan, Muhammad & Alwadei, Mesfer & Alahmari, Ayedh & Algarni, Saleh & Alotaibi, Tareq & Alqahtani, Mohammed & Alqahtani, Naji & Alqahtani, Jaber & Aldhahir, Abdulelah & Homoud, Mazen & Alhammad, Saad. Relationship Between Heart Disease and Obesity Indicators Among Adults: A Secondary Data Analysis. *Cureus*, 2023, vol. 15, 10.7759/cureus.36738.

7. Canoy, D., Cairns, B. J., Balkwill, A. et al. Body mass index and incident coronary heart disease in women: a population-based prospective study. *BMC Med.*, 2013, vol. 11, article 87. <https://doi.org/10.1186/1741-7015-11-87>.

8. Bell, J. A., Carslake, D., O'Keeffe, L. M., et al. Associations of Body Mass and Fat Indexes With Cardiometabolic Traits. *J. Am. Coll. Cardiol.*, 2018, vol. 72 (24), pp. 3142–3154. DOI: 10.1016/j.jacc.2018.09.066.

9. Romero-Corral, A., Somers, V. K., Sierra-Johnson, J. Accuracy of body mass index in diagnosing obesity in the adult general population. *Int. J. Obes.*, 2008, vol. 32, pp. 959–966.

10. Tan, L., Long, L. Z., Ma, X. C. et al. Association of body mass index trajectory and hypertension risk: A systematic review of cohort studies and network meta-analysis of 89,094 participants. *Front Cardiovasc Med.*, 2023, vol. 9, p. 941341. DOI:10.3389/fcvm.2022.941341.

11. Watson, C., Renehan, A. G. & Geifman, N. Associations of specific-age and decade recall body mass index trajectories with obesity-related cancer. *BMC Cancer*, 2021, vol. 21, article 502. <https://doi.org/10.1186/s12885-021-08226-4>.

12. Ruixuan, Dong. Explore the Characteristics of Age, BMI and Blood Composition of Breast Cancer Patients Based on Multivariate Statistical Analysis. *Applied and Computational Mathematics*, August 2020, vol. 9, issue 4, pp. 130–145. DOI: 10.11648/j.acm.20200904.15.

Статья поступила в редакцию 13.03.2024; одобрена после рецензирования 08.04.2024; принята к публикации 12.04.2024.

The article was submitted 13.03.2024; approved after reviewing 08.04.2024; accepted for publication 12.04.2024.

УДК 538.9, 577.322

### МОРФОЛОГИЧЕСКИЙ АНАЛИЗ СТРУКТУР ВЫСУШЕННЫХ КАПЕЛЬ ВОДНО-СОЛЕВОГО РАСТВОРА СЫВОРОТОЧНОГО АЛЬБУМИНА, ОБРАБОТАННОГО ХОЛОДНОЙ ПЛАЗМОЙ

*Ильясов Фардин Касаинович*, Астраханский государственный университет им. В. Н. Татищева, 414056, Российская Федерация, г. Астрахань, ул. Татищева, 20а,

младший научный сотрудник, ORCID: 0000-0003-0253-7445, e-mail: fardin@inbox.ru

*Аюпова Адилья Камильевна*, Астраханский государственный университет им. В. Н. Татищева, 414056, Россия, г. Астрахань, ул. Татищева, 20а; Астраханский государственный медицинский университет, 414000, Российская Федерация, г. Астрахань, ул. Бакинская, 121,

кандидат медицинских наук, ORCID: 0000-0001-9872-1401, e-mail: adilya\_ayupova@mail.ru

В ходе экспериментов осуществлялась обработка водно-солевого раствора бычьего сывороточного альбумина в концентрации 45 г/л потоком (струей) холодной газоразрядной плазмой с плазмообразующим газом гелием. Установлено, что динамические изменения в структуре фаций данной модельной жидкости наблюдаются в течение двух часов (время эксперимента) после воздействия холодной плазмой. Показано, что через 5 минут после воздействия на раствор, содержащий бычий сывороточный альбумин в концентрации 45 г/л, морфологическая картина его фаций не имеет существенных отличий от контрольных образцов (иррадиальный тип фаций сохраняется), через 1 час наблюдается тенденция к формированию частично-радиального типа фаций, а через два часа определяется сформированный частично-радиальный тип. Количество конкреций увеличивается через час после воздействия, а через два часа – соответствует контрольным показателям. Полученные результаты свидетельствуют о модулирующем эффекте холодной плазмы в отношении некоторых твердофазных параметров (трещин, конкреций) водно-солевого раствора альбумина.

**Ключевые слова:** холодная плазма, гелий, твердофазные структуры, альбумин, высушенная капля, фация, метод клиновидной дегидратации

*Финансирование:* исследование выполнено при поддержке Программы развития Астраханского государственного университета им. В. Н. Татищева (Приоритет – 2030), проект № 1372694632.

### MORPHOLOGICAL ANALYSIS OF THE STRUCTURES OF DRIED DROPLETS OF A WATER-SALT SOLUTION OF SERUM ALBUMIN TREATED WITH COLD PLASMA

*Ilyasov Fardin K.*, Astrakhan Tatishchev State University, 20a Tatischev St., Astrakhan, 414056, Russian Federation,

ORCID: 0000-0003-0253-7445, e-mail: fardin@inbox.ru

*Ayupova Adilya K.*, Astrakhan Tatishchev State University, 20a Tatischev St., Astrakhan, 414056, Russian Federation; Astrakhan State Medical University, 121 Bakinskaya St., Astrakhan, 414000, Russian Federation,

Cand. Sci. (Medicine), ORCID: 0000-0001-9872-1401, e-mail: adilya\_ayupova@mail.ru

During the experiments, a water-salt solution of bovine serum albumin at a concentration of 45 g/l was treated with a flow (jet) of cold gas-discharge plasma with plasma-forming gas helium. It has been established that dynamic changes in the facies structure of this model fluid are observed within two hours (experiment time) after exposure to cold plasma. It has been shown that 5 minutes after exposure to a solution containing bovine serum albumin at a concentration of 45 g/l, the morphological picture of its facies does not differ significantly from the control samples (the irradial type of facies is preserved), after 1 hour there is a tendency to the formation of a partially radial facies type, and after two hours, the formed partially radial type is determined. The number of nodules increases an hour after exposure, and after two hours, it corresponds to control indicators. The results obtained indicate the modulating effect of cold plasma on some solid-phase parameters (cracks, nodules) of a water-salt solution of albumin.

**Keywords:** cold plasma, helium, solid-phase structures, albumin, dried drop, facies, wedge-shaped dehydration method

*Financial support:* this research was supported by the Development Program (Priority – 2030) of Astrakhan Tatishchev State University, project № 1372694632.

## **ВВЕДЕНИЕ**

В настоящее время технологии, основанные на использовании эффектов холодной плазмы (ХП), представляют значительный интерес для современной медицины, биологии, экологии, сельского хозяйства, пищевой промышленности и являются перспективными для решения целого ряда актуальных проблем [1–4].

Для генерации тепловой плазмы необходима высокая температура, достигающая нескольких тысяч градусов Цельсия, а ХП получается при температуре окружающей среды, что делает её пригодной для обработки чувствительных к нагреву материалов [5].

ХП создается путем подачи электрической энергии на один или несколько рабочих газов (кислород (O<sub>2</sub>), азот (N<sub>2</sub>), аргон (Ar), воздух и т. д.), находящихся между двумя электродами под низким или атмосферным давлением. Благодаря возможности ионизации газового потока при атмосферном давлении становится возможным контролировать перенос вещества на нужное расстояние – от нескольких миллиметров до десятков сантиметров – для доставки ионизированных нейтральных частиц материала на необходимую глубину [6–8].

Существует мнение, что даже небольшие изменения условий получения обработанных ХП растворов могут существенным образом сказаться на их свойствах: например, изменение расстояния между факелом и поверхностью жидкости всего с 10 до 15 мм значительно меняет химический состав получаемого раствора [9].

Известно, что эффекты ХП могут существенным образом изменяться в присутствии металлов с переменной валентностью, в частности железа [10].

Температура ХП не превышает 40 °С, т. е. имеет температуру, близкую к таковой биологических объектов. Данное обстоятельство во многом обуславливает возможность применения ХП к живым системам [1]. В настоящее время технологии, основанные на использовании эффектов ХП, представляют огромный интерес для современной медицины, биологии, экологии и являются перспективными для решения целого ряда актуальных проблем. В области медицины активно изучаются возможности применения ХП для лечения онкопатологии, воспалительных процессов, длительно незаживающих ран, стерилизации инструментария и различных поверхностей. Формируется новое научное направление – плазменная хирургия, одной из задач которой является разработка и использование новых инструментов (плазменный скальпель, плазменная игла) для обработки ран и остановки кровотечения во время хирургических вмешательств, ускорения клеточной регенерации, тканевого инжиниринга [11].

Перспективной сферой применения ХП является также природоохранная деятельность. В этом аспекте ХП используется для решения экологических проблем, связанных с разрушением низкоконцентрированных органических и неорганических примесей, содержащихся в промышленных выбросах/отходах, которые негативно влияют на живые системы [3]. Преимуществом низкотемпературной плазменной обработки является быстрый распад продуктов ионизации в газовой смеси, что предотвращает нежелательное загрязнение облучаемого объекта и окружающей среды [12, 13].

Большие возможности для разработки технологий применения ХП открываются и в сельском хозяйстве, поскольку данная отрасль неразрывно связана с живыми системами. Одна из особенностей использования ХП в сельском хозяйстве заключается в том, что обработка биологического материала может проводиться не непосредственно плазмой, а водой, предварительно подвергнутой воздействию ХП [4].

Показано, что воздействие ХП на ткань может быть осуществлено не только непосредственно разрядом, но и растворами, обработанными ХП, состав и рН которых после обработки ХП будут существенно различаться, что определяет новые возможности применения самой ХП и растворов, обработанных ею, в различных областях науки [14].

Значительное количество работ посвящено изучению влияния ХП на микроорганизмы. Установлено, что ХП обладает выраженным антибактериальным эффектом. Она способна инактивировать бактерии, грибы и их споры путем влияния на функции важнейших белков, включая мембранные белки и внутриклеточные белки или активируя апоптоз микробных клеток. Интенсивное или продолжительное облучение ХП может разрушить мембраносвязанные белки микроорганизмов и в дальнейшем способствовать полному разрушению бактерий [15].

Установлено, что ХП инактивировала нуклеотиды ДНК и модифицировала белок в клеточной среде, способствуя инактивации бактерий. ХП также обладает ингибирующей активностью в отношении прионных белков, поверхностных белков бактерий и ферментов, таких как лактатдегидрогеназа, щелочная фосфатаза и пероксидаза [16–21].

Выявлено, что вирусы также проявляет высокую чувствительность к действию ХП. Под влиянием ХП капсидные белки и гликопротеины, входящие в состав мембраны вирусной клетки, могут модифицироваться или даже разрушаться, что в конечном итоге приводит к ее гибели [22].

А. К. Мартусевичем и соавторами методом клиновидной дегидратации проведено изучение кристаллогенных свойств сыворотки крови человека, модифицированных ХП [23, 24]. Полученные результаты свидетельствуют о наличии изменений структурно-оптических свойств исследованной биологической жидкости, происходящих под влиянием ХП.

В связи с этим представляется перспективным изучение влияния ХП на структурные параметры модельных жидкостей, содержащих отдельные компоненты жидких сред организма, для получения дополнительных сведений о механизмах ее действия на живые системы.

Из белков в сыворотке крови больше всего содержится альбумина (около 45 г/л), который играет существенную роль в поддержании коллоидно-осмотического давления в крови и служит для организма важным резервом аминокислот (в его составе 585 аминокислотных остатков). Альбумин обладает способностью связывать липофильные вещества, вследствие чего он является «транспортным средством» для жирных кислот, билирубина, лекарственных веществ, некоторых стероидных гормонов и витаминов. Кроме того, альбумин связывает ионы  $\text{Ca}^{2+}$  и  $\text{Mg}^{2+}$ .

Альбумин относится к простым белкам. Его молекулярная масса составляет 66 кДа. Наличие у альбумина 17 дисульфидных мостиков обеспечивает его способность образовывать димеры и полимеры. По сравнению с глобулинами альбумин обладает более высокими осмотической активностью и степенью гидрофильности.

Учитывая высокое функциональное значение альбумина в жизнедеятельности живого организма и информативность метода клиновидной дегидратации в оценке его возможных конформаций, представляется перспективным изучение влияния ХП на твердофазные характеристики растворов альбумина.

Анализ твердофазных структур, сформированных в условиях клиновидной дегидратации исследуемой жидкости, позволяет оценивать изменения ее молекулярных структур, происходящих при внешних воздействиях любых физических факторов [25]. Известно, что конформационные изменения макромолекулы белка находят свое отражение в количестве и характере расположений трещин фаций его раствора [26]. В этой связи представляется перспективным изучение влияния ХП на структурные параметры модельных растворов, содержащих отдельные компоненты жидких сред организма, для получения дополнительных сведений о механизмах ее действия на живые системы.

Цель исследования – изучить закономерности процесса структурирования водно-солевого раствора альбумина под воздействием холодной плазмы.

#### **МАТЕРИАЛЫ И МЕТОДЫ**

Модельные жидкости готовили, растворяя лиофилизированный бычий сывороточный альбумин (БСА) в 0,9% растворе хлорида натрия. Концентрация альбумина составляла 45 г/л.

Получение гелиевой плазмы осуществлялось с помощью экспериментальной установки, состоящей из двух основных частей – камеры, где происходит ионизация газа, и системы подачи газа. Работа установки основана на принципе электрического разряда в газе. При подаче напряжения на электроды между ними возникает разряд, который ионизирует газ, превращая его в плазму.

В ходе эксперимента исследуемую жидкость помещали в микропробирку и подвергали воздействию ХП в течение 3 минут. Влияние ХП на изучаемый раствор оценивали с помощью метода клиновидной дегидратации [25]. Обработанный потоком гелиевой ХП водно-солевой раствор альбумина в форме капли объемом 0,01 мл наносили на предметное стекло через 5 минут, один и два часа после воздействия. Полученные тест-капли (опытные образцы) высушивали при температуре  $+20 \dots +25$  °С, относительной влажности 60 % и минимальной подвижности воздуха до образования тонкой пленки – фации. Контролем являлись высушенные капли водно-солевого раствора альбумина, не подвергшегося воздействию ХП. Морфологическую картину фаций изучали с помощью стереомикроскопа BS3060. Определяли количество, направленность и соотношение радиальных и иррадиальных трещин белковой матрицы в исследуемых образцах, а также конфигурации конкреций, представляющих собой скопление однородного вещества в фации. Существует мнение, что степень сформированности конкреций обусловлена энергетической активностью белковых молекул. Различают завершенные, частично завершенные, незавершенные и несформированные конкреции [25].

Полученные результаты были статистически обработаны с применением t-критерия Стьюдента. Использовали пакеты программ Microsoft Excel. Статистически значимыми различия считали при  $p \leq 0,05$ .

#### **РЕЗУЛЬТАТЫ ИССЛЕДОВАНИЯ**

Проведен сравнительный анализ морфологической картины контрольного и опытных образцов высушенных капель водно-солевого раствора альбумина. Контрольные фации характеризовались наличием крупных трещин, имеющих как радиальную, так и иррадиальную направленность. Количество иррадиальных трещин было в полтора раза больше, чем радиальных (рис. 1А). Из других структурных компонентов наблюдались конкреции различного вида: завершенные, незавершенные и несформированные. Общее количество данных структурных элементов в контрольных образцах составило  $8,6 \pm 0,8$  единиц. При этом преобладали конкреции незавершенной формы (рис. 2А).

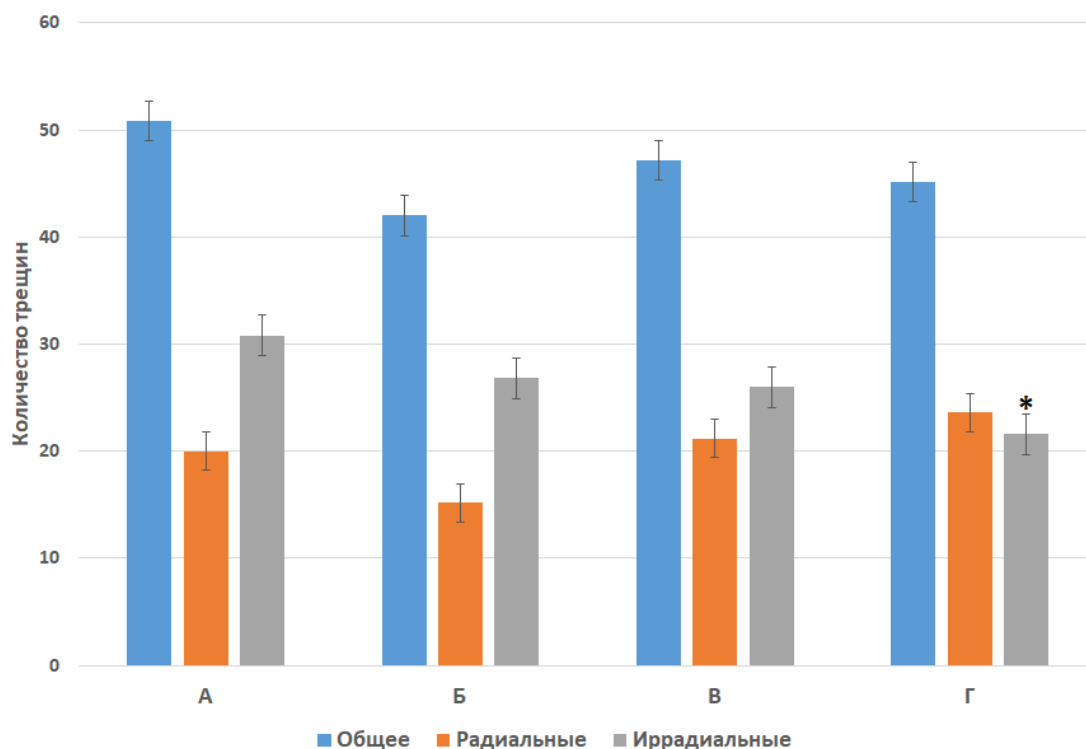


Рисунок 1 – Количество трещин в периферической зоне фаций 4,5%-го водно-солевого раствора БСА: А – в фациях интактного раствора; Б–Г – в фациях, приготовленных из обработанного ХП раствора через 5 минут, 1 час, 2 часа после воздействия соответственно (\* – различия по сравнению с контрольным образцом статистически значимы,  $p < 0,05$ )

В структурной организации капель водно-солевого раствора альбумина, обработанного ХП, помещенных на предметное стекло через 5 минут после воздействия, общее количество крупных трещин достоверно не изменялось по сравнению с контрольными образцами (рис. 1Б). Фации имели иррадиальный тип: соотношение радиальных/иррадиальных трещин соответствовало 0,6. Количество конкреций было сопоставимо с таковыми интактного раствора, их морфологическое разнообразие также не претерпело каких-либо изменений (рис. 2Б).

Через час после воздействия общее число крупных трещин было аналогично таковому в контроле. Наблюдалась тенденция к формированию частично радиального типа фаций: соотношение радиальных/иррадиальных трещин менялось в сторону увеличения радиальных и составляло 0,8 (рис. 1В). Общее количество конкреций, а также число их завершенных и незавершенных форм (рис. 2В), достоверно отличалось от таковых контрольных образцов ( $p < 0,05$ ).

В фациях, приготовленных через 2 часа после воздействия ХП на водно-солевые растворы альбумина, общее количество трещин соответствовало таковому контрольных образцов. Количество иррадиальных трещин статистически достоверно уменьшалось по сравнению с контролем ( $p < 0,05$ ). Наблюдался частично радиальный тип фаций: соотношение радиальных/иррадиальных трещин составило 1,07 (рис. 1Г). Количество наблюдаемых конкреций снижалось до контрольных показателей (рис. 2Г).

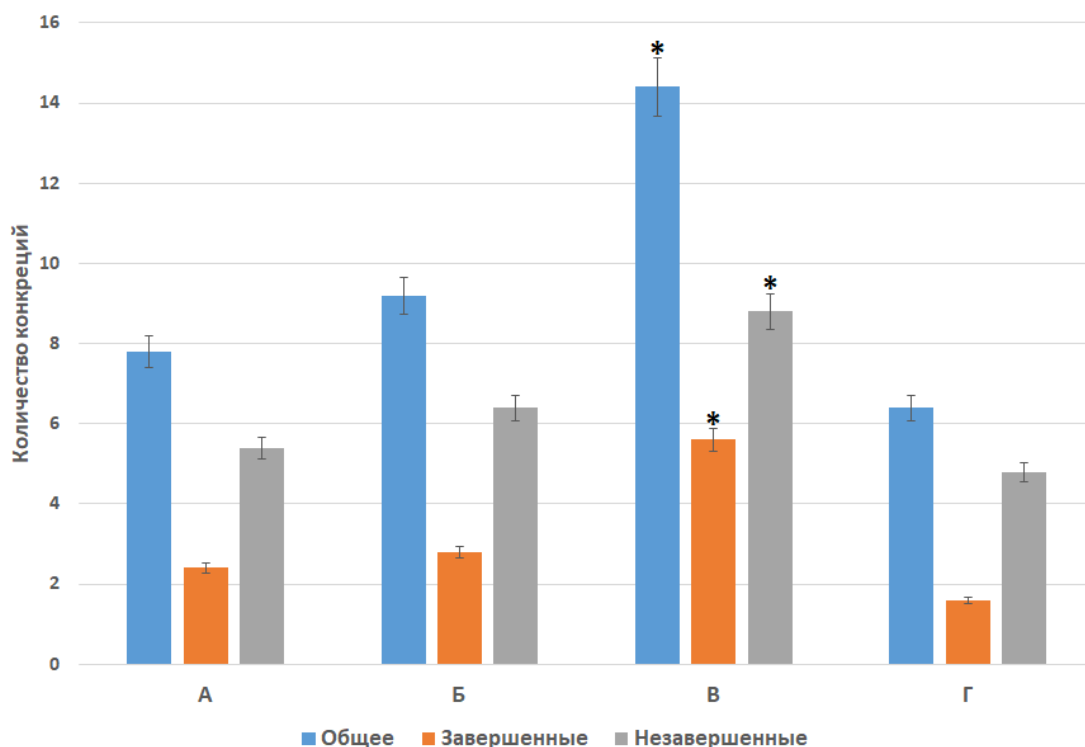


Рисунок 2 – Количество конкреций в периферической зоне фаций 4,5%-го водно-солевого раствора БСА: А – в фациях интактного раствора; Б–Г – в фациях, приготовленных из обработанного ХП раствора через 5 минут, 1 час, 2 часа после воздействия соответственно (\* – различия по сравнению с контрольным образцом статистически значимы,  $p < 0,05$ )

## ЗАКЛЮЧЕНИЕ

Установлено, что динамические изменения в структуре фаций данной модельной жидкости наблюдаются в течение двух часов (время эксперимента) после воздействия ХП. Через 5 минут после воздействия на 4,5%-й водно-солевой раствор БСА морфологическая картина его фаций не имеет существенных отличий от контрольных образцов (иррадиальный тип фаций сохраняется), через 1 час наблюдается тенденция к формированию частично-радиального типа фаций, а через два часа определяется сформированный частично-радиальный тип. Количество конкреций увеличивается через час после воздействия, а через два часа – соответствует контрольным показателям.

Полученные результаты свидетельствуют о модулирующем эффекте ХП в отношении некоторых твердофазных параметров (трещин, конкреций) водно-солевого раствора альбумина.

## Список литературы

1. Герасименко, М. Ю. Низкотемпературная плазма – перспективный метод реабилитации / М. Ю. Герасименко, Т. Н. Зайцева, И. С. Евстигнеева // Физическая и реабилитационная медицина, медицинская реабилитация. – 2019. – № 3. – С. 79–89.
2. Yong, Xu. Effects of non-thermal atmospheric plasma on protein / Yong Xu, Yu Bai, Chenwei Dai, Han Lv, Xiuhong Zhou, Qinghua Xu // Journal of Clinical Biochemistry and Nutrition. – 2022. – Vol. 71, № 3. – P. 173–184.
3. Акишев, Ю. С. Низкотемпературная плазма при атмосферном давлении и ее возможности для приложений / Ю. С. Акишев // Известия высших учебных заведений. Химия и химическая технология. – 2019. – Т. 62, вып. 8. – С. 26–60.
4. Бурмистров, Д. Е. Получение и применение воды, обработанной холодной атмосферной плазмой / Д. Е. Бурмистров, Е. М. Коченков, К. В. Артемьев, В. А. Козлов, Л. В. Колик, С. В. Гудков // Физика водных растворов : сборник трудов конференции. – 2021. – С. 71.
5. López, M. A Review on Non-thermal Atmospheric Plasma for Food Preservation: Mode of Action, Determinants of Effectiveness, and Applications / M. López, T. Calvo, M. Prieto, R. Múgica-Vidal, I. Muro-Fraguas, F. Alba-Ellías, A. Alvarez-Ordóñez // Frontiers in Microbiology. – 2019. – Vol. 10. – P. 622.
6. Harry, J. E. Introduction to Plasma Technology-Science, Engineering and Applications / J. E. Harry. – Weinheim : Wiley-VCH, 2010. – P. 123–153.
7. Mir, S. A. Understanding the Role of Plasma Technology in Food Industry / S. A. Mir, M. A. Shah, M. M. Mir // Food and Bioprocess Technology. – 2016. – Vol. 9. – P. 734–750.
8. Muhammad, A. I. Understanding the Impact of Nonthermal Plasma on Food Constituents and Microstructure – A Review / A. I. Muhammad, Q. Xiang, X. Liao // Food and Bioprocess Technology. – 2018. – Vol. 11. – P. 463–486.

9. Van, Boxem W. Anti-cancer capacity of plasma-treated PBS: effect of chemical composition on cancer cell cytotoxicity / W. Van Boxem, J. Van der Paal, Y. Gorbanev, S. Vanuytsel, E. Smits., S. Dewilde, A. Bogaerts // *Scientific Reports*. – 2017. – Vol. 7 (1). – P. 16478.
10. Lu, X. Transcutaneous plasma stress: From soft-matter models to living tissues / X. Lu, M. Keidar, M. Laroussi, E. Choi, E. J. Szili, K. Ostrikov // *Materials Science and Engineering : reports*. – 2019. – Vol. 138. – P. 36–59.
11. Алейник, А. Н. Использование холодной плазмы для ускорения регенерации кожных покровов после механических повреждений / А. Н. Алейник, В. В. Великая, Тургунова Н. Д., Старцева Ж. А. // *Медицинская техника*. – 2020. – № 4. – с. 29–30.
12. Меньшикова, Е. Б. Окислительный стресс – прооксиданты и антиоксиданты / Е. Б. Меньшикова, В. З. Ланкин, Н. К. Зенков, И. А. Бондарь, Н. Ф. Круговых, В. А. Труфакин. – Москва : Слово, 2006. – 556 с.
13. Riabyi, V. A. A Long-Lifetime Plasmatron for Use in Medicine / V. A. Riabyi, V. P. Savinov, V. G. Yakunin // *Biomedical Engineering*. – 2017. – Vol. 51. – P. 195–199.
14. Оловяникова, Р. Я. Химические механизмы действия холодной плазмы на клетки / Р. Я. Оловяникова, Т. А. Макаренко, Е. В. Лычковская, Е. С. Гудкова, Г. А. Мурадян, Н. Н. Медведева, Т. Н. Чекишева, С. И. Бердников, Е. В. Семичев, Н. А. Малиновская, А. Б. Салмина, В. В. Салмин // *Фундаментальная и клиническая медицина*. – 2020. – № 5 (4). – С. 104–115.
15. Hou, Y. M. Disintegration of biomacromolecules by dielectric barrier discharge plasma in helium at atmospheric pressure / Y. M. Hou, X. Y. Dong, H. Yu // *IEEE Transactions on Plasma Science*. – 2008. – Vol. 36. – P. 1633–1637.
16. Lackmann, J. W. Photons and particles emitted from cold atmospheric-pressure plasma inactivate bacteria and biomolecules independently and synergistically / J. W. Lackmann, S. Schneider, E. Edengeiser // *Journal of the Royal Society Interface*. – 2013. – Vol. 10. – P. 20130591.
17. Elmoualij, B. Decontamination of prions by the flowing afterglow of a reduced-pressure N<sub>2</sub>-O<sub>2</sub> cold-plasma / B. Elmoualij, O. Thellin, S. Gofflot, E. Heinen, P. Levif, J. Séguin // *Plasma Processes and Polymers*. – 2012. – Vol. 9 (6). – P. 612–618.
18. Digel, I. Bactericidal effects of plasma-generated cluster ions / I. Digel, A. T. Artmann, K. Nishikawa, M. Cook, E. Kurulgan, G. M. Artmann // *Medical & Biological Engineering & Computing*. – 2005. – Vol. 43 (6). – P. 800–807.
19. Zhang, H. Effects and mechanism of atmospheric-pressure dielectric barrier discharge cold plasma on lactate dehydrogenase (LDH) enzyme / H. Zhang, Z. Xu, J. Shen, X. Li, L. Ding, J. Ma // *Scientific Reports*. – 2015. – Vol. 5. – P. 10031.
20. Segat, A. Effect of atmospheric pressure cold plasma (ACP) on activity and structure of alkaline phosphatase / A. Segat, N. N. Misra, P. J. Cullen, N. Innocente // *Food and Bioproducts Processing*. – 2016. – Vol. 98. – P. 181–188.
21. Surowsky, B. Cold plasma effects on enzyme activity in a model food system / B. Surowsky, A. Fischer, O. Schlueter, D. Knorr // *Innovative Food Science & Emerging Technologies*. – 2013. – Vol. 19. – P. 146–152.
22. Aboubakr, H. A. Cold argon-oxygen plasma species oxidize and disintegrate capsid protein of feline calicivirus / H. A. Aboubakr, S. K. Mor, L. Higgins // *PLoS One*. – 2018. – Vol. 13 (3). – P. 0194618.
23. Мартусевич, А. К. Влияние холодной гелиевой плазмы на метаболические и физико-химические параметры крови человека *in vitro* / А. К. Мартусевич, А. Г. Соловьева, С. Ю. Краснова, Д. В. Янин, А. Г. Галка, А. В. Костров // *Биомедицина*. – 2018. – № 2. – С. 47–58.
24. Мартусевич, А. К. Модификация физико-химических свойств крови при действии гелиевой холодной плазмы / А. К. Мартусевич, А. Г. Соловьева, С. Ю. Краснова, А. Г. Галка, Д. В. Янин // *Биорадикалы и антиоксиданты*. – 2018. – № 3. – С. 45–48.
25. Шатохина, С. Н. Атлас структур неклочных тканей человека в норме и патологии : в 3 т. Том II. Морфологические структуры сыворотки крови / С. Н. Шатохина, В. Н. Шабалин. – Москва – Тверь : ООО «Триада», 2013. – 240 с.
26. Белова, Л. М. Исследование конформационных изменений молекулы альбумина в различных условиях методом клиновидной дегидратации / Л. М. Белова, Ю. П. Потехина // *Нижегородский медицинский журнал*. – 2003. – № 3–4. – С. 89–90.

#### References

1. Gerasimenko, M. Yu., Zaitseva, T. N., Evstigneeva, I. S. Low-temperature plasma – a promising method of rehabilitation. *Physical and Rehabilitation Medicine, Medical Rehabilitation*, 2019, no. 3, pp. 79–89 (In Russ.).
2. Yong, Xu, Yu, Bai, Chenwei, Dai, Han, Lv, Xiuhong, Zhou, Qinghua, Xu. Effects of non-thermal atmospheric plasma on protein. *Journal of Clinical Biochemistry and Nutrition*, 2022, Vol. 71, no. 3, pp. 173–184.
3. Akishev, Yu. S. Non-thermal plasma at atmospheric pressure and its opportunities for applications. *ChemChemTech*, 2019, Vol. 62, no. 8, pp. 26–60 (In Russ.).
4. Burmistrov, D. E., Kochenkov, E. M., Artemyev, K.V., Kozlov, V. A., Kolik, L. V., Gudkov, S. V. Production and use of water treated with cold atmospheric plasma. *Physics of Aqueous Solutions : collection of proceedings of the conference*, 2021, p. 71 (In Russ.).
5. López, M., Calvo, T., Prieto, M., Múgica-Vidal, R., Muro-Fraguas, I., Alba-Elías, F., Alvarez-Ordóñez, A. A Review on Non-thermal Atmospheric Plasma for Food Preservation: Mode of Action, Determinants of Effectiveness, and Applications. *Frontiers in Microbiology*, 2019, vol. 10, p. 622.
6. Harry, J. E. *Introduction to Plasma Technology-Science, Engineering and Applications*. Weinheim: Wiley-VCH, 2010, pp. 123–153.
7. Mir, S. A., Shah, M. A., Mir, M. M. Understanding the Role of Plasma Technology in Food Industry. *Food and Bioprocess Technology*, 2016, vol. 9, pp. 734–750.

8. Muhammad, A. I., Xiang, Q., Liao, X. Understanding the Impact of Nonthermal Plasma on Food Constituents and Microstructure – A Review. *Food and Bioprocess Technology*, 2018, vol. 11, pp. 463–486.
9. Van Boxem, W., Van, der Paal J., Gorbanev, Y., Vanuysel, S., Smits, E., Dewilde, S., Bogaerts, A. Anti-cancer capacity of plasma-treated PBS: effect of chemical composition on cancer cell cytotoxicity. *Scientific Reports*, 2017, vol. 7 (1), p. 16478.
10. Lu, X., Keidar, M., Laroussi, M., Choi, E., Szili, E. J., Ostrikov, K. Transcutaneous plasma stress: From soft-matter models to living tissues. *Materials Science and Engineering : reports*, 2019, vol. 138, pp. 36–59.
11. Aleynik, A. N., Velikaya, V. V., Turgunova, N. D., Startseva, Zh. A. The use of cold plasma to accelerate the regeneration of the skin after mechanical damage. *Medical Technology*, 2020, no. 4, pp. 29–30 (In Russ.).
12. Menshchikova, E. B., Lankin, V. Z., Zenkov, N. K., Bondar, I. A., Krugovykh, N. F., Trufakin, V. A. Oxidative stress – pro-oxidants and antioxidants. Moscow, Slovo Publ., 2006. 556 p. (In Russ.).
13. Riabiy, V. A., Savinov, V. P., Yakunin, V. G. A Long-Lifetime Plasmatron for Use in Medicine. *Biomedical Engineering*, 2017, vol. 51, pp. 195–199.
14. Olovyannikova, R. Ya., Makarenko, T. A., Lychkovskaya, E. V., Gudkova, E. S., Muradyan, G. A., Medvedeva, N. N., Chekischeva, T. N., Berdnikov, S. I., Semichev, E. V., Malinovskaya, N. A., Salmina, A. B., Salmin, V. V. Chemical mechanisms of cold plasma action on cells. *Fundamental and Clinical Medicine*, 2020, no. 5 (4), pp. 104–115 (In Russ.).
15. Hou, Y. M., Dong, X. Y., Yu, H. Disintegration of biomacromolecules by dielectric barrier discharge plasma in helium at atmospheric pressure. *IEEE Transactions on Plasma Science*, 2008, vol. 36, pp. 1633–1637.
16. Lackmann, J. W., Schneider, S., Edengeiser, E. Photons and particles emitted from cold atmospheric-pressure plasma inactivate bacteria and biomolecules independently and synergistically. *Journal of the Royal Society Interface*, 2013, vol. 10, p. 20130591.
17. Elmoulaj, B., Thellin, O., Gofflot, S., Heinen, E., Levif, P., Séguin, J. Decontamination of prions by the flowing afterglow of a reduced-pressure N<sub>2</sub>-O<sub>2</sub> cold-plasma. *Plasma Processes and Polymers*, 2012, vol. 9 (6), pp. 612–618.
18. Digel, I., Artmann, A. T., Nishikawa, K., Cook, M., Kurulgan, E., Artmann, G. M. Bactericidal effects of plasma-generated cluster ions. *Medical & Biological Engineering & Computing*, 2005, vol. 43 (6), pp. 800–807.
19. Zhang, H., Xu, Z., Shen, J., Li, X., Ding, L., Ma, J. Effects and mechanism of atmospheric-pressure dielectric barrier discharge cold plasma on lactate dehydrogenase (LDH) enzyme. *Scientific Reports*, 2015, vol. 5, p. 10031.
20. Segat, A., Misra, N. N., Cullen, P. J., Innocente, N. Effect of atmospheric pressure cold plasma (ACP) on activity and structure of alkaline phosphatase. *Food and Bioproducts Processing*, 2016, vol. 98, pp. 181–188.
21. Surowsky, B., Fischer, A., Schlueter, O., Knorr, D. Cold plasma effects on enzyme activity in a model food system. *Innovative Food Science & Emerging Technologies*, 2013, vol. 19, pp. 146–152.
22. Aboubakr, H. A., Mor, S. K., Higgins, L. Cold argon-oxygen plasma species oxidize and disintegrate capsid protein of feline calicivirus. *PLoS One*, 2018, vol. 13 (3), p. 0194618.
23. Martusevich, A. K., Solovyova, A. G., Krasnova, S. Yu., Yanin, D. V., Galka, A. G., Kostrov, A. V. The influence of cold helium plasma on the metabolic and physicochemical parameters of human blood in vitro. *Biomedicine*, 2018, no. 2, pp. 47–58 (In Russ.).
24. Martusevich, A. K., Solovyova, A. G., Krasnova, S. Yu., Galka, A. G., Yanin, D. V. Modification of physicochemical properties of blood under the action of cold helium plasma. *Bioradicals and antioxidants*, 2018, no. 3, pp. 45–48 (In Russ.).
25. Shatokhina, S. N., Shabalin, V. N. *Atlas of structures of non-cellular human tissues in normal and pathological conditions : in 3 vol. Volume II. Morphological structures of blood serum*. Moscow – Tver, Triada LLC, 2013. 240 p. (In Russ.).
26. Belova, L. M., Potekhina, Yu. P. Study of conformational changes in the albumin molecule under various conditions using the wedge-shaped dehydration method. *Nizhny Novgorod Medical Journal*, 2003, no. 3–4, pp. 89–90. (In Russ.).

Статья поступила в редакцию 11.07.2024; одобрена после рецензирования 06.09.2024; принята к публикации 09.09.2024.

The article was submitted 11.07.2024; approved after reviewing 06.09.2024; accepted for publication 09.09.2024.

# **ПРИБОРОСТРОЕНИЕ, МЕТРОЛОГИЯ И ИНФОРМАЦИОННО-ИЗМЕРИТЕЛЬНЫЕ ПРИБОРЫ И СИСТЕМЫ**

## **ПРИБОРЫ И МЕТОДЫ ИЗМЕРЕНИЯ (ПО ВИДАМ ИЗМЕРЕНИЙ)**

УДК 004.004

### **РАЗРАБОТКА И ПРОГРАММНО-АППАРАТНАЯ РЕАЛИЗАЦИЯ СДВОЕННОГО 14-РАЗРЯДНОГО АЦП**

*Брысин Андрей Николаевич*, РТУ МИРЭА, 119454, Российская Федерация, г. Москва, пр-т Вернадского, 78; Институт машиноведения им. А.А. Благонравова Российской академии наук, 101000, Российская Федерация, г. Москва, Малый Харитоньевский переулок, 4,

кандидат технических наук, доцент, ORCID: 0000-0001-8870-5037, e-mail: brysin@mirea.ru

*Арсентьев Даниил Романович*, РТУ МИРЭА, 119454, Российская Федерация, г. Москва, пр-т Вернадского, 78,

студент, ORCID: 0009-0002-9972-1993, e-mail: danrom11pro@gmail.com

*Новикова Анастасия Сергеевна*, РТУ МИРЭА, 119454, Российская Федерация, г. Москва, пр-т Вернадского, 78,

ассистент, ORCID: 0009-0002-7451-0743, e-mail: 79163864045@ya.ru

*Журавлева Юлия Алексеевна*, РТУ МИРЭА, 119454, Российская Федерация, г. Москва, пр-т Вернадского, 78; «Национальный исследовательский университет «МЭИ», 111250, Российская Федерация, г. Москва, вн. тер. г. муниципальный округ Лефортово, ул. Красноказарменная, 14, стр. 1,

кандидат технических наук, доцент, ORCID: 0000-0003-3919-5127, e-mail: ulypil@mail.ru

*Железникова Ольга Евгеньевна*, Мордовский государственный университет им. Н. П. Огарёва, 430005, Российская Федерация, г. Саранск, ул. Большевикская, 68/1,

кандидат технических наук, директор Института электроники и светотехники, ORCID: 0000-0003-3919-5127, e-mail: sarstf@mail.ru

*Коваленко Ольга Юрьевна*, Мордовский государственный университет им. Н. П. Огарёва, 430005, Российская Федерация, г. Саранск, ул. Большевикская, 68/1,

доктор технических наук, доцент, ORCID: 0000-0001-7870-5949, e-mail: crystall2000@mail.ru

В данной работе рассматривается разработка и оптимизация схемы работы сдвоенного 14-разрядного аналого-цифрового преобразователя, предназначенного для приложений, требующих высокую точность. Актуальность данного исследования обусловлена необходимостью обеспечения высокой точности и надежности аналого-цифрового преобразования цифровых сигналов и данных в разнообразных системах. В предложенной схеме работы сдвоенного аналого-цифрового преобразователя уделяется особое внимание адаптации к экстремальным условиям, повышенной стойкости к помехам и обеспечению высокой разрядности для точного измерения и анализа параметров. Разработанные технологии и решения могут быть применены в области разработки и совершенствования, радиосвязи, системы навигации и контроля, обеспечивая высокий уровень эффективности.

**Ключевые слова:** аналого-цифровой преобразователь, схема, точность, надежность, помехозащищенность, разрядность

### **DEVELOPMENT AND SOFTWARE-HARDWARE REALIZATION OF A DOUBLE 14-bit ADC**

*Brysin Andrey N.*, RTU MIREA, 78 Vernadsky Avenue, Moscow, 119454, Russian Federation; Institute of Machines Science named after A. A. Blagonravov of the Russian Academy of Sciences,

Cand. Sci. (Engineering), Associate Professor, ORCID: 0000-0001-8870-5037, e-mail: brysin@mirea.ru

*Arsentiev Daniil R.*, RTU MIREA, 78 Vernadsky Avenue, Moscow, 119454, Russian Federation, student, ORCID: 0009-0002-9972-1993, e-mail: danrom11pro@gmail.com

*Novikova Anastasia S.*, RTU MIREA, 78 Vernadsky Avenue, Moscow, 119454, Russian Federation Assistant, ORCID: 0009-0002-7451-0743, e-mail: 79163864045@ya.ru

*Zhuravleva Yulia A.*, RTU MIREA, 78 Vernadsky Avenue, Moscow, 119454, Russian Federation; National Research University «Moscow Power Engineering Institute,

Cand. Sci. (Engineering), Associate Professor, ORCID: 0000-0003-3919-5127, e-mail: ulypil@mail.ru

*Zheleznikova Olga E.*, Ogarev Mordovia State University, 68/1 Bolshevistskaya St., Saransk, 430005, Russian Federation,

Cand. Sci. (Engineering), Director of the Institute of Electronics and Lighting Engineering, ORCID: 0000-0003-3919-5127, e-mail: sarstf@mail.ru

**Kovalenko Olga Yu.**, Ogarev Mordovia State University, 68/1 Bolshhevistskaya St., Saransk, 430005, Russian Federation,

Doct. Sci. (Engineering), Associate Professor, ORCID: 0000-0003-3919-5127, e-mail: crystall2000@mail.ru

This paper discusses the design and optimization of a dual 14-bit analog-to-digital converter designed for applications requiring high accuracy. The relevance of this research is due to the need to ensure high accuracy and reliability of analog-to-digital conversion of digital signals and data in a variety of systems. The proposed dual analog-to-digital converter design places special emphasis on adapting to extreme conditions, improving noise immunity, and providing high bit resolution for accurate measurement and analysis. The developed technologies and solutions can be applied in the field of development and improvement, radio communications, navigation and control systems, providing a high level of efficiency.

**Keywords:** analog-to-digital converter, circuit, accuracy, reliability, noise immunity, bit depth

**Graphical annotation (Графическая аннотация)**



## ВВЕДЕНИЕ

В настоящее время используются различные программы с защитой личной данных. Аналого-цифровой преобразователь (АЦП) имеет критическое значение в устройствах, где точность и скорость обработки данных играют решающую роль, в том числе и для защиты информации [1–5]. После конвертации сигнала в цифровой формат данные могут быть переданы в программируемую логическую интегральную схему (ПЛИС) для дальнейшей обработки. В ПЛИС может быть реализован алгоритм, который анализирует цифровые данные, определяет амплитуду сигнала, вычисляет разность фаз и строит пеленг, позволяя точно определить местоположение цели. Этот процесс позволяет ГСН быстро и точно анализировать сигналы от целей и реагировать на изменения в реальном времени. В ходе работы было проведено моделирование и разработка схемы двояного 14-разрядного аналого-цифрового преобразователя высокой точности с целью:

- повышения точности и разрядности аналогово-цифрового преобразования для более точного измерения аналоговых сигналов;
- исследования возможности применения разрабатываемой схемы в приложениях, требующих высокой точности измерений, включая военные, медицинские и промышленные области;
- разработки алгоритмов обработки данных для повышения надежности и точности измерений.

## РАЗРАБОТКА АЦП

В основе устройства лежит двояный 14-разрядный АЦП со скоростью 3 Гбит/с и интерфейсом JESD. Синхронизацию JESD обеспечивают стабилизаторы напряжения LMK и TI. Вход для каждого канала АЦП по умолчанию подключен к входной цепи трансформатора и может быть подключен к несимметричному источнику сигнала 50 Ом. Опорный вход тактового сигнала осуществляется через трансформаторный вход и также может быть подключен к несимметричному источнику тактового сигнала 50 Ом. Модуль имеет возможность генерации необходимых JESD часов с помощью бортового LMK.

Конфигурация модуля осуществляется через встроенное USB-соединение, для которого была написана библиотека для быстрой прошивки регистров LMX LMK ADC. Стандартное расположение контактов JESD на разъеме FMC позволяет прямое подключение к карте захвата TSW и другим коммерчески доступным платформам разработки FPGA (рис. 1).

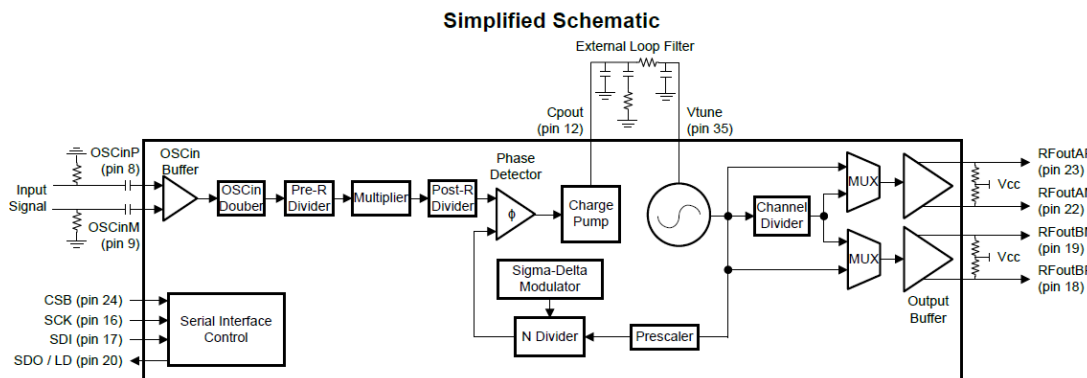


Рисунок 1 – Упрощенная схема двояного 14-разрядного АЦП

Подробно рассмотрим каждый блок устройства АЦП и его функциональность:

**OSCin (Осциллятор входной):** представляет собой входной осциллятор, который может использоваться для предоставления тактового сигнала устройству. Генерирует стабильный сигнал с заданной частотой.

**Doubler (Удвоитель частоты):** удваивает частоту сигнала, который поступает на его вход. Принимает сигнал от OSCin и создает выходной сигнал, частота которого вдвое выше.

**Post-R (Пост-делитель):** производит деление частоты сигнала после удвоителя. Разделяет входной сигнал на более низкочастотный сигнал.

**Divider (Делитель):** осуществляет деление частоты сигнала. Делит входной сигнал на частоту, соответствующую заданному коэффициенту деления.

**Multiplier (Умножитель):** умножает входной сигнал на заданный коэффициент.

**Pre-R (Пред-делитель):** производит предварительное деление частоты входного сигнала. Разделяет входной сигнал на более низкочастотный сигнал перед его дальнейшей обработкой.

**Divider (Делитель второй):** делит предварительно разделенный сигнал на частоту, соответствующую заданному коэффициенту деления.

**I (Текущий):** относится к аналоговой части устройства, связанной с измерением и управлением током. Осуществляет операции, связанные с текущими значениями величин.

**Prescaler (Предварительный делитель):** производит предварительное деление сигнала. Разделяет входной сигнал на более низкочастотный сигнал перед дальнейшей обработкой.

**Channel (Канал):** связан с множеством каналов для аналогово-цифрового преобразования. Мультиплексирует и управляет разными входными каналами.

**MUX Divider (Делитель мультиплексора):** связан с делением сигнала, используемого мультиплексором. Делит сигнал перед его использованием мультиплексором.

**MUX (Мультиплексор):** выбирает один из нескольких входных сигналов для дальнейшей обработки. Мультиплексирует различные входы.

**OSC\_2X PLL\_R\_PRE (Умножитель частоты PLL):** используется для увеличения частоты внутреннего тактового сигнала. Умножает входной сигнал на заданный коэффициент.

**CP\_ICOARSE, CP\_IUP, CP\_IDN (Управление фазой и усилением):** связаны с управлением фазой и усилением сигнала внутри устройства. Осуществляют управление усилением и фазой.

**REF\_EN (Включение опорного напряжения):** контролирует включение опорного напряжения для аналогово-цифрового преобразования. Включает или отключает опорное напряжение.

**PLL\_N (Число делений PLL):** управляет числом делений внутреннего PLL (фазовая автоподстройка) для генерации стабильного тактового сигнала. Устанавливает необходимое число делений для PLL.

**PFD\_DLY (Задержка фазового детектора):** регулирует задержку фазового детектора.

**MASH\_ORDER (Порядок MASH-фильтра):** относится к фильтрации и обработке данных. Устанавливает порядок MASH-фильтра.

**PLL\_N\_PRE (Предварительное умножение PLL):** управляет предварительным умножением внутреннего PLL. Осуществляет предварительное умножение для PLL.

**OUTx\_PD (Отключение выходов):** служит для отключения определенных выходов устройства. Отключает или включает определенные выходы.

**OUTx\_POW (Управление мощностью выходов):** регулирует мощность сигнала на выходе. Управляет мощностью выходных сигналов.

OUTx\_MUX (Мультиплексор выходов): служит для выбора определенного выходного сигнала. Мультиплексирует различные выходы.

CHDIV\_SEG1, CHDIV\_SEG2, CHDIV\_SEG3 (Делители сегментов канала) относится к управлению делителями для разных сегментов канала.

CHDIV\_SEG\_SEL (Выбор сегмента канала): определяет, какой сегмент канала будет использоваться. Выбирает сегмент канала для обработки.

Благодаря всем этим встроенным механизмам и технологиям, АЦП обеспечивает высокую устойчивость к помехам и шумам, обеспечивая точные и надежные результаты измерений даже в сложных условиях.

Принцип работы: устройство принимает входной аналоговый сигнал, который требуется измерить и преобразовать в цифровой формат. Устройство имеет встроенные усилители и фильтры, которые могут усилить или фильтровать входной сигнал для лучшей обработки. Аналого-цифровое преобразование выполняется путем сравнения аналогового сигнала с опорным напряжением и последующего дискретизации аналогового сигнала. Этот процесс включает в себя следующие шаги:

- сравнение: аналоговый сигнал сравнивается с опорным напряжением;
- дискретизация: сигнал аналогового сигнала в моменты времени и запись значения амплитуды на каждом из этих моментов. Эти значения представляют собой цифровые отсчеты, которые образуют выходное цифровое слово;

- разрядность: устройство имеет разрядность 14 бит, что означает, что каждый цифровой отсчет представляется 14-ю битами данных. Это позволяет устройству обеспечивать более высокую точность при преобразовании аналоговых сигналов;

- процессирование данных: цифровые отсчеты могут подвергаться дополнительной обработке, если это необходимо, например, фильтрации или усилению. Это может выполняться встроенными цифровыми блоками или с использованием дополнительных компонентов;

- выходные данные: преобразованные цифровые данные могут быть выведены через интерфейс JESD204, он достаточно эффективен и обеспечивает преимущества в скорости, занимаемой площади и стоимости по сравнению со своими предшественниками.

Общий принцип работы заключается в том, что устройство берет входной аналоговый сигнал, измеряет его и преобразует в цифровой формат, который может быть использован для анализа, обработки и/или хранения. Ключевыми характеристиками этого АЦП являются его разрядность (14 бит) и скорость преобразования, которые в сочетании обеспечивают высокую точность и производительность при измерениях аналоговых сигналов.

Преобразование аналогового сигнала в цифровой формат в аналого-цифровых преобразователях (АЦП) происходит в несколько этапов и может быть описано следующим образом:

Сэмплирование (Sampling):

- сначала аналоговый сигнал, который нужно измерить, сэмплируется. Это означает, что аналоговый сигнал берется в определенные моменты времени;

- сэмплирование производится с определенной частотой, которая называется частотой дискретизации (sampling rate). Частота дискретизации определяет, сколько раз в секунду будет сделана выборка аналогового сигнала.

Квантование (Quantization):

- после сэмплирования значение амплитуды сигнала в каждый момент времени представляется цифровым кодом. Этот процесс называется квантованием;

- разрядность АЦП определяет, сколько бит используется для представления каждого отсчета. В 14-разрядном АЦП каждый отсчет представляется 14 битами данных. Так как мы используем два ядра АЦП на одном кристалле в одном корпусе. Каждое может работать отдельно, т. е. в два канала. Также есть возможность установить работу на одном такте синхронно и развести один относительно другого на полпериоде, тогда будет одноканальный режим [7]. Схема преобразования аналогового сигнала в цифровой в двоекном 14-разрядном АЦП представлена на рисунке 2.

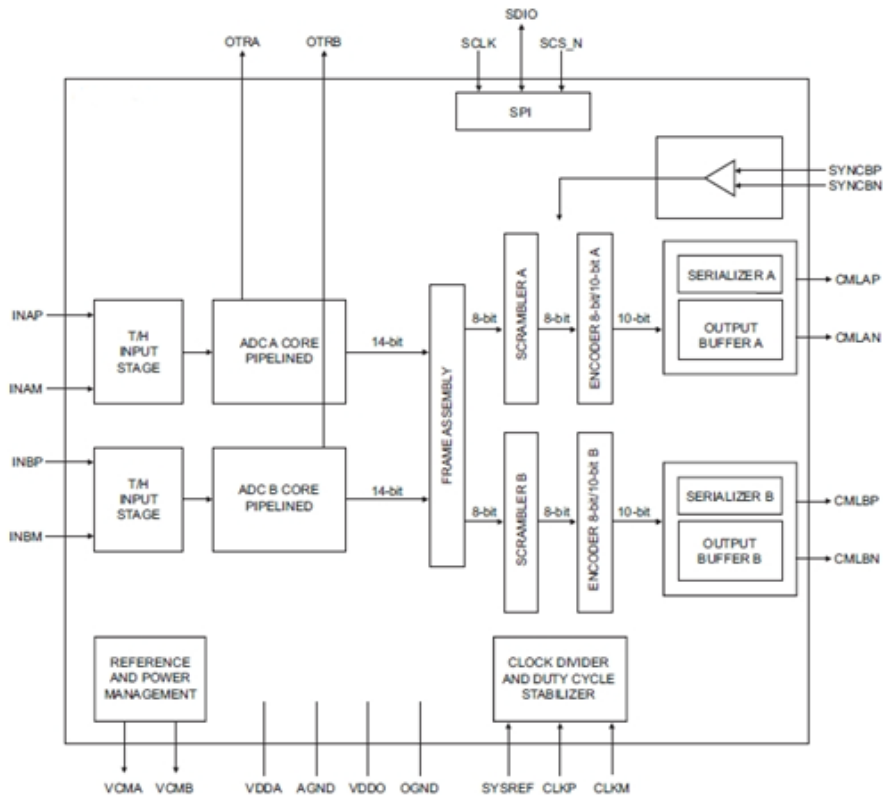


Рисунок 2 – Схема преобразования аналогового сигнала в цифровой в двоядном 14-разрядном АЦП

**Кодирование (Encoding):**

- полученный аналоговый сигнал преобразуется в цифровой формат путем анализа амплитуды и ее представления в виде двоичного кода;
- цифровой код, представляющий амплитуду, сохраняется в регистре или буфере данных АЦП.

**Процессирование (Processing):**

- данные подвергаются дополнительной обработке внутри АЦП, такой как фильтрация, усиление и коррекция ошибок.

**Выход цифровых данных (Digital Output):**

- после завершения процесса преобразования данные могут быть выведены через интерфейс.

Процесс преобразования аналогового сигнала в цифровой формат является ключевой операцией во многих измерительных и управляющих системах, и точность этого процесса имеет большое значение для точности и надежности системы.

Для правильной работы устройства и блокировки фазовой автоподстройки (PLL) необходимо обеспечить наличие входного сигнала. Этот входной сигнал также используется для калибровки управляемого генератора частоты (VCO). Прежде чем начать программирование или настройку устройства, следует подать соответствующий входной сигнал.

Входной сигнал подается на выходы OSCinP и OSCinM устройства. Эти выходы представляют собой дифференциальный буфер, что означает, что общий размах сигнала рассчитывается как разница между сигналом на выводе OSCinP и сигналом на выводе OSCinM. Сигналы могут быть как дифференциальными, так и однополярными, в зависимости от режима работы.

Важно отметить, что для обеспечения правильной работы и согласования сигналов на входе требуется правильное оформление и согласование на обоих концах соединения. Кроме того, также указывает на важность наличия внутреннего смещения и необходимость использования последовательно соединенных конденсаторов переменного тока перед выводами, что также может влиять на корректную работу и согласование сигнала (рис. 3).

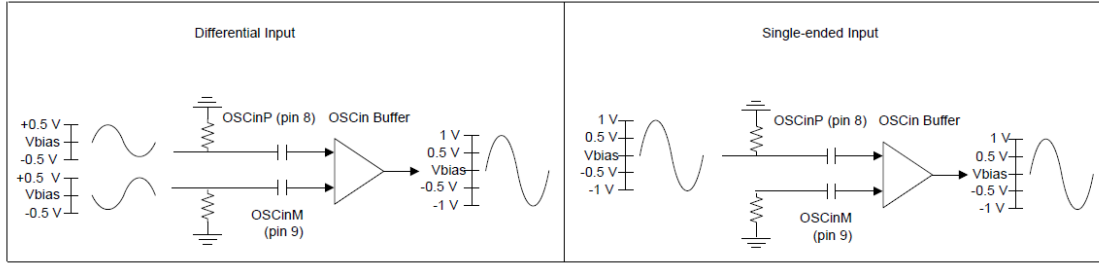


Рисунок 3 – Дифференциальный и односторонний режимы

В функциональном описании продолжается рассмотрение тракта входного сигнала устройства. Этот тракт играет важную роль в обеспечении наилучшей работы фазовой автоподстройки (PLL) и гибкости при решении различных задач.

Чтобы достичь оптимального уровня шума в PLL, важно подавать входной сигнал с частотой 200 МГц, особенно для высшей частоты двухфазного детектора. Для обеспечения гибкости в решении разнообразных задач предусмотрены следующие компоненты в тракте входного сигнала, которые могут быть активированы или отключены по необходимости.

Удвоитель OSCin является низкочастотным удвоителем частоты, предназначенным для умножения входных частот на два. Он способен работать как с нарастающим, так и с спадающим фронтом входного сигнала, но при его включении входной сигнал должен иметь рабочий цикл 50%. Удвоитель полезен в случаях, когда требуется достичь определенной частоты PFD, которая не может быть достигнута только делением или удвоением исходного сигнала.

Предварительный делитель частоты – высокочастотный делитель, способный разделять входной сигнал даже при очень высоких частотах до 1400 МГц. Он обеспечивает возможность деления входного сигнала на более низкие частоты, если это необходимо.

Умножитель в схеме является программируемым умножителем с низким уровнем шума. Совместно с предварительным и после-делителями, он предоставляет гибкость установки PFD на частотах, которые могут создавать критические разрывы в целых числах между частотами VCO и выходными частотами. Это особенно важно для избежания частотных помех и обеспечения стабильной работы [6].

Делитель Post-R используется для дополнительного снижения частоты, особенно в расширенном режиме PFD.

Пользователь должен учитывать характеристики и требования своей конкретной задачи при выборе и настройке этих компонентов (рис. 4) для обеспечения оптимальной работы фазовой автоподстройки (PLL).

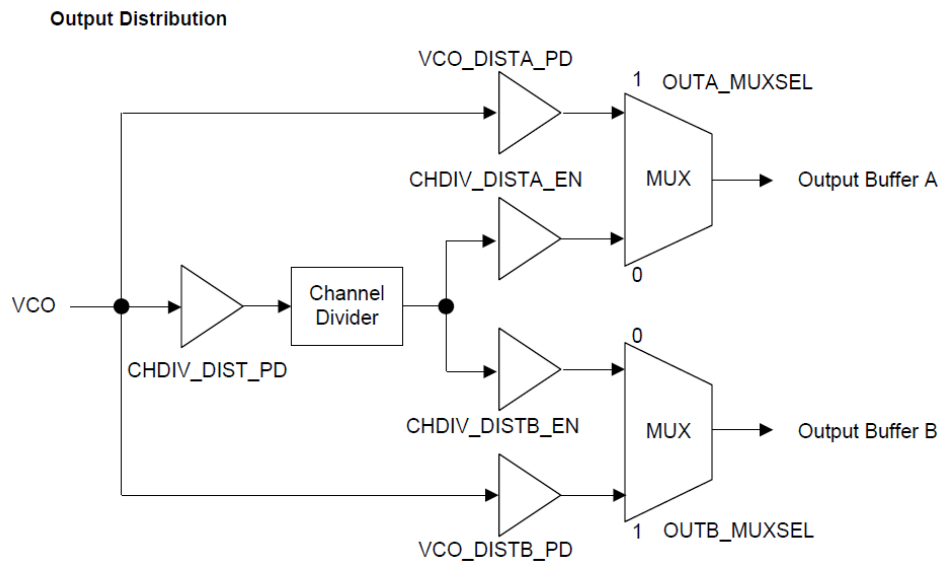


Рисунок 4 – Диаграмма распределения выходов

Для каждого из выходов А и В доступен мультиплексор (MUX), который позволяет выбирать между выходом VCO и выходом канального делителя. Перед этими MUX расположено несколько буферов в тракте распределения сигнала, и их настройка зависит от выбранного маршрута передачи сигнала.

Важно отметить, что отключение неиспользуемых буферов имеет большое значение, поскольку это позволяет изолировать нежелательные сигналы и снизить ненужное потребление электроэнергии. Это действие способствует эффективному управлению энергопотреблением.

У каждого из выходов А и В имеется мультиплексор, который выбирает между выходом VCO и выходом канального делителя. Перед этими мультиплексорами расположены несколько буферов, которые можно настроить в зависимости от выбранного маршрута сигнала. Отключение неиспользуемых буферов помогает изолировать нежелательные сигналы и снижает потребление электроэнергии.

Каждый из выходных буферов А и В обладает программно-управляемым коэффициентом усиления, настраиваемым с использованием регистров OUTA\_POW и OUTB\_POW. Важно отметить, что выход VCO является открытым коллектором и требует внешней подтяжки от вывода RFout к VCC. Существует два варианта подтяжки – резисторная и индуктивная, которые могут быть использованы в зависимости от конкретных требований и приложения. Установка резистора с сопротивлением 50 Ом позволяет согласовать выходной импеданс с 50 Ом, однако максимальная выходная мощность ограничена. Параметры выходного буферного тока должны быть установлены так, чтобы избежать насыщения, поскольку выходная мощность увеличивается с увеличением значения выходного тока. Установка индуктивной подтяжки создает резонанс на интересующей частоте и обеспечивает более высокую выходную мощность при том же токе. Однако это также приводит к более высокому выходному импедансу, и может потребоваться дополнительное согласование.

В режиме фракционной фазовой подстройки можно достичь очень высокой точности в изменении фазового соотношения между выходом и входом. При каждой записи в регистр MASH\_SEED происходит сдвиг фазы согласно уравнению, описанному в документации. Начальное значение MASH\_SEED должно быть меньше знаменателя регистра PLL\_N\_DEN. Фактический фазовый сдвиг можно рассчитать с использованием следующего уравнения: Фазовый сдвиг (градусы) =  $360 \times \text{MASH\_SEED} \times \text{PLL\_N\_PRE} / \text{PLL\_N\_DEN} / [\text{Значение делителя канала}]$ . Это позволяет точно настраивать фазовое соотношение с высоким разрешением в зависимости от конкретных требований и задачи.

Для активации и деактивации питания устройства можно использовать вывод CE, на который подается логическое напряжение HIGH или LOW, либо можно воспользоваться соответствующим битом регистра POWERDOWN, устанавливая его в 0 или 1. Когда необходимо выйти из режима пониженного питания, это можно сделать, вернув вывод CE в состояние HIGH (если включение/выключение питания осуществлялось через вывод CE), либо установив бит регистра POWERDOWN в 0 (если устройство было отключено с помощью этого бита). После выхода из режима пониженного питания для повторной калибровки устройства необходимо перепрограммировать регистр R0.

Выход MUXout может быть настроен для выдачи сигнала, который указывает на состояние блокировки PLL. Если функция обнаружения блокировки разрешена (LD\_EN = 1) и вывод MUXout настроен как выход для индикации блокировки (MUXOUT\_SEL = 1), то, когда устройство находится в состоянии блокировки, на выходе MUXout будет установлен логический уровень HIGH, а когда устройство разблокировано, выход MUXout будет находиться в логическом состоянии LOW.

Выход MUXout можно настроить (MUXOUT\_SEL = 0) на использование последовательного вывода данных при считывании регистров. Для того чтобы считать определенное значение из регистра, нужно установить бит R/W в 1, содержимое поля данных будет игнорироваться. Если запрограммировать этот регистр в устройство, то последовательные данные для считывания будут выведены начиная с 9-го такта. Временная диаграмма считывания регистров представлена на рисунке 5.

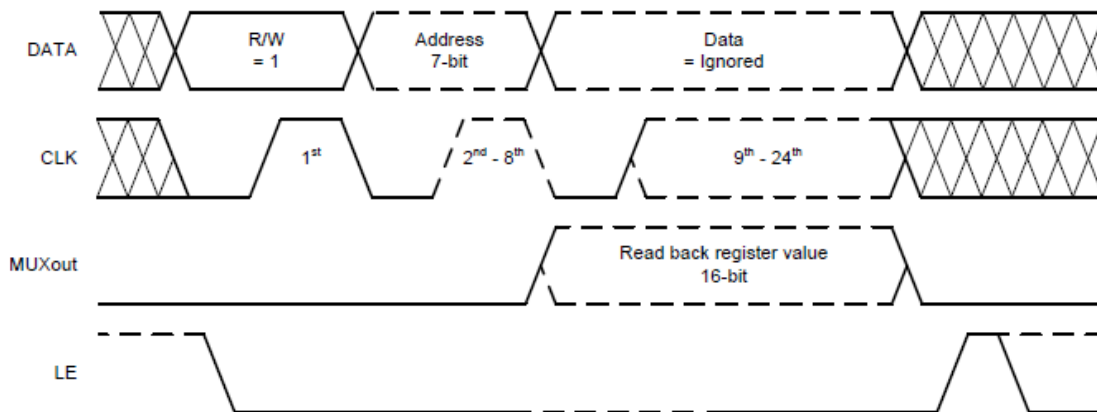


Рисунок 5 – Временная диаграмма считывания регистров

Для осуществления программирования используются 24-разрядные сдвиговые регистры, состоящие из следующих компонентов: бита R/W (самый старший бит MSB), за которым идет 7-битное поле адреса, а затем 16-битное поле данных. Бит R/W (бит 23) определяет операцию: значение 1 обозначает чтение, а 0 – запись. Поле адреса ADDRESS (биты 22:16) применяется для дешифровки адреса внутреннего регистра. Оставшиеся 16 бит формируют поле данных DATA (биты 15:0). Пока сигнал CSB находится в низком состоянии, последовательные данные поступают в сдвиговый регистр при нарастающем фронте тактового импульса (при этом данные подгружаются, начиная с самого старшего бита – MSB). Когда сигнал CSB переходит в высокий уровень, данные из поля данных передаются в выбранный банк регистров.

АЦП обладает рядом встроенных механизмов, спроектированных для обеспечения устойчивости к помехам и шумам. Эти механизмы важны для обеспечения высокой точности и надежности измерений.

**Фильтрация:** АЦП оснащен встроенными фильтрами, способными уменьшать высокочастотные помехи и шумы. Эти фильтры исключают быстро изменяющиеся компоненты сигнала, позволяя нам сосредотачиваться на низкочастотных составляющих и устранять влияние помех [6].

**Дифференциальный вход:** использование дифференциальных входов, которые позволяют нам компенсировать общие помехи. Это достигается путем измерения разницы между двумя входными сигналами, что позволяет игнорировать общие помехи, так как они одинаково воздействуют на оба входа, встроенные алгоритмы для обнаружения и подавления шумов помогают повысить качество измерений. Используются медианные фильтры, средние фильтры и алгоритмы численного усреднения для эффективного удаления шумовых компонентов из сигнала.

**Защита от переполнения:** чтобы избежать искажения данных из-за переполнения, АЦП оборудован механизмами обнаружения и управления переполнением. Это гарантирует сохранение точности измерений и предотвращает потерю данных.

**Компенсация ошибок:** Для коррекции системных ошибок, вызванных помехами и шумами, используется калибровка. Этот процесс автоматически корректирует коэффициенты усиления, смещения и другие параметры АЦП, обеспечивая точность даже в условиях воздействия помех.

**Электромагнитная совместимость (ЭМС):** АЦП разработан с учетом требований по снижению электромагнитных помех. Эффективная конструкция печатных плат и применение экранирования компонентов помогают минимизировать внешние воздействия.

**Дизайн схемы питания:** АЦП предусматривает эффективное планирование и разработку схемы питания, что способствует снижению шумов и помех, поступающих в устройство.

## РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

В ходе работы были исследованы зависимости фазового шума от частоты в схемах с обратной связью и без нее (рис. 6–9).

**График TClosed-Loop Phase Noise:** представляет зависимость фазового шума от частоты с учетом закрытой обратной связи. Горизонтальная ось представляет частоту, а вертикальная ось измеряет индекс фазового шума. Зависимость TClosed-Loop Phase Noise имеет низкие значения фазового шума на низких частотах и постепенно увеличивается с увеличением частоты. Обратная связь помогает снизить фазовый шум, особенно на более высоких частотах.

**График Open-Loop Phase Noise:** представляет зависимость фазового шума от частоты без использования обратной связи. Горизонтальная ось представляет частоту, а вертикальная ось измеряет индекс фазового шума. Зависимость Open-Loop Phase Noise имеет более высокие значения фазового шума на всем диапазоне частот по сравнению с TClosed-Loop Phase Noise. Отсутствие обратной связи означает, что фазовые ошибки и шум от источника частоты передаются напрямую на выход.

График TClosed-Loop Phase Noise демонстрирует, как обратная связь помогает снизить фазовый шум на высоких частотах, что делает его более подходящим для приложений, где требуется высокая стабильность частоты. В то время как график Open-Loop Phase Noise показывает более высокий фазовый шум на всем диапазоне частот из-за отсутствия обратной связи и коррекции фазовых ошибок.

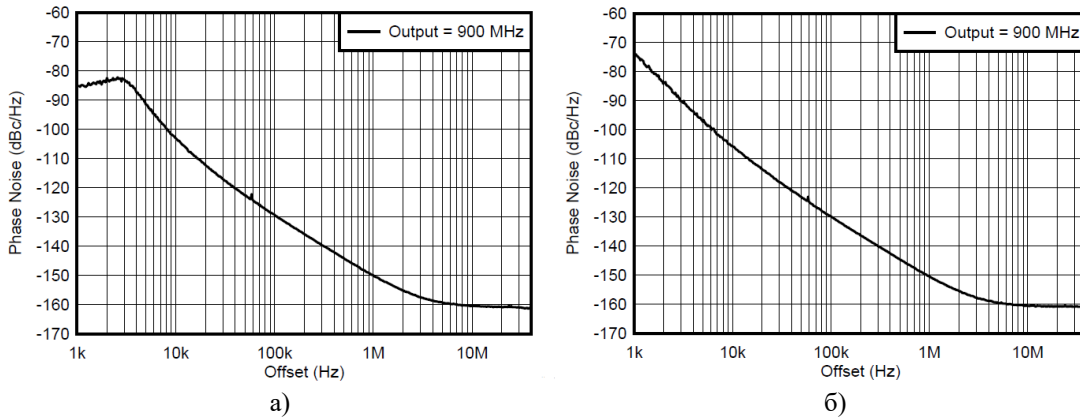


Рисунок 6 – Выходные сигналы 900 MHz: а) Closed-Loop Phase Noise (Фазовый шум в замкнутом контуре) и б) Open-Loop Phase Noise (Фазовый шум разомкнутого контура)

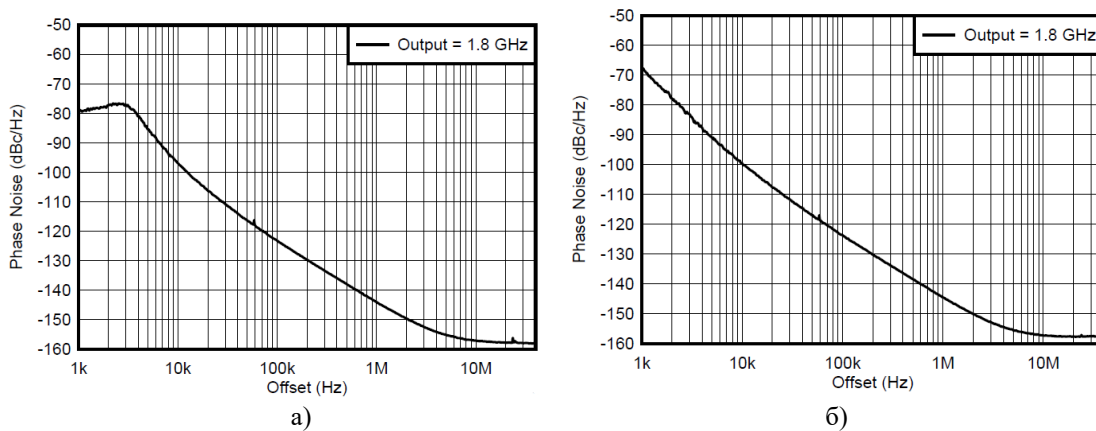


Рисунок 7 – Выходные сигналы 1,8 GHz: а) Closed-Loop Phase Noise (Фазовый шум в замкнутом контуре) и б) Open-Loop Phase Noise (Фазовый шум разомкнутого контура)

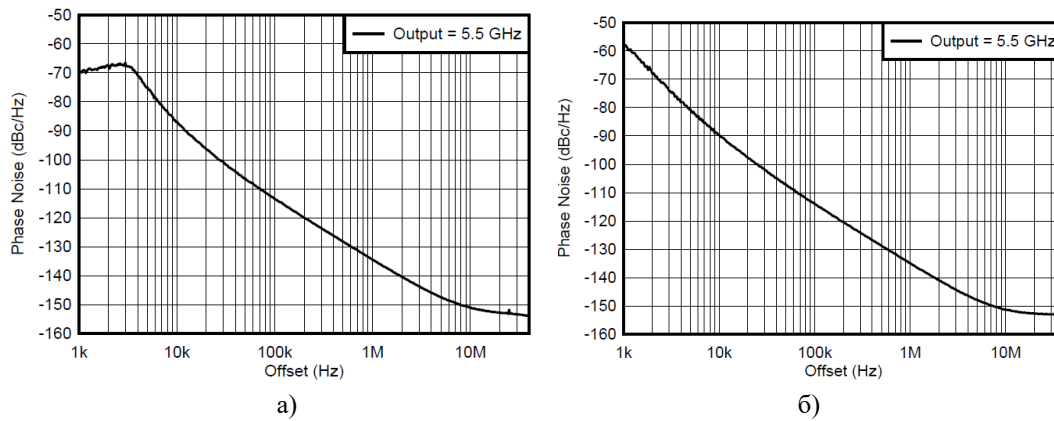


Рисунок 8 – Выходные сигналы 5,5 GHz: а) Closed-Loop Phase Noise (Фазовый шум в замкнутом контуре) и б) Open-Loop Phase Noise (Фазовый шум разомкнутого контура)

Полоса пропускания выходного сигнала 5,4 ГГц и отображение производительности PLL – это график, который исследует производительность фазового замкнутого контура (PLL) при работе с сигналами частотой 5,4 ГГц и при использовании широкой полосы пропускания (Wide Loop Bandwidth). Такой анализ проводится для понимания, как PLL справляется с задачей стабилизации частоты в этих условиях и насколько эффективно его параметры настройки обеспечивают нужную производительность.

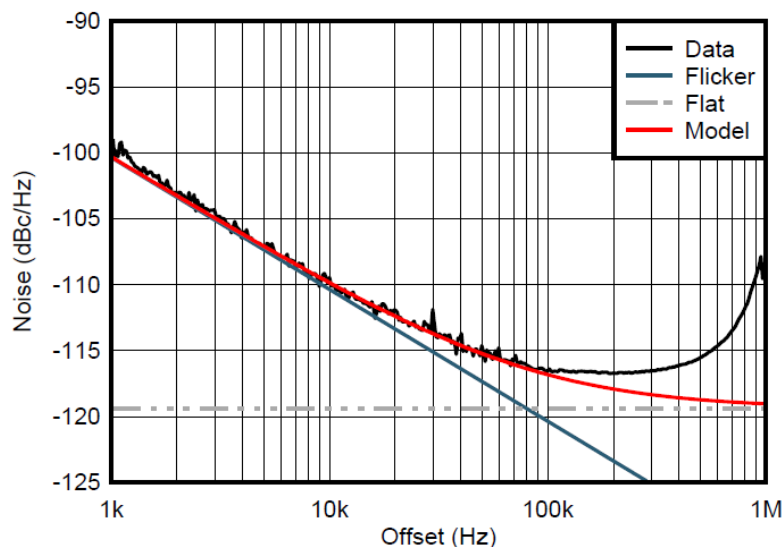


Рисунок 9 – Полоса пропускания выходного сигнала 5,4 ГГц

Отображение производительности PLL:

– Data (Данные): показывает измерения с работой PLL при частоте 5,4 ГГц и широкой полосе пропускания;

– Flicker (Фликер-шум): уровень фликер-шума в выходном сигнале PLL. Фликер-шум является низкочастотным компонентом фазового шума;

– Flat (Флэтность): указывает на равномерность производительности PLL по всей широкой полосе пропускания;

– Model (Модель): график, который отображает теоретическую модель производительности PLL при заданных условиях частоты и ширины полосы пропускания.

Анализ этих зависимостей позволяет оценить, насколько PLL соответствует требованиям конкретного приложения, в данном случае при работе с сигналами частотой 5,4 ГГц и широкой полосой пропускания. Это позволяет инженерам и исследователям оптимизировать настройки PLL и при необходимости внести коррективы, чтобы обеспечить желаемую производительность в конкретном контексте.

### ЗАКЛЮЧЕНИЕ

ADCRU14 имеет более сложную и точную архитектуру АЦП. Это включает в себя использование более продвинутых алгоритмов дискретизации и фильтрации, которые спроектированы для минимизации влияния различных видов шума, таких как квантование, термический и аналоговый шум.

Устойчивость к шуму и точность дискретизации в ADCRU14 могут быть объяснены рядом факторов и особенностей данного АЦП, таких как:

– использование высококачественных компонентов: ADCRU14 построен с использованием высококачественных компонентов, таких как резисторы и конденсаторы, что способствует уменьшению аналогового шума и увеличению точности дискретизации;

– более низкое энергопотребление: меньшее потребление энергии способствует уменьшению электромагнитных помех и уровня шума внутри устройства, что положительно влияет на качество сигнала и точность дискретизации;

– более сложные схемы фильтрации и компенсации: ADCRU14 включает в себя более продвинутые методы компенсации ошибок и шума, такие как цифровая коррекция, линеаризация и устранение систематических искажений;

– коррекция метрологических характеристик: ADCRU14 имеет встроенные механизмы для коррекции метрологических характеристик, такие как нелинейность, дрейф и смещение нуля, что способствует увеличению точности измерений,

– шумоподавление в цифровой области: ADCRU14 также предоставляет возможности для обработки шума в цифровой области путем применения цифровых фильтров и алгоритмов шумоподавления.

В итоге комбинация вышеупомянутых факторов и особенностей делает ADCRU14 более устойчивым к шуму и более точным в дискретизации, что делает его подходящим для приложений, где высокая точность и минимизация шума являются приоритетами.

**Список источников**

1. Брысин, А. Н. Промышленная электроника: импульсные устройства и основы микропроцессорной техники / А. Н. Брысин, С. А. Микаева. – Казань, 2019.
2. Журавлева, Ю. А. Методический лабораторный комплекс по специальным дисциплинам с углубленным изучением схемотехники и программирования электронных и микропроцессорных систем / Ю. А. Журавлева, Э. Л. Ларшина, А. Н. Брысин, С. А. Микаева. – Казань, 2022.
3. Ткачев, С. Е. Цифро-аналоговые преобразователи / С. Е. Ткачев, С. А. Микаева // Наукосфера. – 2022. – № 12–1. – С. 305–308
4. Шельпук, М. И. Разработка программного обеспечения для мониторинга и редактирования информации о студентах / М. И. Шельпук, С. А. Микаева, Ю. А. Журавлева, В. А. Шигапова, О. Ю. Коваленко // Прикаспийский журнал: управление и высокие технологии. – 2022. – № 4 (60). – С. 77–88.
5. Арсентьев, Д. Р. Сравнительный анализ АЦП высокой точности дискретизации сигнала / Д. Р. Арсентьев, С. А. Микаева, Ю. А. Журавлева, Н. В. Абрамова // Фундаментальные, поисковые, прикладные исследования и инновационные проекты : сборник трудов Национальной научно-практической конференции. – Москва, 2023. – С. 294–298.
6. Корн, Г. А. Моделирование случайных процессов на аналого-цифровых машинах / Г. А. Корн. – Москва : Мир, 1968. – 316 с.
7. Горбунов, Ю. Н. Алгоритмы и устройства цифровой стохастической обработки сигналов в радиолокации : учеб. пос. / Ю. Н. Горбунов, А. В. Бондарев. – Москва : НИЦЭВТ, ИПК МРП, 1990. – 144 с.
8. Горбунов, Ю. Н. Цифровая обработка радиолокационных сигналов в условиях использования грубого (малозарядного) квантования / Федеральное космическое агентство ФГУП «Центральный научно-исследовательский институт (ЦНИРТИ) им. академика А. И. Берга». – 2007. – 87 с.

**References**

1. Brysin, A. N., Mikaeva, S. A. *Industrial electronics: pulse devices and fundamentals of microprocessor technology*. Kazan, 2019 (In Russ.).
2. Zhuravleva, Yu. A., Larshina, E. L., Brysin, A. N., Mikaeva, S. A. *Methodological laboratory complex in special disciplines with in-depth study of circuitry and programming of electronic and microprocessor systems*. Kazan, 2022 (In Russ.).
3. Tkachev, S. E., Mikaeva, S. A. Digital-to-analog converters. *Scienceosphere*, 2022, no. 12–1, pp. 305–308 (In Russ.).
4. Shelpuk, M. I., Mikaeva, S. A., Zhuravleva, Yu. A., Shigapova, V. A., Kovalenko, O. Yu. Development of software for monitoring and editing information about students. *Caspian Journal: Control and High Technologies*, 2022, no. 4 (60), pp. 77–88 (In Russ.).
5. Arsentiev, D. R., Mikaeva, S. A., Zhuravleva Yu. A., Abramova N. V. Comparative analysis of high-precision signal sampling ADCs. *Fundamental, exploratory, applied research and innovative projects : collection of proceedings of the National Scientific and Practical Conference*. Moscow, 2023. pp. 294–298 (In Russ.).
6. Korn, G. A. *Modeling random processes on analog-digital machines*. Moscow, Mir Publ., 1968. 316 p. (In Russ.).
7. Gorbunov, Yu. N., Bondarev, A. V. *Algorithms and devices for digital stochastic signal processing in radar : textbook*. Moscow, NICEVT, IPK MRP, 1990. 144 p. (In Russ.).
8. Gorbunov, Yu. N. *Digital processing of radar signals in the conditions of using coarse (low-charge) quantization*. Federal Space Agency FSUE Central Research Institute (TSNIRTI) named after Academician A.I. Berg, 2007. 87 p. (In Russ.).

Статья поступила в редакцию 18.06.2024; одобрена после рецензирования 01.07.2024; принята к публикации 10.07.2024.

The article was submitted 18.06.2024; approved after reviewing 01.07.2024; accepted for publication 10.07.2024.

## ПРАВИЛА ДЛЯ АВТОРОВ

1. В журнале публикуются материалы на английском и русском языках по тематике, соответствующей утвержденным для журнала отраслям наук, группам специальностей.

2. В список соавторов работ включаются только те лица, которые внесли творческий вклад в подготовку представленных материалов. Лицам, оказавшим только техническую помощь, можно выразить благодарность в конце статьи. Один человек может быть автором (соавтором) не более чем двух статей в одном номере журнала, причем единственным автором он может быть только в одной статье.

3. Объем публикаций для научных статей должен быть не менее 8 страниц, а количество источников в библиографическом списке (списке литературы) – не менее 10 позиций.

4. Содержание каждой статьи должно включать следующие элементы: УДК; название статьи; сведения об авторах, включая их место работы, должность, адрес электронной почты; аннотацию объемом от 100 до 250 слов, ключевые слова (от 9 до 13); графическую аннотацию, отражающую содержание статьи; название статьи, сведения об авторах, аннотацию и ключевые слова на английском языке (для англоязычных статей – на русском языке); введение – оно должно заканчиваться формулировкой цели работы в явной форме; собственно текст статьи – очень желательна его сегментация на разделы, имеющие содержательные заголовки; выводы или заключение (должны соответствовать формулировке цели статьи).

5. Для русскоязычных статей приводится два библиографических списка: на языке оригинала статьи; список с транслитерацией русскоязычных источников на латиницу и (дополнительно) приведением в квадратных скобках переводов названий статей и названий источников на английский язык.

В «русскоязычном» библиографическом списке (списке литературы) порядок следования источников – по алфавиту фамилий авторов (сначала русскоязычные источники, потом иноязычные). На все источники, включенные в библиографический список, должны быть даны ссылки в тексте статьи в квадратных скобках. При необходимости авторы могут указывать номера страниц в источниках, на которые даются ссылки. Приветствуются ссылки на иноязычные источники, а также на материалы, опубликованные ранее в журнале «Прикаспийский журнал: управление и высокие технологии». Однако в последнем случае количество таких ссылок не должно превышать 20 % от общего количества источников, включенных в библиографический список. Для источников, имеющих DOI, целесообразно его указывать. При ссылках на статьи, опубликованные в журнале «Прикаспийский журнал: управление и высокие технологии», целесообразно в конце библиографического описания источника в круглых скобках указывать гиперссылку, указывающую на место размещения статьи на странице сайта Астраханского государственного университета.

Ссылки в библиографическом списке на материалы, размещенные в интернете, допускаются при соблюдении следующих условий: если у материала, на который дается ссылка, имеется автор и/или название, то они должны быть указаны для этого источника; должен быть приведен полный маршрут доступа к источнику в интернете; должна быть указана дата обращения (доступа) к источнику.

Ограничения по списку литературы: доля самоцитирований для любого из авторов статьи, а также по совокупности всех авторов статьи, не должна превышать 25 %; доля ссылок на статьи с участием одного автора, не являющегося автором (соавтором) статьи, не должна превышать 25 %.

6. Суммарная доля таблиц и иллюстраций в общем объеме представляемой статьи не должна превышать 40 %. Под иллюстрациями понимаются следующие объекты: диаграммы; графики; рисунки; эскизы; фотографии; карты и т. п.

7. Доля оригинального текста в статьях (оцениваемого через систему «Антиплагиат» на сайте [www.antiplagiat.ru](http://www.antiplagiat.ru)) должна быть не менее 80 %.

8. Указание на то, что работа финансируется по какому-либо гранту, в рамках Федеральной целевой программы, государственного заказа и пр. дается в виде постраничной сноски после заголовка (названия) работы.

9. В сведения об авторах работ помимо места работы и должности целесообразно включать ORCID автора и гиперссылку на страничку с его личными наукометрическими показателями на сайте [www.elibrary.ru](http://www.elibrary.ru). По желанию можно привести также ссылки на странички с наукометрическими показателями на Scopus, в ResearchGate; на личную страничку, размещенную на сайте организации.

10. Основные технические требования к оформлению статей (материалов):

10.1. Текст должен быть расположен по ширине страницы формата А4 с учётом полей (все поля по 2,5 см), набран шрифтом Times New Roman, кегль 12, межстрочный интервал 1,0. В таблицах, подрисуночных надписях допускается уменьшенный шрифт – вплоть до 10 кегля. Альбомная ориентация страниц допускается только в порядке исключения для следующих случаев: широкоформатные таблицы с большим количеством колонок; иллюстрации большого размера, которые не умещаются на странице с книжной ориентацией.

Абзацные отступы одинаковы по всему тексту – 0,75 см. Кавычки («»), скобки ([ ], ( )), маркеры и другие знаки должны быть аналогичными на протяжении всего предоставляемого для публикации материала.

# **ПРИКАСПИЙСКИЙ ЖУРНАЛ: управление и высокие технологии**

НАУЧНО-ТЕХНИЧЕСКИЙ ЖУРНАЛ

**2024  
№ 3 (67)**

Свидетельство о регистрации средства массовой информации  
Федеральной службы по надзору в сфере массовых коммуникаций,  
связи и охраны культурного наследия  
ПИ № ФС77-31932 от 16 мая 2008 г.

Учредитель  
Астраханский государственный университет имени В. Н. Татищева  
Российская Федерация, 414056, г. Астрахань, ул. Татищева, 20а

Адрес редакции:  
Российская Федерация, 414056, г. Астрахань, ул. Татищева, 20

Адрес издателя:  
Российская Федерация, 414056, г. Астрахань, ул. Татищева, 20а

Издание включено в Интернет-каталог  
ООО «Агентство «Книга-Сервис» 2022/1

**Главный редактор И.М. Ажмухамедов**

Редактирование,  
компьютерная правка, верстка *Н.Н. Сахно*

Дата выхода в свет 30.11.2024 г.

Цена свободная  
Уч.-изд. 12,0. Усл. печ. л. 16,8.  
Заказ № 4635. Тираж 500 экз. (первый завод – 22 экз.)

---

Астраханский государственный университет имени В. Н. Татищева  
414056, г. Астрахань, ул. Татищева, 20а  
тел. (8512) 24-66-60 (доб. 3; издательско-полиграфический отдел)  
Отпечатано в Астраханской цифровой типографии  
414040, г. Астрахань, пл. К. Маркса, 33  
тел./факс (8512) 54-00-11, 73-40-40,  
E-mail: a-d-t@mail.ru