

ISSN 2518-1726 (Online),  
ISSN 1991-346X (Print)



«ҚАЗАҚСТАН РЕСПУБЛИКАСЫ  
ҰЛТТЫҚ ҒЫЛЫМ АКАДЕМИЯСЫ» РҚБ  
«ХАЛЫҚ» ЖҚ

# Х А Б А Р Л А Р Ы

---

---

## ИЗВЕСТИЯ

РОО «НАЦИОНАЛЬНОЙ  
АКАДЕМИИ НАУК РЕСПУБЛИКИ  
КАЗАХСТАН»  
ЧФ «Халық»

## N E W S

OF THE ACADEMY OF SCIENCES  
OF THE REPUBLIC OF  
KAZAKHSTAN  
«Halyk» Private Foundation

**SERIES**  
**PHYSICS AND INFORMATION TECHNOLOGY**

**3 (347)**

**JULY – SEPTEMBER 2023**

PUBLISHED SINCE JANUARY 1963  
PUBLISHED 4 TIMES A YEAR

ALMATY, NAS RK



## ЧФ «ХАЛЫҚ»

В 2016 году для развития и улучшения качества жизни казахстанцев был создан частный Благотворительный фонд «Халык». За годы своей деятельности на реализацию благотворительных проектов в областях образования и науки, социальной защиты, культуры, здравоохранения и спорта, Фонд выделил более 45 миллиардов тенге.

Особое внимание Благотворительный фонд «Халык» уделяет образовательным программам, считая это направление одним из ключевых в своей деятельности. Оказывая поддержку отечественному образованию, Фонд вносит свой посильный вклад в развитие качественного образования в Казахстане. Тем самым способствуя росту числа людей, способных менять жизнь в стране к лучшему – профессионалов в различных сферах, потенциальных лидеров и «великих умов». Одной из значимых инициатив фонда «Халык» в образовательной сфере стал проект *Ozgeris powered by Halyk Fund* – первый в стране бизнес-инкубатор для учащихся 9-11 классов, который помогает развивать необходимые в современном мире предпринимательские навыки. Так, на содействие малому бизнесу школьников было выделено более 200 грантов. Для поддержки талантливых и мотивированных детей Фонд неоднократно выделял гранты на обучение в Международной школе «Мирас» и в *Astana IT University*, а также помог казахстанским школьникам принять участие в престижном конкурсе «*USTEM Robotics*» в США. Авторские работы в рамках проекта «Тәлімгер», которому Фонд оказал поддержку, легли в основу учебной программы, учебников и учебно-методических книг по предмету «Основы предпринимательства и бизнеса», преподаваемого в 10-11 классах казахстанских школ и колледжей.

Помимо помощи школьникам, учащимся колледжей и студентам Фонд считает важным внести свой вклад в повышение квалификации педагогов, совершенствование их знаний и навыков, поскольку именно они являются проводниками знаний будущих поколений казахстанцев. При поддержке Фонда «Халык» в южной столице был организован ежегодный городской конкурс педагогов «*Almaty Digital Ustaz*».

Важной инициативой стал реализуемый проект по обучению основам финансовой грамотности преподавателей из восьми областей Казахстана, что должно оказать существенное влияние на воспитание финансовой грамотности и предпринимательского мышления у нового поколения граждан страны.

Необходимую помощь Фонд «Халык» оказывает и тем, кто особенно остро в ней нуждается. В рамках социальной защиты населения активно проводится работа по поддержке детей, оставшихся без родителей, детей и взрослых из социально уязвимых слоев населения, людей с ограниченными возможностями, а также обеспечению нуждающихся социальным жильем, строительству социально важных объектов, таких как детские сады, детские площадки и физкультурно-оздоровительные комплексы.

В копилку добрых дел Фонда «Халык» можно добавить оказание помощи детскому спорту, куда относится поддержка в развитии детского футбола и карате в нашей стране. Жизненно важную помощь Благотворительный фонд «Халык» оказал нашим соотечественникам во время недавней пандемии COVID-19. Тогда, в разгар тяжелой борьбы с коронавирусной инфекцией Фонд выделил свыше 11 миллиардов тенге на приобретение необходимого медицинского оборудования и дорогостоящих медицинских препаратов, автомобилей скорой медицинской помощи и средств защиты, адресную материальную помощь социально уязвимым слоям населения и денежные выплаты медицинским работникам.

В 2023 году наряду с другими проектами, нацеленными на повышение благосостояния казахстанских граждан Фонд решил уделить особое внимание науке, поскольку она является частью общественной культуры, а уровень ее развития определяет уровень развития государства.

Поддержка Фондом выпуска журналов Национальной Академии наук Республики Казахстан, которые входят в международные фонды Scopus и Wos и в которых публикуются статьи отечественных ученых, докторантов и магистрантов, а также научных сотрудников высших учебных заведений и научно-исследовательских институтов нашей страны является не менее значимым вкладом Фонда в развитие казахстанского общества.

**С уважением,  
Благотворительный Фонд «Халык»!**

### **БАС РЕДАКТОР:**

**МУТАНОВ Ғалымқайыр Мұтанұлы**, техника ғылымдарының докторы, профессор, ҚР ҰҒА академигі, ҚР БҒМ ҒК «Ақпараттық және есептеу технологиялары институты» бас директорының м.а. (Алматы, Қазақстан), **Н-5**

### **БАС РЕДАКТОРДЫҢ ОРЫНБАСАРЫ:**

**МАМЫРБАЕВ Өркен Жұмажанұлы**, ақпараттық жүйелер мамандығы бойынша философия докторы (Ph.D), ҚР БҒМ Ғылым комитеті «Ақпараттық және есептеуші технологиялар институты» РМК жауапты хатшысы (Алматы, Қазақстан), **Н=5**

### **РЕДАКЦИЯ АЛҚАСЫ:**

**ҚАЛИМОЛДАЕВ Мақсат Нұрәділұлы**, физика-математика ғылымдарының докторы, профессор, ҚР ҰҒА академигі (Алматы, Қазақстан), **Н=7**

**БАЙГУНЧЕКОВ Жұмаділ Жанабайұлы**, техника ғылымдарының докторы, профессор, ҚР ҰҒА академигі, Кибернетика және ақпараттық технологиялар институты, Сатпаев университетінің Қолданбалы механика және инженерлік графика кафедрасы, (Алматы, Қазақстан), **Н=3**

**ВОЙЧИК Вальдемар**, техника ғылымдарының докторы (физика), Люблин технологиялық университетінің профессоры (Люблин, Польша), **Н=23**

**БОШКАЕВ Қуантай Авғазыұлы**, Ph.D. Теориялық және ядролық физика кафедрасының доценті, әл-Фараби атындағы Қазақ ұлттық университеті (Алматы, Қазақстан), **Н=10**

**QUEVEDO Nemando**, профессор, Ядролық ғылымдар институты (Мехико, Мексика), **Н=28**

**ЖҮСІПОВ Марат Абжанұлы**, физика-математика ғылымдарының докторы, теориялық және ядролық физика кафедрасының профессоры, әл-Фараби атындағы Қазақ ұлттық университеті (Алматы, Қазақстан), **Н=7**

**КОВАЛЕВ Александр Михайлович**, физика-математика ғылымдарының докторы, Украина ҰҒА академигі, Қолданбалы математика және механика институты (Донецк, Украина), **Н=5**

**РАМАЗАНОВ Тілекқабұл Сәбитұлы**, физика-математика ғылымдарының докторы, профессор, ҚР ҰҒА академигі, әл-Фараби атындағы Қазақ ұлттық университетінің ғылыми-инновациялық қызмет жөніндегі проректоры, (Алматы, Қазақстан), **Н=26**

**ТАКИБАЕВ Нұрғали Жабағаұлы**, физика-математика ғылымдарының докторы, профессор, ҚР ҰҒА академигі, әл-Фараби атындағы Қазақ ұлттық университеті (Алматы, Қазақстан), **Н=5**

**ТИГИНЯНУ Ион Михайлович**, физика-математика ғылымдарының докторы, академик, Молдова Ғылым Академиясының президенті, Молдова техникалық университеті (Кишинев, Молдова), **Н=42**

**ХАРИН Станислав Николаевич**, физика-математика ғылымдарының докторы, профессор, ҚР ҰҒА академигі, Қазақстан-Британ техникалық университеті (Алматы, Қазақстан), **Н=10**

**ДАВЛЕТОВ Асқар Ербуланович**, физика-математика ғылымдарының докторы, профессор, әл-Фараби атындағы Қазақ ұлттық университеті (Алматы, Қазақстан), **Н=12**

**КАЛАНДРА Пьетро**, Ph.D (физика), Нанокұрылымды материалдарды зерттеу институтының профессоры (Рим, Италия), **Н=26**

**«ҚР ҰҒА Хабарлары. Физика және информатика сериясы».**

**ISSN 2518-1726 (Online),**

**ISSN 1991-346X (Print)**

Меншіктеуші: «Қазақстан Республикасының Ұлттық ғылым академиясы» РҚБ (Алматы қ.). Қазақстан Республикасының Ақпарат және қоғамдық даму министрлігінің Ақпарат комитетінде 14.02.2018 ж. берілген **№ 16906-Ж** мерзімдік басылым тіркеуіне қойылу туралы куәлік.

Тақырыптық бағыты: *физика және ақпараттық коммуникациялық технологиялар сериясы*. Қазіргі уақытта: *«ақпараттық технологиялар» бағыты бойынша ҚР БҒМ БҒСБК ұсынған журналдар тізіміне енді.*

Мерзімділігі: *жылына 4 рет.*

Тиражы: *300 дана.*

Редакцияның мекен-жайы: *050010, Алматы қ., Шевченко көш., 28, 219 бөл., тел.: 272-13-19*  
*<http://www.physico-mathematical.kz/index.php/en/>*

### ГЛАВНЫЙ РЕДАКТОР:

**МУТАНОВ Галимжаир Мутанович**, доктор технических наук, профессор, академик НАН РК, и.о. генерального директора «Института информационных и вычислительных технологий» КН МОН РК (Алматы, Казахстан), **Н=5**

### ЗАМЕСТИТЕЛЬ ГЛАВНОГО РЕДАКТОРА:

**МАМЫРБАЕВ Оркен Жумажанович**, доктор философии (PhD) по специальности Информационные системы, ответственный секретарь РГП «Института информационных и вычислительных технологий» Комитета науки МОН РК (Алматы, Казахстан), **Н=5**

### РЕДАКЦИОННАЯ КОЛЛЕГИЯ:

**КАЛИМОЛДАЕВ Максат Нурадилович**, доктор физико-математических наук, профессор, академик НАН РК (Алматы, Казахстан), **Н=7**

**БАЙГУНЧЕКОВ Жумадил Жанабаевич**, доктор технических наук, профессор, академик НАН РК, Институт кибернетики и информационных технологий, кафедра прикладной механики и инженерной графики, Университет Сагпаева (Алматы, Казахстан), **Н=3**

**ВОЙЧИК Вальдемар**, доктор технических наук (физ.-мат.), профессор Люблинского технологического университета (Люблин, Польша), **Н=23**

**БОШКАЕВ Куантай Авгазыевич**, доктор Ph.D, преподаватель, доцент кафедры теоретической и ядерной физики, Казахский национальный университет им. аль-Фараби (Алматы, Казахстан), **Н=10**

**QUEVEDO Hemando**, профессор, Национальный автономный университет Мексики (UNAM), Институт ядерных наук (Мехико, Мексика), **Н=28**

**ЖУСУПОВ Марат Абжанович**, доктор физико-математических наук, профессор кафедры теоретической и ядерной физики, Казахский национальный университет им. аль-Фараби (Алматы, Казахстан), **Н=7**

**КОВАЛЕВ Александр Михайлович**, доктор физико-математических наук, академик НАН Украины, Институт прикладной математики и механики (Донецк, Украина), **Н=5**

**РАМАЗАНОВ Тлексабул Сабитович**, доктор физико-математических наук, профессор, академик НАН РК, проректор по научно-инновационной деятельности, Казахский национальный университет им. аль-Фараби (Алматы, Казахстан), **Н=26**

**ТАКИБАЕВ Нурғали Жабағевич**, доктор физико-математических наук, профессор, академик НАН РК, Казахский национальный университет им. аль-Фараби (Алматы, Казахстан), **Н=5**

**ТИГИНЯНУ Ион Михайлович**, доктор физико-математических наук, академик, президент Академии наук Молдовы, Технический университет Молдовы (Кишинев, Молдова), **Н=42**

**ХАРИН Станислав Николаевич**, доктор физико-математических наук, профессор, академик НАН РК, Казахстанско-Британский технический университет (Алматы, Казахстан), **Н=10**

**ДАВЛЕТОВ Аскар Ербуланович**, доктор физико-математических наук, профессор, Казахский национальный университет им. аль-Фараби (Алматы, Казахстан), **Н=12**

**КАЛАНДРА Пьетро**, доктор философии (Ph.D, физика), профессор Института по изучению наноструктурированных материалов (Рим, Италия), **Н=26**

### «Известия НАН РК. Серия физика и информатики».

ISSN 2518-1726 (Online),

ISSN 1991-346X (Print)

Собственник: *Республиканское общественное объединение «Национальная академия наук Республики Казахстан» (г. Алматы).*

Свидетельство о постановке на учет периодического печатного издания в Комитете информации Министерства информации и общественного развития Республики Казахстан **№ 16906-Ж** выданное 14.02.2018 г.

Тематическая направленность: *серия физика и информационные коммуникационные технологии.* В настоящее время: *вошел в список журналов, рекомендованных ККСОН МОН РК по направлению «информационные коммуникационные технологии».*

Периодичность: *4 раз в год.*

Тираж: *300 экземпляров.*

Адрес редакции: *050010, г. Алматы, ул. Шевченко, 28, оф. 219, тел.: 272-13-19*

*<http://www.physico-mathematical.kz/index.php/en/>*

#### **EDITOR IN CHIEF:**

**MUTANOV Galimkair Mutanovich**, doctor of technical Sciences, Professor, Academician of NAS RK, acting director of the Institute of Information and Computing Technologies of SC MES RK (Almaty, Kazakhstan), **H=5**

#### **DEPUTY EDITOR-IN-CHIEF**

**MAMYRBAYEV Orken Zhumazhanovich**, Ph.D. in the specialty "Information systems, executive secretary of the RSE "Institute of Information and Computational Technologies", Committee of Science MES RK (Almaty, Kazakhstan) **H=5**

#### **EDITORIAL BOARD:**

**KALIMOLDAYEV Maksat Nuradilovich**, doctor in Physics and Mathematics, Professor, Academician of NAS RK (Almaty, Kazakhstan), **H=7**

**BAYGUNCHEKOV Zhumadil Zhanabayevich**, doctor of Technical Sciences, Professor, Academician of NAS RK, Institute of Cybernetics and Information Technologies, Department of Applied Mechanics and Engineering Graphics, Satbayev University (Almaty, Kazakhstan), **H=3**

**WOICIK Waldemar**, Doctor of Phys.-Math. Sciences, Professor, Lublin University of Technology (Lublin, Poland), **H=23**

**BOSHKAYEV Kuantai Avgazievich**, PhD, Lecturer, Associate Professor of the Department of Theoretical and Nuclear Physics, Al-Farabi Kazakh National University (Almaty, Kazakhstan), **H=10**

**QUEVEDO Hemando**, Professor, National Autonomous University of Mexico (UNAM), Institute of Nuclear Sciences (Mexico City, Mexico), **H=28**

**ZHUSSUPOV Marat Abzhanovich**, Doctor in Physics and Mathematics, Professor of the Department of Theoretical and Nuclear Physics, Al-Farabi Kazakh National University (Almaty, Kazakhstan), **H=7**

**KOVALEV Alexander Mikhailovich**, Doctor in Physics and Mathematics, Academician of NAS of Ukraine, Director of the State Institution «Institute of Applied Mathematics and Mechanics» DPR (Donetsk, Ukraine), **H=5**

**RAMAZANOV Tlekkabul Sabitovich**, Doctor in Physics and Mathematics, Professor, Academician of NAS RK, Vice-Rector for Scientific and Innovative Activity, Al-Farabi Kazakh National University (Almaty, Kazakhstan), **H=26**

**TAKIBAYEV Nurgali Zhabagaevich**, Doctor in Physics and Mathematics, Professor, Academician of NAS RK, Al-Farabi Kazakh National University (Almaty, Kazakhstan), **H=5**

**TIGHINEANU Ion Mikhailovich**, Doctor in Physics and Mathematics, Academician, Full Member of the Academy of Sciences of Moldova, President of the AS of Moldova, Technical University of Moldova (Chisinau, Moldova), **H=42**

**KHARIN Stanislav Nikolayevich**, Doctor in Physics and Mathematics, Professor, Academician of NAS RK, Kazakh-British Technical University (Almaty, Kazakhstan), **H=10**

**DAVLETOV Askar Erbulanovich**, Doctor in Physics and Mathematics, Professor, Al-Farabi Kazakh National University (Almaty, Kazakhstan), **H=12**

**CALANDRA Pietro**, PhD in Physics, Professor at the Institute of Nanostructured Materials (Monterotondo Station Rome, Italy), **H=26**

#### **News of the National Academy of Sciences of the Republic of Kazakhstan.**

**Series of physics and informatics.**

**ISSN 2518-1726 (Online),**

**ISSN 1991-346X (Print)**

Owner: RPA «National Academy of Sciences of the Republic of Kazakhstan» (Almaty). The certificate of registration of a periodical printed publication in the Committee of information of the Ministry of Information and Social Development of the Republic of Kazakhstan **No. 16906-ЖК**, issued 14.02.2018  
Thematic scope: *series physics and information technology.*

Currently: *included in the list of journals recommended by the CCSES MES RK in the direction of «information and communication technologies».*

Periodicity: *4 times a year.*

Circulation: *300 copies.*

Editorial address: *28, Shevchenko str., of. 219, Almaty, 050010, tel. 272-13-19*

*<http://www.physico-mathematical.kz/index.php/en/>*

NEWS OF THE NATIONAL ACADEMY OF SCIENCES OF THE REPUBLIC OF KAZAKHSTAN  
PHYSICO-MATHEMATICAL SERIES

ISSN 1991-346X

Volume 3. Number 347 (2023). 7–17

<https://doi.org/10.32014/2023.2518-1726.200>

UDC 004.942

© G. Abdikalyk\*, A. Mukanova, A. Nazyrova, 2023

Astana International University, Astana, Kazakhstan.

E-mail: [gulnazymabdikalik@gmail.ru](mailto:gulnazymabdikalik@gmail.ru)

### NAMED ENTITY RECOGNITION FOR KAZAKH LANGUAGE USING CRF AND RANDOM FOREST MODELS: A COMPARATIVE STUDY

**Abdikalyk Gulnazym** — grad student of the Astana International University. 010000. Astana, Kazakhstan

E-mail: [gulnazymabdikalik@gmail.ru](mailto:gulnazymabdikalik@gmail.ru). ORCID ID: <https://orcid.org/0009-0008-5216-0707>;

**Assel Mukanova** — PhD, assoc. professor of the Astana International University. 010000. Astana, Kazakhstan

E-mail: [asiserikovna@gmail.com](mailto:asiserikovna@gmail.com). ORCID ID: <https://orcid.org/0000-0002-8964-3891>;

**Nazyrova Aizhan** — senior lecturer of the Astana International University. 010000. Astana, Kazakhstan

E-mail: [ayzhan.nazyrova@gmail.com](mailto:ayzhan.nazyrova@gmail.com). ORCID ID: <https://orcid.org/0000-0002-9162-6791>.

**Abstract.** The activity of recognizing and categorizing named items in text is known as named entity recognition (NER), and it is essential to natural language processing. However, due to morphological complexity and scarce linguistic resources, NER for under-resourced languages like Kazakh presents distinct difficulties. The performance of two well-known machine learning models, Conditional Random Fields (CRF) and Random Forest, for NER in the Kazakh language was thoroughly compared in this scientific work. The work addresses feature engineering strategies catered to the morphological complexity of Kazakh and makes use of a benchmark dataset generated specifically for Kazakh NER. While Random Forest models manage high-dimensional feature spaces and intricate interactions in the data, CRF models capture sequential dependencies and contextual information. The efficiency of both the CRF and Random Forest models for Kazakh NER is shown by experimental findings. However, the scarcity of labeled data has an impact on how well these models work. Future research directions include extending annotated datasets through partnerships with linguists and native speakers in order to address this constraint. The study also emphasizes how crucial it is to deal with Kazakh's complicated morphology in NER. Word stems and part-of-speech tags are among the morphological qualities that CRF and Random Forest models integrate, which enhances the recognition of named entities

in a variety of inflections and variants. The comparison analysis sheds light on the advantages and disadvantages of the Random Forest and CRF models for Kazakh NER. While Random Forest models may manage complex linkages and feature interactions, CRF models excel at capturing sequential dependencies and utilizing contextual information. The NER task's specific needs and characteristics determine the model to use. In conclusion, by offering information on the effectiveness of CRF and Random Forest models, this comparative study advances the field of NER for the Kazakh language. It shows the value of handling morphological complexity, the necessity of annotated data, and directs future study aimed at enhancing Kazakh NER systems.

**Keywords:** Named Entity Recognition (NER), Kazakh language, Conditional Random Fields (CRF), Random Forest, comparative study

*This research has been funded by the Science Committee of the Ministry of Science and Higher Education of the Republic of Kazakhstan (Grant No. AP19577922)*

© Г. Әбдіқалық\*, Ә. Мұқанова, А. Назырова, 2023

Астана Халықаралық Университеті, Астана, Қазақстан.

E-mail: [gulnazymabdikalik@gmail.ru](mailto:gulnazymabdikalik@gmail.ru)

## **CRF ЖӘНЕ RANDOM FOREST МОДЕЛДЕРІНІҢ КӨМЕГІМЕН ҚАЗАҚ ТІЛІНДЕ АТАЛҒАН ОБЪЕКТІЛЕРДІ ТАҢУ: САЛЫСТЫРМАЛЫ ЗЕРТТЕУ**

**Әбдіқалық Гүлназым** — Астана Халықаралық университетінің магистранты. 010000. Астана, Қазақстан

E-mail: [gulnazymabdikalik@gmail.ru](mailto:gulnazymabdikalik@gmail.ru). ORCID ID: <https://orcid.org/0009-0008-5216-0707>;

**Әсел Мұқанова** — Астана Халықаралық университетінің доценті, PhD. 010000. Астана, Қазақстан

E-mail: [asiserikovna@gmail.com](mailto:asiserikovna@gmail.com). ORCID ID: <https://orcid.org/0000-0002-8964-3891>;

**Назырова Айжан Есболовна** — Астана Халықаралық университетінің аға оқытушысы, 010000. Астана, Қазақстан

E-mail: [ayzhan.nazyrova@gmail.com](mailto:ayzhan.nazyrova@gmail.com). ORCID ID: <https://orcid.org/0000-0002-9162-6791>.

**Аннотация.** Мәтіндегі аталған объектілерді таңу және санаттау әрекеті аталған объектілерді таңу (NER) деп аталады және табиғи тілді өңдеу үшін үлкен маңызға ие. Алайда, морфологияның күрделілігіне және лингвистикалық ресурстардың тапшылығына байланысты, қазақ сияқты ресурстары шектеулі тілдердегі аталған объектілерді таңу белгілі бір қиындықтармен байланысты. Бұл ғылыми жұмыста пер үшін машиналық оқытудың екі белгілі моделінің — шартты кездейсоқ өрістердің (CRF) және кездейсоқ орманның қазақ тіліндегі тиімділігін салыстыру жүргізілді. Жұмыста қазақ тілінің морфологиялық күрделілігін ескере отырып, белгілерді іріктеу стратегиялары қарастырылады және пер қазақ тілі үшін арнайы жасалған эталондық деректер жиынтығы пайдаланылады. Random Forest модельдері жоғары өлшемді белгілер кеңістігін және деректердегі күрделі өзара әрекеттесуді басқарса, CRF



модельдері дәйекті тәуелділіктер мен контекстік ақпаратты көрсетеді. Қазақ NER үшін CRF және Random Forest модельдерінің тиімділігі эксперименттік нәтижелермен расталады. Алайда, бұл модельдердің тиімділігіне таңбаланған деректердің жетіспеушілігі әсер етеді. Зерттеудің болашақ бағыттары осы мәселені шешу үшін лингвистермен және ана тілінде сөйлейтіндермен ынтымақтастық арқылы аннотацияланған деректер жиынтығын кеңейтуді қамтиды. Зерттеу сонымен қатар NER-де қазақ тілінің күрделі морфологиясын ескеру қаншалықты маңызды екенін көрсетеді. CRF және Random Forest модельдерін ескеретін морфологиялық қасиеттердің ішінде сөздерде сөйлеу бөліктерінің сабақтары мен тегтерінің болуын атап өтуге болады, бұл әртүрлі ауытқулар мен нұсқаларда аталған нысандарды тануды жақсартады. Салыстырмалы талдау қазақстандық NER үшін Random Forest және CRF модельдерінің артықшылықтары мен кемшіліктеріне жарық түсіреді. Random Forest модельдері күрделі байланыстар мен белгілердің өзара әрекеттесуін басқара алатын болса, CRF модельдері дәйекті тәуелділіктерді анықтауда және контекстік ақпаратты пайдалануда жақсы жұмыс істейді. NER тапсырмасының нақты қажеттіліктері мен сипаттамалары модель таңдауын анықтайды. Қорытындылай келе, CRF және Random Forest модельдерінің тиімділігі туралы ақпаратты ұсына отырып, бұл салыстырмалы зерттеу Қазақ тілі үшін NER дамуына ықпал ететінін атап өткен жөн. Ол морфологиялық күрделілікпен жұмыстың құндылығын, аннотацияланған деректердің қажеттілігін көрсетеді және қазақ тілінің NER жүйелерін жетілдіруге бағытталған одан әрі зерттеулерді бағыттайды.

**Түйін сөздер:** Аталған объектілерді тану (NER), қазақ тілі, шартты кездейсоқ өрістер (CRF), кездейсоқ орман, салыстырмалы зерттеу

*Бұл жұмысты Қазақстан Республикасы Ғылым және жоғары білім министрлігінің Ғылым комитеті қаржылай қолдады (грант AP19577922).*

© Г. Абдикалык\*, А. Муканова, А. Назырова, 2023

Международный университет Астана, Астана, Казахстан.

E-mail: [gulnazymabdikalik@gmail.ru](mailto:gulnazymabdikalik@gmail.ru)

## РАСПОЗНАВАНИЕ ИМЕНОВАННЫХ ИМЕНОВАННЫХ ОБЪЕКТОВ В КАЗАХСКОМ ЯЗЫКЕ С ПОМОЩЬЮ МОДЕЛЕЙ CRF И RANDOM FOREST: СПРАВНИТЕЛЬНОЕ ИССЛЕДОВАНИЕ

**Абдикалык Гульназым** — магистрант Международного университета Астаны. 010000. Астана, Казахстан

E-mail: [gulnazymabdikalik@gmail.ru](mailto:gulnazymabdikalik@gmail.ru). ORCID ID: <https://orcid.org/0009-0008-5216-0707>;

**Муканова Асель** — PhD, доцент Международного университета Астаны. 010000. Астана, Казахстан

E-mail: [asiserikovna@gmail.com](mailto:asiserikovna@gmail.com). ORCID ID: <https://orcid.org/0000-0002-8964-3891>;

**Назырова Айжан Есболовна** — старший преподаватель Международного университета Астаны. 010000. Астана, Казахстан

E-mail: [ayzhan.nazyrova@gmail.com](mailto:ayzhan.nazyrova@gmail.com). ORCID ID: <https://orcid.org/0000-0002-9162-6791>.

**Аннотация.** Деятельность по распознаванию и категоризации именованных объектов в тексте называется распознаванием именованных объектов (NER) и имеет большое значение для обработки естественного языка. Однако из-за сложности морфологии и скудности лингвистических ресурсов распознавание именованных объектов в языках с ограниченными ресурсами, таких как казахский, сопряжено с определенными трудностями. В данной научной работе проведено сравнение эффективности двух известных моделей машинного обучения - условных случайных полей (CRF) и случайного леса — для NER на казахском языке. В работе рассматриваются стратегии подбора признаков с учетом морфологической сложности казахского языка и используется эталонный набор данных, созданный специально для NER казахского языка. В то время как модели Random Forest управляют высокоразмерными пространствами признаков и сложными взаимодействиями в данных, модели CRF отражают последовательные зависимости и контекстную информацию. Эффективность моделей CRF и Random Forest для казахского NER подтверждается экспериментальными результатами. Однако на эффективность работы этих моделей влияет нехватка помеченных данных. Будущие направления исследований включают расширение аннотированных наборов данных за счет сотрудничества с лингвистами и носителями языка для решения этой проблемы. В исследовании также подчеркивается, насколько важно учитывать в NER сложную морфологию казахского языка. Среди морфологических качеств, которые учитывают модели CRF и Random Forest, можно выделить наличие в словах стеблей и тегов частей речи, что улучшает распознавание именованных Именованных объектов в различных склонениях и вариантах. Сравнительный анализ проливает свет на преимущества и недостатки моделей Random Forest и CRF для казахстанской NER. В то время как модели Random Forest могут управлять сложными связями и взаимодействием признаков, модели CRF лучше справляются с выявлением последовательных зависимостей и использованием контекстной информации. Специфические потребности и характеристики задачи NER определяют выбор модели. В заключение следует отметить, что, предлагая информацию об эффективности моделей CRF и Random Forest, данное сравнительное исследование способствует развитию NER для казахского языка. Оно показывает ценность работы с морфологической сложностью, необходимость аннотированных данных и направляет дальнейшие исследования, направленные на совершенствование систем NER казахского языка.

**Ключевые слова:** Распознавание именованных объектов (NER), казахский язык, условные случайные поля (CRF), случайный лес, сравнительное исследование.

*Эта работа была финансово поддержана Комитетом науки Министерства науки и высшего образования Республики Казахстан (грант AP19577922).*

## Introduction

The task of identifying and categorizing named entities, such as names of people, companies, places, and times in text that is unorganized, is known as named entity recognition (NER) (Sang et al., 2003). It is important in many applications, such as text mining (Cheng et al., 2020), information retrieval, and question-answering (Aliod et al., 2006). Despite the fact that NER has been thoroughly explored for major languages, the lack of linguistic resources and the intricacy of morphology make extracting named entities from under-resourced languages particularly difficult.

The Kazakh language, which is a Turkic language with limited resources and is mostly spoken in Kazakhstan and its nearby territories (Eryiğit et al., 2013), is the subject of this research article. The Kazakh language possesses a wide range of morphological traits, such as intricate inflections, agglutination, and different grammatical constructions. Given that named entities can differ greatly in form and context, these linguistic complexities make it difficult to effectively identify and categorize them.

This study examines the performance of Conditional Random Fields (CRF) and Random Forest, two well-known machine learning models, to address the NER difficulties in the Kazakh language. A sequence labeling approach called CRF uses contextual data to accurately identify entities while also capturing sequential interdependence. The Random Forest ensemble learning method, on the other hand, mixes various decision trees to generate predictions, managing high-dimensional feature spaces and capturing complicated relationships in the data.

The objective of the comparison study is to assess the effectiveness of Random Forest and CRF models for NER in Kazakh. Comparing the models is part of the study, which also covers other NER-related topics including entity detection. We evaluate the precision of named entity extraction from Kazakh texts using these models on a benchmark Kazakh NER dataset.

The results of this study, which specifically focused on the Kazakh language, help us comprehend NER in languages with little resources. They provided information on the effectiveness and applicability of CRF and Random Forest models for NER tasks in Kazakh, as well as their advantages and disadvantages. The findings will aid researchers and practitioners in selecting suitable models for NER in the Kazakh language and provide a foundation for further improvements in NER methods for languages with limited resources.

In the parts that follow, we will go over the methodology for the comparative analysis as well as a summary of similar work in NER for languages with limited resources. Following a presentation and analysis of the experimental findings, we will talk about the implications and potential future developments for NER in Kazakh.

*Ner challenges in Kazakh language.* The activity of recognizing and categorizing named items in text is known as named entity recognition (NER), and it is essential to natural language processing. However, due primarily to the language's

morphological complexity, extracting named items from under-resourced languages (Yeniterzi, 2011) like Kazakh presents special difficulties. We go into the special difficulties posed by morphological complexity in NER for the Kazakh language in this section.

The Kazakh language possesses a wide variety of grammatical features, such as intricate inflections, agglutination, and rich morphological traits (Tolegen et al., 2019). The morphological intricacy that makes accurate NER challenging is exacerbated by these linguistic traits (Yergesh et al., 2016). Named Entity Recognition (NER), also known as named entity identification and classification in text, is a crucial problem in natural language processing (Nadeau et al., 2007). The scarcity of annotated data, however, is one of the biggest obstacles to NER for the Kazakh language.

#### 1. Complex Irregularities

The inflection system used in Kazakh is intricate and includes different verb conjugations, noun declensions, and adjectival agreements. Each entity has multiple word forms as a result of these inflections, necessitating the use of NER algorithms to take these various inflected versions into consideration (Iskhakova et al., 2020).

#### 2. Agglutination

In order to express grammatical information, affixes are frequently appended to the root word in Kazakh, a process known as agglutination. The extensive and complex word forms that result from this agglutinative nature make it difficult to effectively identify and extract named items (Bilakhanova et al., 2023).

#### 3. The use of language Features

A wide range of grammatical traits, such as case markers, tense markers, and person markers, are used in Kazakh. The interplay between these markers and their effects on named entity recognition must be taken into account because these traits add to the language's morphological complexity (Tolegen et al., 2023).

#### 4. Lack of Annotated Data

The availability of high-quality Annotated Corpora is crucial for the creation of accurate and reliable NER models. Nevertheless, finding enough annotated data for languages with limited resources, like Kazakh, is a considerable difficulty (Bogdanchikov et al., 2022). The availability of Kazakh-specific annotated datasets frequently restricts the range for developing and testing NER models.

*Conditional random fields for kazakh ner.* The activity of recognizing and categorizing named items in text is known as named entity recognition (NER), and it is essential to natural language processing. For sequence labeling tasks like NER, Conditional Random Fields (CRF) is a well-liked and efficient machine learning model that is frequently utilized.

A probabilistic graphical model called Conditional Random Fields is employed for applications involving sequence labeling. It is a discriminative model that takes into account the relationships between adjacent labels in a sequence, so capturing the contextual data required for precise labeling (Lample et al., 2016). For NER tasks, where the types and limits of named entities frequently depend on the context of the words around them, CRF models are particularly well suited.

*Random forest for kazakh ner.* A well-liked machine learning technique called Random Forest is renowned for its capacity to manage high-dimensional feature spaces and capture intricate relationships in the data (Belgiu et al., 2016). An ensemble learning technique called Random Forest uses several decision trees to produce predictions. It is frequently used for classification and regression problems and is a member of the family of tree-based algorithms. The way Random Forest works is by using various subsets of the training data to train a group of decision trees, then combining those predictions to get the outcome.

A set of features taken from the input data are used as the basis for the Kazakh Random Forest models for NER. Word IDs, part-of-speech tags, morphological characteristics, and contextual information from nearby words are a few examples of these features. Multiple decision trees are constructed during the training phase utilizing various feature and training data subsets. To ensure variation among the different trees, each decision tree learns to make predictions based on a random selection of features.

*Corpus structure.* As a dataset were utilized part of the train data which is open source in the website github (<https://github.com/IS2AI/KazNERD/tree/main>), training of the models was done on data in Kazakh. The dataset contained 132 phrases, 831 words, and 1489 characters in total. IOB2 tags (Sang et al., 1999) were used to label the data, as seen in Figure 1. Tagger that uses the IOB format, where chunks are labeled by their appropriate category, can be created that tags every word in a sentence.

B - for the word in the Beginning chunk

I - for words Inside the chunk

O - Outside any chunk

	Tokens	IOB2
0	Еуразиялық	B-LAW
1	экономикалық	I-LAW
2	одақ	I-LAW
3	туралы	I-LAW
4	шартқа	I-LAW

Fig. 1. Example of IOB2 scheme labeled data

Unexpectedly, the majority of words are categorized as not belonging to any block. These words can be thought of as placeholders, and the classifier's performance may be impacted by their use. Let's run a different test on the dataset devoid of O tags, you can see this from Figure 2.

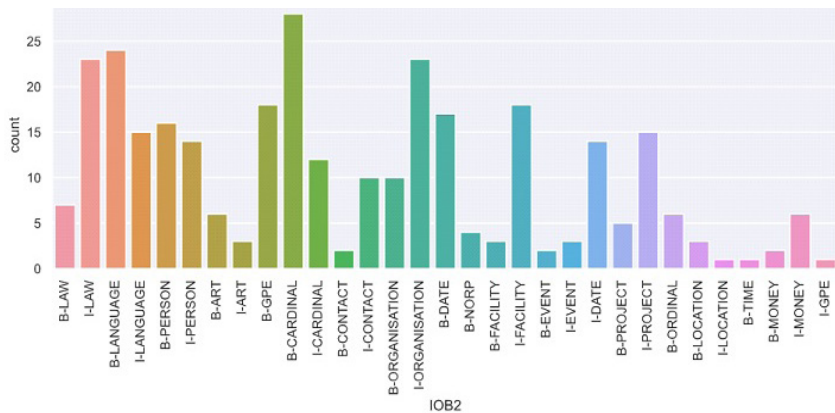


Fig. 2. Words distribution without O tag

*Modeling the data.* In this section, we propose a comparison study to assess Conditional Random Fields (CRF) and Random Forest, two well-known machine learning models, for NER in the Kazakh language. We use a benchmark dataset designed especially for Kazakh NER to conduct an extensive comparison. The dataset comprises of Kazakh-language writings that have been annotated and span a variety of topics and item kinds. In order to ensure a suitable distribution of entities throughout the sets, we separate the dataset into training, validation, and test sets.

We use feature engineering to gather pertinent data for NER in Kazakh for both CRF and Random Forest models. Word identities, part-of-speech labels, and contextual elements from nearby words are included in this. We use common assessment metrics like precision, recall, and F1-score to evaluate the effectiveness of the CRF and Random Forest models. These metrics reveal how well the models are able to recognize and categorize named things in Kazakh texts. Before we move on to the modeling portion, it is essential to understand the performance indicators that will be used to evaluate the models. We will assess the models using the precision (1), recall (2) and F1 score (3), metrics because we are dealing with information extraction.

In order to generate the metrics described above, True/False positives and True/False negatives are used, respectively.

- True Positives (TP) are successfully predicted positive values, which means that both the actual and projected classes have the same value.
- True Negatives (TN) are successfully predicted negative values, which indicates that both the actual and projected class values are negative.
- False Positives (FP) are when the expected class is present but the actual class is absent.
- False Negatives (FN) are when the expected class is no when the actual class is yes.

$$\text{Precision} = \frac{tp}{tp+fp}, \tag{1}$$

$$\text{Recall} = \frac{tp}{tp+fn}, \tag{2}$$

$$F-1 = 2 * \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \tag{3}$$

On the test set, we compute the metrics, and we contrast the outcomes of the two models. The outcomes of the performance of the RF and CRF models are also shown in the pictures below. (Shown in Figure 3.)

	precision	recall	f1-score	support		precision	recall	f1-score	support
B-ART	0.00	0.00	0.00	6	B-ART	1.00	0.33	0.50	6
B-CARDINAL	0.43	0.36	0.39	28	B-CARDINAL	0.80	1.00	0.89	28
B-CONTACT	0.00	0.00	0.00	2	B-CONTACT	0.00	0.00	0.00	2
B-DATE	0.00	0.00	0.00	17	B-DATE	1.00	1.00	1.00	17
B-EVENT	0.00	0.00	0.00	2	B-EVENT	0.00	0.00	0.00	2
B-FACILITY	0.00	0.00	0.00	3	B-FACILITY	0.00	0.00	0.00	3
B-GPE	0.20	0.22	0.21	18	B-GPE	0.75	1.00	0.86	18
B-LANGUAGE	0.00	0.00	0.00	24	B-LANGUAGE	0.96	1.00	0.98	24
B-LAW	0.00	0.00	0.00	7	B-LAW	0.67	0.29	0.40	7
B-LOCATION	0.00	0.00	0.00	3	B-LOCATION	0.00	0.00	0.00	3
B-MONEY	0.00	0.00	0.00	2	B-MONEY	0.00	0.00	0.00	2
B-NORP	0.67	0.50	0.57	4	B-NORP	0.00	0.00	0.00	4
B-ORDINAL	0.50	0.17	0.25	6	B-ORDINAL	1.00	1.00	1.00	6
B-ORGANISATION	0.00	0.00	0.00	10	B-ORGANISATION	0.77	1.00	0.87	10
B-PERSON	0.00	0.00	0.00	16	B-PERSON	0.70	1.00	0.82	16
B-PROJECT	0.00	0.00	0.00	5	B-PROJECT	1.00	0.40	0.57	5
B-TIME	0.00	0.00	0.00	1	B-TIME	0.00	0.00	0.00	1
I-ART	0.00	0.00	0.00	3	I-ART	0.00	0.00	0.00	3
I-CARDINAL	0.00	0.00	0.00	12	I-CARDINAL	1.00	0.75	0.86	12
I-CONTACT	0.00	0.00	0.00	10	I-CONTACT	1.00	1.00	1.00	10
I-DATE	0.00	0.00	0.00	14	I-DATE	1.00	1.00	1.00	14
I-EVENT	0.00	0.00	0.00	3	I-EVENT	1.00	0.33	0.50	3
I-FACILITY	0.00	0.00	0.00	18	I-FACILITY	0.86	1.00	0.92	18
I-GPE	0.00	0.00	0.00	1	I-GPE	0.00	0.00	0.00	1
I-LANGUAGE	0.00	0.00	0.00	15	I-LANGUAGE	0.00	0.00	0.00	15
I-LAW	0.00	0.00	0.00	23	I-LAW	0.79	1.00	0.88	23
I-LOCATION	0.00	0.00	0.00	1	I-LOCATION	0.00	0.00	0.00	1
I-MONEY	0.00	0.00	0.00	6	I-MONEY	1.00	1.00	1.00	6
I-ORGANISATION	0.29	0.17	0.22	23	I-ORGANISATION	0.79	1.00	0.88	23
I-PERSON	0.00	0.00	0.00	14	I-PERSON	0.93	1.00	0.97	14
I-PROJECT	0.00	0.00	0.00	15	I-PROJECT	0.65	1.00	0.79	15
O	0.82	0.99	0.90	1178	O	1.00	1.00	1.00	1178
accuracy			0.79	1490	accuracy			0.96	1490
macro avg	0.09	0.08	0.08	1490	macro avg	0.58	0.57	0.55	1490
weighted avg	0.67	0.79	0.72	1490	weighted avg	0.95	0.96	0.95	1490

Fig. 3. RF and CRF model performance

The experimental results demonstrate that both CRF and Random Forest models compete favorably for Kazakh NER. While CRF is excellent at capturing sequential dependencies, Random Forest does a better job of covering high-dimensional feature spaces.

The model performed poorly even though it had a high average score with an f-1 score of 0.79. For the majority of the classes, precision and recall levels were 0. It appears that the model lacks the characteristics required to make wise selections. The methodology just calls for memory of words and tags, which is insufficient. For the algorithm to produce more precise predictions, the context information surrounding each word must also be supplied. As the scores increased and the

f-1 score reached 0.96, the CRF classifier surpassed the Random Forest classifier. For each class, the precision and recall measurements, however, have only slightly increased. Perhaps the model is repeating words and not fully accounting for context.

*Limitations and directions for the future.* Limitations of CRF and Random Forest models include the necessity for annotated data and issues handling the complicated morphology of Kazakh. Expanding annotated datasets, analyzing deep learning models, and researching transfer learning and domain adaptation techniques in Kazakh NER should be the key objectives of future research. Kazakh's morphological complexity poses difficulties for NER since substantial changes and inflections in named entities may not be properly captured by CRF and Random Forest models. Future research should look at sophisticated feature engineering techniques and Kazakh-specific language resources to increase morphological complexity. As there are just a few annotated corpora developed specifically for Kazakh NER, the availability of annotated data is a significant limitation. Future research should concentrate on improving the usability of annotated datasets and collaborating with linguists, academics, and native speakers to solve this problem.

## **Conclusion**

The use of Conditional Random Fields (CRF) and Random Forest models for Named Entity Recognition (NER) in the Kazakh language is examined in this study. Its performance is assessed, and its advantages, disadvantages, and potential directions are covered in a comparison analysis. The results demonstrate that while Random Forest models manage high-dimensional feature spaces and intricate interactions in the data, CRF models excel at capturing sequential dependencies and contextual information. Both models successfully extract named entities from texts written in Kazakh. However, it is difficult to create reliable and accurate models because there is a lack of annotated data for training these models.

To improve NER performance in Kazakh, future research should concentrate on increasing the accessibility of annotated datasets, investigating domain adaptation strategies, and applying deep learning models. The performance of NER systems can be greatly enhanced by integrating linguistic resources and expertise particular to the Kazakh language. The advancement of NER in Kazakh depends on using unsupervised and semi-supervised learning methodologies and assessing models on various domains and text genres. Overall, the comparison study offers insightful information about the effectiveness of CRF and Random Forest models for Named Entity Recognition in Kazakh.

## **REFERENCES**

- Aliod D.M., van Zaanen M. and Smith D. (2006). Named entity recognition for question answering. In Proceedings of the Australasian Language Technology Workshop (ALTA). Pp. 51–58. ALTA. (in Eng.).
- Belgiu M. and Dragut L. (2016). Random Forest in Remote Sensing: A Review of Applications and Future Directions. ISPRS Journal of Photogrammetry and Remote Sensing, 114, 24–31. <https://doi.org/10.1016/j.isprsjprs.2016.01.011>. (in Eng.).



- Bilakhanova A., Ydyrys A. & Sultanova N. (2023). KAZAKH LANGUAGE-BASED QUESTION ANSWERING SYSTEM USING DEEP LEARNING APPROACH. *Suleyman Demirel University Bulletin: Natural And Technical Sciences*, 62(1), 113–121. doi:10.47344/sdubnts.v62i1.974. (in Eng.).
- Bogdanchikov A., Ayazbayev D., Varlamis I. (2022). Classification of Scientific Documents in the Kazakh Language Using Deep Neural Networks and a Fusion of Images and Text. *Big Data Cogn. Comput.* 2022, 6, 123. <https://doi.org/10.3390/bdcc6040123>. (in Eng.).
- Cheng P. and Erk K. (2020). Attending to entities for better text understanding. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 34. Pp. 7554–7561. (in Eng.).
- Eryiğit G., Çetin F., Yanık M., Temel T., Çiçekli I. (2013). TURKSENT: a sentiment annotation tool for social media. In: *Proceedings of the 7th Linguistic Annotation Workshop & Interoperability with Discourse, ACL 2013, Sofia, Bulgaria*. (in Eng.).
- Iskhakova G.R., Kirillova Z.N. & Yerbulatova I.K. (2020). Paired Combinations in Kazakh Language and its Ways of Translation into Russian. *Utopía Y Praxis Latinoamericana*, 25(1). Pp. 396-401. (in Eng.).
- Lample G., Ballesteros M., Subramanian S., Kawakami K. and Dyer C. (2016). "Neural architectures for named entity recognition", arXiv vol. 1603.01360: 1–1. (in Eng.).
- Nadeau D., Sekine S. (2007). A survey of named entity recognition and classification. In: *Lingvisticae Investigationes*. (in Eng.).
- Sang E.F.T.K. and Meulder F.D. (2003). Introduction to the CoNLL-2003 shared task: Language-independent named entity recognition. In *Proceedings of the Conference on Natural Language Learning (CoNLL) at HLT-NAACL*. Pp. 142–147. ACL. (in Eng.).
- Sang E.F.T.K. and Veenstra J. (1999). Representing text chunks. In *Proceedings of the European Chapter of the Association for Computational Linguistics (EACL)*. Pp. 173–179. ACL. (in Eng.).
- Tolegen G., Toleu A., Mamyrbayev O., Mussabayev R. (2023). Neural Named Entity Recognition for Kazakh. In: Gelbukh, A. (eds) *Computational Linguistics and Intelligent Text Processing. CICLing 2019. Lecture Notes in Computer Science*. Vol. 13452. Springer, Cham. [https://doi.org/10.1007/978-3-031-24340-0\\_1](https://doi.org/10.1007/978-3-031-24340-0_1). (in Eng.).
- Tolegen G., Toleu A., Mamyrbayev O. and Mussabayev R. (2019). "Named Entity Recognition for Kazakh Using Conditional Random Fields". *CICLing2019: Springer Lecture Notes in Computer Science (La Rochelle, France, 2019)*. (in Eng.).
- Yeniterzi Reyyan (2011). Exploiting Morphology in Turkish Named Entity Recognition System. In *Proceedings of the ACL 2011 Student Session*. Pp. 105–110. (in Eng.).
- Yergesh B., Sharipbay A., Bekmanova G., Lipnitskii S. (2016). Sentiment analysis of Kazakh phrases based on morphological rules. *J. Kyrgyz State Tech. Univ. Named After I. Razzakov. Theor. Appl. Sci. Tech. J.* 2(38). Pp. 39–42. (in Eng.).

NEWS OF THE NATIONAL ACADEMY OF SCIENCES OF THE REPUBLIC OF  
KAZAKHSTAN  
PHYSICO-MATHEMATICAL SERIES  
ISSN 1991-346X  
Volume 3. Number 347 (2023). 18–27  
<https://doi.org/10.32014/2023.2518-1726.201>

УДК 004.931

© **G.B. Abdikerimova<sup>1</sup>, M.B. Yessenova<sup>1\*</sup>, T.T. Ospanova<sup>1</sup>, U.Zh Aitimova<sup>2</sup>,  
M. Aitimov<sup>2</sup>, 2023**

<sup>1</sup>Eurasian National University named after L.N. Gumilyov, Astana, Kazakhstan;  
<sup>2</sup>Kazakh Agrotechnical University named after S. Seifullin, Astana, Kazakhstan.  
E-mail: moldir\_11.92@mail.ru

## **USE OF INFORMATION TEXTURE LAWS MASK METHODS IN SPACE IMAGE PROCESSING**

**Yessenova Moldir** — Doctoral student of the Department of Information Systems of the Eurasian National University named after L.N. Gumilyov, Astana, Kazakhstan, st. Satpaeva, 2, 010000  
E-mail: moldir\_11.92@mail.ru, <https://orcid.org/0000-0002-2644-0966>;

**Abdikerimova Gulzira** — Associate Professor, Department of Information Systems, Eurasian National University named after L.N. Gumilyov, PhD, Astana, Kazakhstan, st. Satpaeva, 2, 010000  
E-mail: gulzira1981@mail.ru, <https://orcid.org/0000-0002-4953-0737>;

**Ospanova Tileugaisha Topanbaevna** — Associate Professor of the Department of Information Systems of the Eurasian National University named after L.N. Gumilyov, candidate of technical sciences, Astana, Kazakhstan, st. Satpayeva, 2, 010000  
E-mail: tleu@mail.ru, <https://orcid.org/0000-0002-1729-1321>;

**Aitimova Ulzada Zholdasbekovna** — Associate Professor of the Department of Information Systems of the Kazakh Agrotechnical Research University named after S. Seifullin, Candidate of Physical and Mathematical Sciences, st. Zhenis, 62, Astana, Kazakhstan, 010000  
E-mail: uaitimova@mail.ru, <https://orcid.org/0000-0002-0803-7137>;

**Murat Aitimov** — Director of the Kyzylorda Regional Branch of the Academy of Public Administration under the President of the Republic of Kazakhstan, PhD, Kyzylorda, Kazakhstan  
E-mail: aitimovmurat07@gmail.com, <https://orcid.org/0000-0002-8397-8914>.

**Abstract.** Image processing systems are currently used to solve many applied problems. The article conducted a study using a program implemented in the Matlab software environment to determine the factors that negatively affect the growth of wheat in different growing seasons according to aerospace images. The program is based on the Laws texture mask method and clustering. This paper presents the algorithm of the program and shows the results of image processing by highlighting the homogeneity of the image. To solve the problem, the spectral luminance coefficient (SBC), NDVI, Loves mask method and clustering are used.

This method is universal and has great potential for determining the boundaries of objects and areas with different properties on aerospace images using clustering based on object images obtained in different vegetation periods. That is, in the future, the possibility of using the Laws texture mask method for analyzing experimental data is being studied in order to identify homogeneous areas on aerospace images that may be associated with foci of weeds and pests on wheat.

**Key words:** Law's textural masks, NDVI, clustering, image processing, satellite images, orthogonal transformation

**Conflict of interest:** The authors declare that there is no conflict of interest.

© Г.Б. Абдикеримова<sup>1</sup>, М.Б. Есенова<sup>1\*</sup>, Т.Т. Оспанова<sup>1</sup>, У.Ж. Айтимова<sup>2</sup>,  
М. Айтимов<sup>2</sup>, 2023

<sup>1</sup>Л.Н. Гумилев атындағы Еуразия ұлттық университеті, Астана, Қазақстан;

<sup>2</sup>С. Сейфуллин атындағы Қазақ агротехникалық университеті,  
Астана, Қазақстан.

E-mail: moldir\_11.92@mail.ru

## ҒАРЫШТЫҚ КЕСКІНДЕРДІ ӨНДЕУДЕ АҚПАРАТТЫҚ ТЕКСТУРАЛЫҚ ЛАВС МАСКАЛАР ӘДІСТЕРІН ҚОЛДАНУ

**Есенова Молдир Балканровна** — Л.Н. Гумилев атындағы Еуразия ұлттық университетінің Ақпараттық жүйелер кафедрасының докторанты, Астана, Қазақстан, Сатпаев к., 2, 010000  
E-mail: moldir\_11.92@mail.ru, <https://orcid.org/0000-0002-2644-0966>;

**Абдикеримова Гульзира Бахытбековна** — Л.Н. Гумилев атындағы Еуразия ұлттық университетінің Ақпараттық жүйелер кафедрасының м.а. доценты, PhD, Астана, Қазақстан, Сатпаев к., 2, 010000  
E-mail: gulzira1981@mail.ru, <https://orcid.org/0000-0002-4953-0737>;

**Оспанова Тилеугайша Топанбаевна** — Л.Н. Гумилев атындағы Еуразия ұлттық университетінің Ақпараттық жүйелер кафедрасының м.а. доценты, т. ғ. к., Астана, Қазақстан, Сатпаев к., 2, 010000  
E-mail: tleu@mail.ru, <https://orcid.org/0000-0002-1729-1321>;

**Айтимова Улдада Жолдасбековна** — С. Сейфуллин атындағы Қазақ агротехникалық зерттеу университеті ақпараттық жүйелер кафедрасында, м.а. қауымдастырылған профессор, ф.-м.ғ.к., Астана, Қазақстан, Женіс к., 62, 010000  
E-mail: uaitimova@mail.ru, <https://orcid.org/0000-0002-0803-7137>;

**Айтимов Мурат** — Қазақстан Республикасы Президентінің жанындағы Мемлекеттік басқару академиясының Қызылорда облыстық филиалының директоры, PhD, Қызылорда, Қазақстан  
E-mail: aitimovmurat07@gmail.com, <https://orcid.org/0000-0002-8397-8914>.

**Аннотация.** Кескінді өңдеу жүйелері қазіргі уақытта көптеген қолданбалы есептерді шешу үшін қолданылады. Мақалада аэроғарыштық суреттерде әртүрлі вегетациялық кезеңдерде бидайдың өсуіне теріс әсер ететін факторларды анықтау үшін Matlab бағдарламалық ортасында енгізілген бағдарламаны қолдану арқылы зерттеу жүргізілді. Бағдарлама

Лавс текстуралық маска әдісіне және кластерлеуге негізделген. Бұл жұмыста бағдарламаның алгоритмі берілген және кескіннің біртектілігін ерекшелеу арқылы кескінді өңдеу нәтижелері көрсетілген. Мәселені шешу үшін спектрлік жарықтандыру коэффициенті (SBC)NDVI, Лавс маскалар әдісі және кластерлеу қолданылады. Бұл әдіс әмбебап болып табылады және әртүрлі вегетация кезеңдерінде алынған объектінің суреттері негізінде кластерлеуді пайдалана отырып, объектілер мен аэроғарыштық кескінде әртүрлі қасиеттері бар аймақтардың шекараларын анықтау үшін үлкен мүмкіндікке ие. Яғни, келешекте бидайдағы арамшөптер, зиянкестердің ошақтарымен байланысты болуы мүмкін аэроғарыштық суреттердегі біртекті аймақтарды анықтау мақсатында тәжірибелік деректерді талдау үшін Лавс текстуралық маскалар әдісін қолдану мүмкіндігі зерттелуде.

**Түйін сөздер:** Лавс текстуралық маскалары, NDVI, кластерлеу, кескінді өңдеу, спутниктік кескіндер, ортогоналды түрлендіру

**Мүдделер қақтығысы:** Авторлар осы мақалада мүдделер қақтығысы жоқ деп мәлімдемейді.

**©Г.Б. Абдикеримова<sup>1</sup>, М.Б. Есенова<sup>1\*</sup>, Т.Т. Оспанова<sup>1</sup>, У.Ж. Айтимова<sup>2</sup>,  
М. Айтимов<sup>2</sup>, 2023**

<sup>1</sup>Евразийский национальный университет имени Л.Н. Гумилева,  
Астана, Казахстан;

<sup>2</sup>Казахский агротехнический университет имени С. Сейфуллина,  
Астана, Казахстан.

E-mail: moldir\_11.92@mail.ru

## **ИСПОЛЬЗОВАНИЕ МЕТОДОВ ИНФОРМАТИВНОЙ ТЕКСТУРНОЙ МАСОК ЛАВСА ПРИ ОБРАБОТКЕ КОСМИЧЕСКИХ ИЗОБРАЖЕНИЙ**

**Есенова Молдир Балкаировна** — докторант кафедры информационных систем Евразийского национального университета имени Л.Н. Гумилева, Астана, Казахстан, ул. Сагпаева, 2, 010000

E-mail: moldir\_11.92@mail.ru, <https://orcid.org/0000-0002-2644-0966>;

**Абдикеримова Гульзира Бахытбековна** — и.о. доцента кафедры информационных систем Евразийского национального университета имени Л.Н. Гумилева, PhD, Астана, Казахстан, ул. Сагпаева, 2, 010000

E-mail: gulzira1981@mail.ru, <https://orcid.org/0000-0002-4953-0737>;

**Оспанова Тилеугайша Топанбаевна** — и.о. доцента кафедры информационных систем Евразийского национального университета имени Л.Н. Гумилева, к.т.н., Астана, Казахстан, ул. Сагпаева, 2, 010000

E-mail: tleu@mail.ru, <https://orcid.org/0000-0002-1729-1321>;

**Айтимова Улзада Жолдасбековна** — и.о. ассоциированного профессора кафедры информационных систем Казахский агротехнический исследовательский университет имени С.Сейфуллина, к.ф.-м.н., ул. Жениса, 62, г. Астана, Казахстан, 010000

E-mail: uaitimova@mail.ru, <https://orcid.org/0000-0002-0803-7137>;

**Айтимов Мурат** — директор Кызылординского областного филиала Академии государственного управления при Президенте Республики Казахстан, PhD, Кызылорда, Казахстан  
E-mail: aitimovmurat07@gmail.com, <https://orcid.org/0000-0002-8397-8914>.

**Аннотация.** Системы обработки изображений в настоящее время используются для решения многих прикладных задач. В статье проведено исследование с помощью программы, реализованной в программной среде Matlab. Программа определяет факторы, негативно влияющие на рост пшеницы в разные вегетационные периоды по аэрокосмическим снимкам. Предлагаемая программа основана на методе текстурной маски Лавса и кластеризации. В данной работе представлен алгоритм работы программы и показаны результаты обработки изображения путем выделения однородности изображения. Для решения задачи используются спектральный коэффициент яркости (SBC), NDVI, метод маски Лавса и кластеризация. Этот метод универсален и имеет большие возможности для определения границ объектов и областей с различными свойствами на аэрокосмических снимках с использованием кластеризации на основе изображений объекта, полученных в разные вегетационные периоды. То есть в перспективе изучается возможность использования метода текстурных масок Лавса для анализа экспериментальных данных с целью выявления однородных участков на аэрокосмических снимках, которые могут быть связаны с очагами роста сорняков и вредителей на пшенице.

**Ключевые слова:** текстурные маски Лавса, NDVI, кластеризация, обработка изображений, космические снимки, ортогональное преобразование.

**Конфликт интересов:** авторы заявляют об отсутствии конфликта интересов.

## Кіріспе

Қазақстан Республикасының агроөнеркәсіптік кешенін дамытудың мемлекеттік бағдарламасы аясында айқындалған міндеттердің біріне байланысты мемлекеттік қызмет көрсету сапасын арттыру және агроөнеркәсіптік кешенге цифрлық технологияларды енгізуді қамтамасыз ету болып табылады. Қазіргі уақытта әртүрлі көздерден алынған кескіндерді кластерлерге бөлетін алгоритмдерді құру мәселесі жиі қолданылатын салаларда туындайды, мысалы, жер бетінің фотосуреттерін талдау, кескіндердегі ақаулар мен жарықтарды анықтау және т.б. текстураның ерекшеліктерін зерттеу, олардан дақылдардың сорттарын ажырату және олардың өсуіне теріс әсер ететін факторларды анықтау сияқты ақпаратты алу, деректерді цифрландыру өзекті тақырып болып қала береді (Омарханова, 2022).

А.И. Бараев атындағы астық шаруашылығы ғылыми-өндірістік орталығындағы дақылдардың вегетациялық кезеңіндегі өсу ерекшеліктері мен олардың өсуіне кері әсер ететін факторлар осы зерттеуге негіз болып табылады. Бұл мақалада Matlab бағдарламалық ортасында іске асырылған бағдарламаны пайдаланып, аэроғарыштық кескіндерден арамшөп ошақтарын анықтау әдісі

қарастырылады. Суреттердің текстуралық қасиеттерін анықтаудың негізгі әдістері ретінде бірінші және екінші реттік статистикалық сипаттамаларын, Лавс текстуралық маскалар әдісін, ортогональды түрлендіру әдістерін ажыратуға болады. Бұл жұмыста эксперимент жүргізу барысында Лавс текстуралық маскалар әдісін қолдану кезінде әрбір пиксель үшін кескінді сегменттеуді жүзеге асыруға болатын атрибуттар жиынтығын анықтайды және бұл әдіс есептеу уақыты тұрғысынан да кейбір артықшылықтарға ие, атап айтқанда, екінші ретті статистикалық сипаттамалар әдісімен және құрылымдық сипаттамалармен салыстырғанда (Шарма, 2013). Әртүрлі кескіндер үшін алынған реттілік/тәртіпсіздік қатынасы, «аномальды» текстурасы бар аймақтардың үлесі және т.б. сияқты текстуралық белгілердің сипаттамалары, сондай-ақ әртүрлі спектрлік коэффициенттер әрқайсысымен қосымша корреляциялануы мүмкін деп болжанады. Алынған нәтижелер ғылыми-зерттеу жұмыстарында, атап айтқанда, А.И. Бараев атындағы астық шаруашылығы ғылыми-өндірістік орталығында талдау арқылы бидайдың өсуіне теріс әсер ететін факторларды бақылау барысында пайдаланылады. Атап айтқанда, арамшөптерден зардап шеккен аймақтарда химиялық тыңайтқыштарды қолдану шығынды азайтуға, ал кейбір жағдайларда топырақ жамылғысының табиғи құнарлылығын сақтауға мүмкіндік береді (Романюк, 2014).

Келешекте машиналық оқытуда аталған әдістердің ең тиімдісін пайдалана отырып, әртүрлі ресурстарды оңтайландыру, өнімділікті арттыру және қоршаған ортаға зиян келтірместен жоғары өнімділікке теріс әсер ететін факторларды жою бойынша әрекеттер жүйесін әзірлейтін автоматтандырылған қосымша әзірленетін болады. Қарастырылып отырған мәселені неғұрлым толық сипаттау үшін бірқатар авторлардың еңбектері зерттелді.

#### **Әдістер мен материалдар.**

Бұл жұмыста (Ган, 2017) авторлар арамшөптерді идентификациялауда қолмен жобалау ерекшеліктеріне негізделген белгілерді алуда тұрақсыз сәйкестендіру нәтижелері мен нашар жалпылау мәселесін шешті. Соя өскіндері мен олармен байланысты арамшөптер зерттеу нысаны ретінде қарастырылып, конволюционды нейрондық желімен біріктірілген K-means әдісін зерттеу негізінде арамшөптерді идентификациялау моделі құрылды.

Бұл мақалада (Джин, 2021) авторлар көкөністерді анықтау және олардың айналасында шектейтін қораптарды салу үшін терең оқыту мен кескінді өңдеу технологиясын біріктіретін жаңа әдісті ұсынады. Кейіннен шектейтін жәшіктерден түсіп қалған жасыл заттар арамшөп ретінде өңделді. Фоннан арамшөптерді оқшаулау үшін кескінді өңдеу арқылы түс индексіне негізделген сегменттеу орындалды. Пайдаланылған түс индексі Байес классификациясының қатесіне сәйкес генетикалық алгоритмдер (GAS) арқылы анықталды және бағаланды.

Бұл мақалада (Петейнатос, 2020) авторлар *Zea mays*, *Helianthus annuus*, *Solanum tuberosum*, *Alopecurus myosuroides*, *Amaranthus retroflexus*, *Avena*

fatua, Chenopodium альбомы, Lamium purpureum<sup>22</sup>, Matricaria chamo-milla, Setariaump және Setariaump сияқты RGB камерасымен жиналған суреттерді қарастырды. Stellaria медиасы және конволюционды нейрондық желілерді (CNNs) оқыту үшін берілген. Үш түрлі CNN, атап айтқанда VGG16, ResNet-50 және Xception 93 000 сурет бейімделіп, оқытыған. Жаттығу суреттері әр суретте тек бір түрді қамтитын өсімдік материалының кескіндерінен тұрды. Авторлар алгоритмдер мен жүйелердің және жасанды нейрондық желілердің (ANNS) тану жылдамдығына, сенімділігіне және дәлдігіне қатысты мәселелерді жақсарту әдісін ұсынды. Кескінді сынау өсімдіктерді және арамшөп түрлерін анықтауда 77 %-дан 98 %-ға дейінгі аралықта Топ-1 дәлдігін берді.

Бұл зерттеуде (Сабзи, 2018) авторлар жалпақ жапырақты арамшөптердің тығыздығын анықтап, бидай алқаптарында гербицидтерді қолдануды азайтуға үлес қосты. Осы мақсатта авторлар кескіндерді өңдеу әдістерін қолданды, сонымен қатар арамшөптерді анықтау үшін жасанды нейрондық желілерді (ANN) және регрессия модельдерін әзірледі.

Бұл жұмыста (Агин, 2015) мөлдір электронды микроскоппен алынған кескіндердің әртүрлі ортонормальдық негіздерінде спектрлік түрлендіру қолданылды. Технологиялық процестердің тиімділігі мен реактивтілігі, кеуектілігі, диффузиялық коэффициенті және т.б. микрофотография арқылы өсімдік жасушасының қабырғаларының бұзылу дәрежесін анықтаудың тиімділігін көрсетті. Бұл жұмыста ортогональды түрлендірулерге негізделген әдістер электронды микроскоптағы кескіндерде орындалды, ал аэроғарыштық кескіндердің тиімділігі көрсетілмеді.

Бұл жұмыстың ерекшелігі - аэроғарыштық суреттердегі объектілерді текстуралық белгілер бойынша тану. Яғни, аэроғарыштық кескіндердегі біртекті аймақтарды анықтау үшін эксперименттік мәліметтерді талдау үшін текстуралық белгілердің жиынтығын қолдану мүмкіндігі туралы мәселе зерттелуде.

#### **Нәтижелер және оларды талқылау**

Бұл зерттеу жұмысы 01-012-025-040:42 кадастрлық нөмірі бойынша А.И. Бараев атындағы ғылыми-зерттеу институтына тиесілі жер учаскесінде жүргізілді (Абдикеримова, 2019). Әрбір вегетациялық кезең үшін алынған аэроғарыштық суреттер туралы ақпарат осы ғылыми орталық мамандарының мәліметтерімен анықталды. Зерттеу барысында қолданылған әдістердің бірі Лавс текстуралық маскасы болды. Аэроғарыштық суреттерде әртүрлі текстуралардың болуына байланысты (Хасан, 2019) Лавс жасаған текстуралық маскалар олардың құрылымдық айырмашылығын көрсетті. Бұл әдістің негізгі идеясы - бекітілген өлшемді терезедегі текстура мазмұнының өзгеруін бағалау. Энергетикалық сипаттамаларды есептеу үшін 5×5 өлшемді 25 маска жиынтығы қолданылады. Лавс маскаларын құрастыру кезінде 5 вектор қолданылады (1):.

$L5 = [1 \ 4 \ 6 \ 4 \ 1]$  - Деңгей – орташа сұр деңгей

$E5 = [-1 \ -2 \ 0 \ 2 \ 1]$  - Жиек - шеткі мүмкіндіктерді шығарып алу

$$S5 = [-1 \ 0 \ 2 \ 0 \ 2 \ -1] - \text{Дақ} - \text{дақтарды алу} \quad (1)$$

$$R5 = [1 \ -4 \ 6 \ -4 \ 1] - \text{Толқын} - \text{толқындарды алу}$$

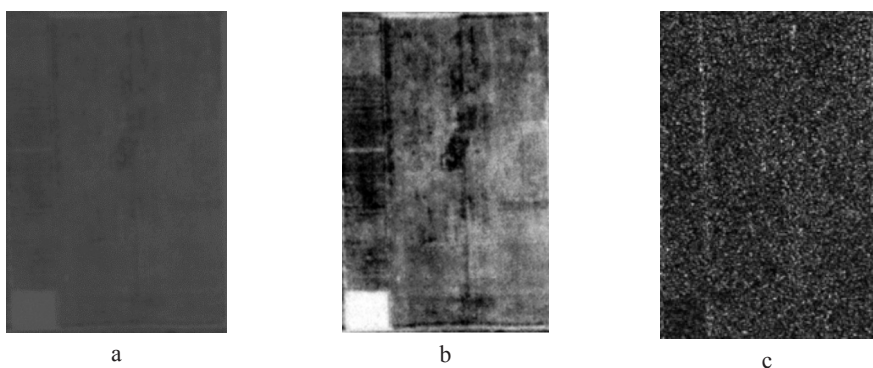
$$W5 = [-1 \ 2 \ 0 \ -2 \ 1] - \text{Толқын} - \text{толқын ерекшеліктерін алу}$$

Келесі қадам - маскалардың симметриялық жұптарын біріктіретін орташалау процедурасы. Енді осы векторларды бір-бірімен көбейтіп, біз жиырма бес түрлі маска алдық (1-кесте).

Кесте 1. Маскаларды симметриялы жұптау

Мүмкін болатын $5 \times 5$ Лавс маскалары				
L5L5	E5L5	S5L5	W5L5	R5L5
L5E5	E5E5	S5E5	W5E5	R5E5
L5S5	E5S5	S5S5	W5S5	R5S5
L5W5	E5W5	S5W5	W5W5	R5W5
L5R5	E5R5	S5R5	W5R5	R5R5

Текстуралық маска — текстура белгілерін алудың дәстүрлі әдісі, оның негізгі ерекшелігі маскалардың бес түрі, атап айтқанда деңгейлер, жиектер, дақтар, толқындар және толқындармен кескіндерді сүзу болып табылады. Бұл маскалардың әрбір комбинациясы бірегей ақпарат береді. Бұл мақалада кескінді өңдеу қолданылды, яғни кескіндердің сапасын жақсарту және бастапқы кескінде төмен текстуралық маскаларды пайдалану. Өсімдік жамылғысының әртүрлі кезеңдеріндегі суреттерге  $5 \times 5$  өлшемдегі 25 текстуралық маска тәжірибеден өтті. Мысалы, 1 (б, с) суретте олардың кейбіреулері көрсетілген. Стандартты тәсілге сәйкес, жұмыс істейтін терезенің өлшемі  $5 \times 5$  болды және әр терезеге Лавс текстуралық маскасы қолданылды. Кескінді өңдеудің бұл қадамында біз текстуралық энергияны өлшеу (ТЕМ) кескіндері деп аталатын кескіндердің жаңа жинағын жасаймыз.

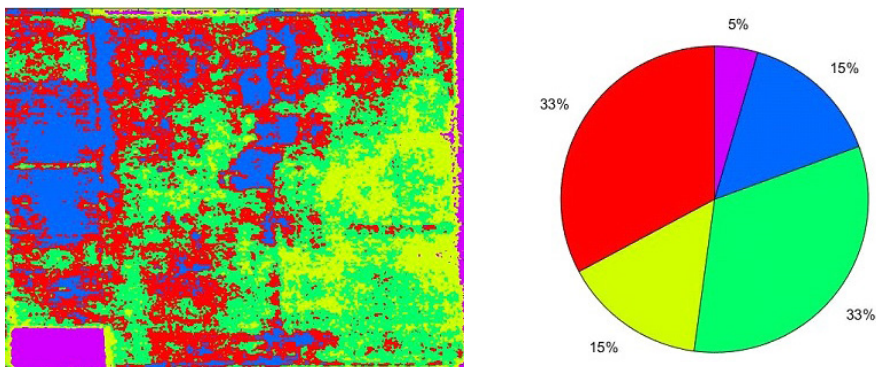


Сур. 1. Бастапқы сурет (а), L5L5 Лавс текстуралық маскасының нәтижесі (б), L5S5 Лавс текстуралық маскасының нәтижесі (с)

(Fig. 1. Original image (a), the result of L5L5 Law's textural mask (b), the result of L5S5 textural Law's mask (c))



Келесі қадам - терезелермен жұмыс істегеннен кейін алынған барлық кескіндерді кескін ретінде жақсы көрсету үшін қалыпқа келтіру керек. Барлық ықтимал 25 комбинация қолдануға рұқсат етіледі, оның ішінде 5×5 масклар 1-кестеде көрсетілген. Суретті оқып, сұр реңкке түрлендіруден кейін MATLAB функциясы кескін матрицасын және сүзгі ядросының матрицасын аргумент ретінде алып, конволюцияны өңдеу үшін қолданылады. Бұл зерттеу жұмысында суретті талдауда қолданылатын ақпараттық текстуралық мүмкіндіктерді алу үшін Лавстың 25 текстуралық маскасын қарастырдық. Нәтижесінде әрбір текстуралық маскадан алынған кескіндерде айтарлықтай айырмашылық болды, яғни кейбір кескіндер біртекті аймақтарды жақсы ажыратады, ал кейбір нәтижелерде гетерогенділік пайда болды. Төмендегі 2-суретте L5L5 текстуралық масканы қолданғаннан кейін алынған кластерлеу нәтижелері бойынша бидайдағы арамшөп ошағы айқынырақ болатынын көрсетілген.



Сур. 1. Лавс L5L5 текстуралық маскасының кластерлеу нәтижесі  
(Fig. 1. Clustering result of Law's L5L5 texture mask)

4 вегетациялық кезеңде жүргізілген тәжірибелер нәтижесінде Лавстың текстуралық маскасы арқылы текстураны талдау әдістерінің бірі жасалды. Аэроғарыштық суреттерді талдау кезінде біз әртүрлі текстуралармен айналысамыз. Текстуралық талдау әдістерін енгізу нәтижелері бойынша қылқан жапырақты немесе жапырақты ормандарды, дәнді немесе бұршақ дақылдары егілген дәнді-дақылдарды және т.б., зиянкестерден зардап шеккен ауыл шаруашылығы дақылдарын, шөлейтті аумақтарды ажыратуға болады.

### Қорытынды

Бұл жұмыста А.И. Бараев атындағы ғылыми-зерттеу институтына тиесілі 01-012-025-040:42 кадастрлық нөмірі бойынша жер учаскесін қарастырдық. 2022 жылы мұнда бидай егілді. Бидайдың өсуіне кері әсер ететін фактор арамшөптердің өсу динамикасы Лавстың текстуралық маскалары арқылы анықталды. Алынған 100 кескіннің ішінен L5L5 текстуралық маска біртекті аймақтарға арналған 25 мүмкін текстуралық маскадан анық ажыратылды.

Кластерлеу L5L5 текстуралық масканы қолдану арқылы алынған кескіндер бойынша жүргізілді. Нәтиже 9 жұмыста ортогональды түрлендіру әдістерін қолдану нәтижесімен сәйкес келді. Лавстың текстуралық маскалар әдісі есептеу уақытының жылдамдығымен ерекшеленеді. Кейінгі зерттеулерде индикаторлардың векторлары теріс факторлармен, атап айтқанда арамшөп ошақтарымен байланысты болуы мүмкін. Бағдарламалық жүйені машиналық оқытуда жиі қолданылатын басқа тәсілдер арқылы оқытуға болады. Жаттығудан кейін жүйе параметрлердің мәндерін болжай алады.

#### ӘДЕБИЕТТЕР

Абдикеримова Г.Б., Мурзин Ф.А., Бычков А.Л., Вэй Х., Рябчикова Е.И., Аязбаев Т. (2019). Ортогоналды түрлендірулер негізінде текстуралық бейнелерді талдау. Теориялық және қолданбалы ақпараттық технологиялар журналы, 97(1), 15–22.

Агин О., Танер А. (2015). Бидай өндірісіндегі арамшөптердің қарқындылығын кескінді өңдеу әдістерін қолдану арқылы анықтау. *Anadolu Tarım Bilimleri Dergisi*, 30(2), 110–117, <https://doi.org/10.7161/anajas.2015.30.2.110-117>.

Джин Х., Че Дж., Чен Ю. (2021). Көкөніс плантациясында терең оқыту және кескінді өңдеу арқылы арамшөптерді анықтау. *IEEE Access*, 9, 10940–10950.

Омарханова З., Мисник О., Матайбаева Г., Мұқашева Г., Жолдаяқова Г. & Рамазанова С. (2022). Агроөнеркәсіптік кешендегі қаржыландыру көлеміне әсер етудің экологиялық факторларының әсері. *Journal of Environmental Management & Tourism*, 13(3), 790–801, [https://doi.org/10.14505/jemt.v13.3\(59\).18](https://doi.org/10.14505/jemt.v13.3(59).18).

Петейнагос Г.Г., Рейхель П., Карута Дж., Андужар Д., Герхардс Р. (2020). Жүгері, күнбағыс және картоптағы арамшөптерді конволюциялық нейрондық желілер көмегімен анықтау. *Қашықтан зондтау*, 12(24), 4185, <https://doi.org/10.3390/rs12244185>.

Романюк Н.Н., Нүкешев С.О., Қақабаев Н.А., Агейчик В.А., Шило И.Н., Тойғамбаев С.К., Есхожин Д.З. Сеялка. (2014). Қазақстан Республикасының патенті 29217, МПК А 01В 49/04, өтінші беруші – С.Сейфуллин атындағы Қазақ агротехникалық университеті, №29217; 11.12.2013 жылы қолданылды; 15.12.2014 ж., 12 шығарылым (орыс тілінде).

Сабзи С., Аббаспур-Гиланде Ю., Гарсиа-Матеос Г. (2018). Метаэвристикалық алгоритмдерді пайдалана отырып, картоп дақылдарындағы арамшөптерді анықтауға арналған жылдам және дәл сараптамалық жүйе. *Өнеркәсіптегі компьютерлер*, 98, 80–89, <https://doi.org/10.1016/j.compind.2018.03.001>.

Тан Дж., Ван Д., Чжан З., Ол Л., Синь Дж., Ху У. (2017). Арамшөптерді идентификациялау К-құралдарына негізделген, конволюционды нейрондық желімен біріктірілген оқыту мүмкіндігі. *Ауыл шаруашылығындағы компьютерлер мен электроника*, 135, 63–70, <https://doi.org/10.1016/j.compag.2017.01.001>.

Хасан М.А., Янг М., Рашид А., Янг Г., Рейнольдс М., Ся Х., & Хэ З., (2019). Көп спектрлі UAV платформасын пайдалана отырып, астық өнімділігін болжау үшін бидайдың өсу циклі бойынша NDVI жылдам мониторингі. *Өсімдіктану*, 282, 95–103.

Шарма Н., Верма А. (2013). Спутниктік суреттегі аймақтарды анықтауға арналған текстуралық тәсілдің өнімділігін салыстыру. *Int. Дж. Есептеу. Қолданба*, 74, 10–15.

#### REFERENCES

Abdikirimova G.B., Murzin F.A., Bychkov A.L., Wei X., Ryabchikova E.I., Ayazbayev T. (2019). The analysis of textural images on the basis of orthogonal transformations. *Journal of Theoretical and Applied Information Technology*, 97(1), 15–22.

Ağın O., Taner A. (2015). Determination of weed intensity in wheat production using image processing techniques. *Anadolu Tarım Bilimleri Dergisi*, 30(2), 110–117, <https://doi.org/10.7161/anajas.2015.30.2.110-117>.

Hassan M.A., Yang M., Rasheed A., Yang G., Reynolds M., Xia X., & He Z., (2019). A rapid monitoring of NDVI across the wheat growth cycle for grain yield prediction using a multi-spectral UAV platform. *Plant science*, 282, 95–103.

Jin X., Che J., Chen Y. (2021). Weed identification using deep learning and image processing in vegetable plantation. *IEEE Access*, 9, 10940–10950.

Omarkhanova Z., Misnik O., Mataibayeva G., Mukasheva G., Zholdoyakova G. & Ramazanova S. (2022). Influence of Environmental Factors of Influence on the Volume of Financing in the Agro-Industrial Complex. *Journal of Environmental Management & Tourism*, 13(3), 790–801, [https://doi.org/10.14505/jemt.v13.3\(59\).18](https://doi.org/10.14505/jemt.v13.3(59).18).

Peteinatos G.G., Reichel P., Karouta J., Andújar D., Gerhards R. (2020). Weed identification in maize, sunflower, and potatoes with the aid of convolutional neural networks. *Remote Sensing*, 12(24), 4185, <https://doi.org/10.3390/rs12244185>.

Romanyuk N.N., Nukeshev S.O., Kakabayev N.A., Ageychik V.A., Shilo I.N., Toigambayev S.K., Eskhozhin D.Z. Seeder. (2014). Republic of Kazakhstan patent 29217, MPK A 01B 49/04, applicant – S. Seifullin Kazakh Agrotechnical University, No.29217; applied on 11.12.2013; published on 15.12.2014, Issue 12 (in Russian).

Sabzi S., Abbaspour-Gilandeh Y., Garcia-Mateos G. (2018). A fast and accurate expert system for weed identification in potato crops using metaheuristic algorithms. *Computers in Industry*, 98, 80–89, <https://doi.org/10.1016/j.compind.2018.03.001>.

Sharma N., Verma A. (2013). Performance comparison of texture-based approach for identification of regions in satellite image. *Int. J. Comput. Appl*, 74, 10–15.

Tang J., Wang D., Zhang Z., He L., Xin J.; Xu Y. (2017). Weed identification based on K-means feature learning combined with convolutional neural network. *Computers and electronics in agriculture*, 135, 63–70, <https://doi.org/10.1016/j.compag.2017.01.001>.

NEWS OF THE NATIONAL ACADEMY OF SCIENCES OF THE REPUBLIC OF  
KAZAKHSTAN  
PHYSICO-MATHEMATICAL SERIES  
ISSN 1991-346X  
Volume 3. Number 347 (2023). 28–43  
<https://doi.org/10.32014/2023.2518-1726.202>

УДК 665.733.3: 519.816

©**B. Assanova**<sup>1</sup>, **B. Orazbayev**<sup>1,2\*</sup>, **Zh. Moldasheva**<sup>1</sup>, **G. Shuitenov**<sup>3</sup>,  
**E. Dyussembina**<sup>3</sup>, 2023

<sup>1</sup>Kh. Dosmukhamedov Atyrau University, Atyrau, Kazakhstan;

<sup>2</sup>L.N. Gumilyov Eurasian National University, Astana, Kazakhstan;

<sup>3</sup>Esil University, Astana, Kazakhstan.

E-mail: batyr\_o@mail.ru,

## **METHODOLOGY FOR DEVELOPING MODELS OF INTERRELATED TECHNOLOGICAL UNITS OF A DELAYED COKING UNIT ON THE BASIS OF AVAILABLE INFORMATION OF A DIFFERENT NATURE**

**Assanova B.U.** — PhD, Dean of the Faculty of Physics, Mathematics and Information Technology, Kh. Dosmukhamedov Atyrau University, st. Students №1 Atyrau, Kazakhstan

E-mail: baha1981\_13@mail.ru, <https://orcid.org/0000-0003-1029-6266>;

**Orazbayev B.B.** — doctor of technical sciences, academician of the Engineering academy of the republic of Kazakhstan, professor of the department of System analysis and Control, L.N. Gumilyov Eurasian National University, Satpayev str.2A, Astana, Kazakhstan

E-mail: batyr\_o@mail.ru, <https://orcid.org/0000-0003-2109-6999>;

**Moldasheva Zh.Zh.** — Deputy Dean of the Faculty of Physics, Mathematics and Information Technologies of Atyrau University named after H. Dosmukhamedov

E-mail: zhadira1985@mail.ru, <https://orcid.org/0000-0002-0559-3410>;

**Shuitenov G.Zh.** — Candidate of Pedagogical Sciences, Associate Professor, Vice-Rector for Strategy and Digitization, Esil University, Zhubanov str.7, Astana, Kazakhstan

E-mail: g.shuitenov@mail.ru, <https://orcid.org/0000-0002-9905-7247>;

**Dyussembina E.** — doctoral student of the Department of System Analysis and Control, L.N. Gumilyov Eurasian National University, Satpayev str.2A, Astana, Kazakhstan

E-mail: e-m-i-k-o\_90@mail.ru, <https://orcid.org/0009-0005-7347-7218>.

**Abstract.** A technique is being developed for constructing models of interconnected units of complex technological systems characterized by a deficit and fuzziness of initial information, for example, a slow coking unit, using various types of available information. The proposed methodology is based on the development of models of interconnected aggregates of technological systems based on the available information of a different nature and the values of the criteria for choosing an effective model of each aggregate and on combining them into a single system (package) of models. The integration of the developed models of individual units into one package is carried out taking into account the flow of the

process in the technological system and the interconnections of its units. Criteria for determining the type of effective model for each unit and the process of choosing a model are carried out on the basis of an expert assessment and system analysis of the operating modes of the technological system. A general block diagram of the created methodology for developing models of aggregates of a technological system is created on the basis of available information of a different nature in conditions of deficiency and fuzziness of the initial information, as well as a block diagram for the development of deterministic, statistical, fuzzy or linguistic models of aggregates. The developed system methodology, with the availability of the necessary information, will allow developing deterministic or statistical models using traditional analytical and experimental-statistical methods, as well as fuzzy or linguistic models based on the proposed methods with fuzzy initial information. The proposed methods are intended for the synthesis of a fuzzy model of an aggregate, when its input parameters are clear, and its output parameters are fuzzy, and a linguistic model of an aggregate, when both the input and output parameters of an object are fuzzy.

**Keywords:** technological system, fuzzy model, system approach, package of models, expert evaluation

© Б.У. Асанова<sup>1,2</sup>, Б.Б. Оразбаев<sup>1,2</sup>, Ж.Ж. Молдашева<sup>1</sup>, Г.Ж. Шүйтенов<sup>3</sup>,  
Э.М. Дюсембина<sup>2</sup>, 2023

<sup>1</sup> Х. Досмухамедов атындағы Атырау университеті, Атырау, Қазақстан;

<sup>2</sup> Л.Н. Гумилев атындағы Еуразия ұлттық университеті, Астана, Қазақстан;

<sup>3</sup> Esil University, Астана, Қазақстан.

E-mail: baha1981\_13@mail.ru

## ТҮРЛІ СИПАТТАҒЫ ҚОЛ ЖЕТІМДІ АҚПАРАТТАР НЕГІЗІНДЕ БАЯУ КОКСТЕУ ҚОНДЫРҒЫСЫНЫҢ ӨЗАРА БАЙЛАНЫСҚАН ТЕХНОЛОГИЯЛЫҚ АГРЕГАТТАРЫ МОДЕЛЬДЕРІН ҚҰРУ ӘДІСТЕМЕСІ

**Асанова Б.У.** — PhD, Х.Досмухамедов атындағы Атырау университеті, физика-математика және ақпараттық технологиялар факультетінің деканы, студенттер даңғылы №1 Атырау қ., Қазақстан

E-mail: baha1981\_13@mail.ru.

**Оразбаев Б.Б.** — техника ғылымдарының докторы, Қазақстан Республикасы Инженерлік академиясының академигі, Жүйелік талдау және басқару кафедрасының профессоры, Л.Н.Гумилев атындағы Еуразия ұлттық университеті, Сәтпаев көш. 2А, Астана қ., Қазақстан

E-mail: batyr\_o@mail.ru

**Молдашева Ж.Ж.** — Х.Досмухамедов атындағы Атырау университеті, физика-математика және ақпараттық технологиялар факультетінің оқу ісі жөніндегі декан орынбасары, студенттер даңғылы №1 Атырау қ., Қазақстан

E-mail: zhadira1985@mail.ru

**Шүйтенов Г.Ж.** — педагогика ғылымдарының кандидаты, доцент, Стратегия және цифровизация бойынша проректор, Esil University, Жұбанов көш. 7, Астана қ., Қазақстан

E-mail: g.shuitenov@mail.ru.

**Дюсембина Э.М.** — Жүйелік талдау және басқару кафедрасының докторанты, Л.Н.Гумилев атындағы Еуразия ұлттық университеті, Сәтпаев көш. 2А, Астана қ., Қазақстан  
E-mail: e-m-i-k-o\_90@mail.ru.

**Аннотация.** Түрлі сипаттағы қол жетімді ақпараттарды қолдана отырып баяу кокстеу қондырғы сияқты бастапқы ақпараттың тапшылы және айқынсыдығымен сипатталатын күрделі технологиялық жүйелердің өзара байланысқан агрегаттарының модельдерін құру әдістемесі әзірленеді. Бұл әдістеме технологиялық жүйенің өзара байланысқан әр негізгі агрегаттарына қол жетімді ақпарат сипаты және оған тиімді модель таңдау критерийлері мәндері бойынша модель құруға және оларды бір жүйеге (пакетке) біріктіруге негізделеді. Жеке агрегаттардың құрылған модельдерін бір пакетке біріктіру технологиялық жүйеде жүретін процесстің өтуін және агрегаттардың өзара байланыстарын ескере отырып орындалады. Әр агрегатқа тиімді модель типін анықтау критерийлері мен модельді таңдау процесі эксперт-мамандар көмегімен жүйелік талдау арқылы жүзеге асырылады. Түрлі қол жетімді ақпараттар негізінде бастапқы ақпараттардың тапшылығы мен айқынсыздығы жағдайында жүйе агрегаттарының модельдерін құру әдістемесінің жалпы блок-схемасы мен детерминді, статистикалық және айқын емес, лингвистикалық модельдерді құру тәсілдерінің блок-схемалары келтіріледі. Әзірленетін жүйелік әдістеме қажетті ақпараттар болған дәстүрлі аналитикалық, эксперименталдық-статистикалық тәсілдерді қолдана отырып, агрегаттың детерминді немесе статистикалық модельдерін, ал бастапқы ақпараттар айқын емес болғанда, ұсынылатын тәсілдер негізінде, айқын емес немесе лингвистикалық модельдерін құруға мүмкіндік беретін болады. Ұсынылатын тәсілдер агрегаттың кірісі айқын, шығысы айқын емес болғанда оның айқын емес моделін, ал кірісі де, шығысы да айқын емес болса, агрегаттың лингвистикалық моделін құруға арналады.

**Түйін сөздер:** технологическая система, нечеткая модель, системный подход, пакет моделей, экспертная оценка

*Зерттеуді Қазақстан Республикасы Ғылым және жоғары білім министрлігінің Ғылым комитеті қаржыландырады (грант №АР19679897 «Мұнай коксын өндіру процесін басқару үшін интеллектуалдандырылған шешім қабылдау жүйесін әзірлеу»).*

© Б.У. Асанова<sup>1</sup>, Б.Б. Оразбаев<sup>1,2\*</sup>, Ж.Ж. Молдашева<sup>1</sup>, Г.Ж. Шуйтенов<sup>3</sup>,  
Э.М. Дюсембина<sup>2</sup>, 2023

<sup>1</sup> Атырауский университет им. Х. Досмухамедова, Атырау, Казахстан;

<sup>2</sup> Евразийский национальный университет им. Л.Н. Гумилева,  
Астана, Казахстан;

<sup>3</sup> Esil University, Астана, Казахстан.

E-mail: baha1981\_13@mail.ru

## МЕТОДИКА РАЗРАБОТКИ МОДЕЛЕЙ ВЗАИМОСВЯЗАННЫХ ТЕХНОЛОГИЧЕСКИХ АГРЕГАТОВ УСТАНОВКИ ЗАМЕДЛЕННОГО КОКСОВАНИЯ НА ОСНОВЕ ДОСТУПНОЙ ИНФОРМАЦИИ РАЗЛИЧНОГО ХАРАКТЕРА

**Асанова Б.У.** — PhD, декан факультета физики-математического и информационных технологий Атырауского университета имени Х. Досмухамедова, ул. Студентов №1 Атырау, Казахстан

E-mail: baha1981\_13@mail.ru.

**Оразбаев Б.Б.** — академик Инженерной академии Республики Казахстан, профессор кафедры Системного анализа и управления, Евразийский национальный университет им. Л.Н. Гумилева, ул. Сапгаева 2А, Астана, Казахстан

E-mail: batyr\_o@mail.ru

**Молдашева Ж.Ж.** — зам. декан по учебной работе факультета физики-математического и информационных технологий Атырауского университета имени Х. Досмухамедова, ул. Студентов №1 Атырау, Казахстан

E-mail: zhadira1985@mail.ru

**Шуйтенов Г.Ж.** — кандидат педагогических наук, доцент, проректор по стратегии и цифровизации, Esil University, ул. Жубанова 7, Астана, Казахстан

E-mail: g.shuitenov@mail.ru.

**Дюсембина Э.М.** — докторант кафедры Системного анализа и управления Евразийский национальный университет им. Л.Н. Гумилева, ул. Сапгаева 2А, Астана, Казахстан

E-mail: e-m-i-k-o\_.90@mail.ru.

**Аннотация.** Разрабатывается методика построения моделей взаимосвязанных агрегатов сложных технологических систем, характеризующихся дефицитом и нечеткостью исходной информации, например установка медленного коксования, с использованием доступной информации различного типа. Предлагаемая методика основывается на разработке моделей взаимосвязанных агрегатов и технологических систем, на основе доступной информации различного характера и значениях критериев выбора эффективной модели каждого агрегата и на объединении их в единую систему (пакет) моделей. Объединение разработанных моделей отдельных агрегатов в один пакет выполняется с учетом протекания процесса в технологической системе и взаимосвязей ее агрегатов. Критерии определения типа эффективной модели для каждого агрегата и процесс выбора модели осуществляются на основе экспертной оценки и системного анализа режимов работы технологической системы. Создается общая блок-схема создаваемой

методики разработки моделей агрегатов технологической системы на основе доступной информации различного характера в условиях дефицита и нечеткости исходной информации, а также блок-схемы разработки детерминированных, статистических, нечетких или лингвистических моделей агрегатов. Разрабатываемая системная методика при доступности необходимой информации позволит разрабатывать детерминированные или статистические модели с использованием традиционных аналитических и экспериментально-статистических методов, а также нечетких или лингвистических моделей на основе предлагаемых методов при нечеткости исходной информации. Предлагаемые методы предназначены для синтеза нечеткой модели агрегата, когда его входные параметры четкие, а его выходные нечеткие, и лингвистической модели агрегата, когда и входные, и выходные параметры объекта нечеткие.

**Ключевые слова:** технологическая система, нечеткая модель, системный подход, пакет моделей, экспертная оценка.

### **Кіріспе**

Кәзіргі уақытта электроникада, космостық технологияда, металлургияда және басқа салаларда қолданатын жоғары сапалы мұнай коксына еген сұраныс қарқынды артып келеді. Мұндай сапалы коксты тиімді өндіру үшін заманауи мұнай өңдеу зауыттарында (МӨЗ) өзара материалдық, ақпараттық ағындармен байланысқан түрлі агрегаттардан тұратын баяу кокстеу технологиялық қондырғылары (БКҚ) қолданылады. Бұл жұмыста зерттеу нысаны болып табылатын Атырау МӨЗ қолданыстағы БКҚ 21–10/6 мысалында өзара байланысқан технологиялық агрегаттардан тұратын күрделі жүйелердің математикалық модельдерінің кешенің (пакетін) құру әдістемесі ұсынылады. Өндірістік жағдайда мұндай күрделі нысандардың модельдерін құруда қажетті бастапқы ақпараттың тапшылығына, айқынсыздығына байланысты және олардың жұмысын жүйелік модельдеуде қиындықтар туындайды (Kafarov, 2018; Zhorov, 2015). БКҚ 21–10/6 қондырғысында өңделген ауыр мұнай өнімдерінен (гудрон, мазут) жоғары сапалы мұнай коксымен қатар түрлі жеңіл мұнай өнімдерін (газ, бензин, жеңіл газойл) өндіріледі.

БКҚ 21–10/6 қондырғысының өзара байланысқан негізгі технологиялық агрегаттарына кокстеу реакторлары, негізгі ректификациондық колонна, алғашқы және қайта қыздыру пештері жатады. Қондырғының жұмысын жүйелік модельдеу үшін осы аталған өзара байланысқан оның негізгі агрегаттарының математикалық модельдерін құрып, алынған модельдерді бір модельдер пакетіне (жүйесіне) біріктіру қажет (Kabibullin және т.б., 2023) Технологиялық қондырғылардың түрлі агрегаттарының күрделілігі, жұмыс сипаты әр түрлі болғандықтан, олардың модельдерін құруда қолжетімді түрлі сипаттағы ақпаратты қолдануға тура келеді. Сондықтан қолжетімді, түрлі сипаттағы ақпараттар негізінде өзара байланысқан агрегаттар модельдерін құрып, оларды бір модельдер пакетіне мүмкіндік беретін әдістеме әсерлеу



қазіргі кезде аса өзекте ғылыми-практикалық проблема болып табылады. Бұл зерттеудің негізгі мақсаты осы проблеманы зерттеп, шешу брoлып табылады.

Ғылыми зерттеулер мен әдебиеттерді талдау нәтижелері технологиялық жүйенің өзара байланысқан агрегаттарының модельдері пакетін ақпарат жетіспеушілігі мен айқынсыздығы жағдайында құру мәселелеріі әлі де жете зерттеліп, шешілмеген көрсетеді. Зерттеу тақырыбы бойынша талданған жұмыстарға жсалған шолудың қорытындысын қарастырайық. Көптеген агрегаттардан тұратын күрделі технологиялық нысандардың математикалық модельдерін құру сұрақтары (Mohaddecy, 2006; Sharikov және т.б., 2018), жұмыстарында зерттелген. Осы және басқа жұмыстарда мұнай өндеу мен басқа өндірістердің технологиялық жүйелерінде өтетін типтік процесстерін теориялық зерттеулер мен олардың кинетикасы негізінде детерминді модельдері келтірілілген. Мұндай типтік поцесстердің детерминді модельдері масса және энергия сақталу заңдарына негізделгендіктен әмбебап болып табылғанымен, олар тек жеке агрегаттардағы типтік процесстерді ғана сипаттайды. Сонымен қатар мұндай модельдердің кемшіліктеріне оларда агрегаттардың өзара байланыстары мен әсерлері және бастапқы ақпараттың жетіспеушілігі, айқын еместігі ескерілмейтіні жатады. Сол себептен бұл модельдерді, практикада жиі орын алытын, оларды құруға қажетті ақпараттың тапшылығы мен айқынсыздығы жағдайында құру мүмкін емес және ондай модельдер технологиялық қондырғылар жұмысын жүйелі модельдеуге жарамсыз болады.

Шетел ғалымдары Pinheiro C.C., Zhi-Wen Zhao және Coleman B. (Pinheiro, 2018; Zhi-Wen Zhao, 2018; Coleman, 2018), жетекшіліктерімен орындалған жұмыстарда мұнай өндеу мен басқа өндірістердің технологиялық жүйелері жұмысын модельдеу мен басқару, соның ішінде статистикалық және стохастикалық тәсілдеме негізінде басқару тәсілдері зерттелген. Технологиялық жүйелер модельдері негізінде оларда жүретін процесстерді туындаған жағдайларға байланысты басқару сұрақтары Поспелов Д. жұмысында жұмысында ұсынылған (Поспелов, 2018).

Алайда бұл зерттеулерде технологиялық жүйе агрегаттарының арасындағы өзара байлыныстары ескерілмеген. Сол себептен бұл жұмыстарда ұсынылған риформинг реакторының моделі оның жұмысын модельдеу, оптимизация және басқаруға жарамды болғанымен, тұтастай риформинг технологиялық жүйесінің жұмысын жүйелі модельдеу арқылы тиімді басқаруға жарамсыз. Ал статискалық, стохастикалық модельдер күрделі нысандар жұмысын модельдеуге жарамды болып табылады. Бірақ мұндай модельдерді құру үшін технологиялық жүйе жұмысын толықтай және сенімді сипаттайтын статистикалық деректер керек. Мұндай сенімді сандық деректерді жинау үшін көптеген эксперименттер, өлшеулер жүргізіп, нәтижерін дұрыс өндеу керек. Алайда кейбір маңызды пара метрлері мен көрсеткіштері сандық түрде өлшенуі өте күрделі, не мүмкін еместігін сипатталатын Атырау МӨЗ БКҚ сияқты көптеген айқынсыздықпен сипатталатын технологиялық

жүйелер үшін мұндай эксперименттер, модель құруға қажетті кейбір маңызды параметрлер өлшенбейтіндіктен, мүмкін емес, не тиімсіз болады. Өндірістік практикада мұндай айқын емес параметрлер менн көрсеткіштер сол өндірістердегі басқаратын тәжіриелі мамандар, яғни шешім қабылдаушы тұлға (ШҚТ), пәндік сала бойынша эксперттер көмегімен табиғи тілде бағаланады. Мұндай айқын емес ақпарат негізінде ШҚТ дұрыс, тиімді шешім қабылдауға қабілетті.

(Dzhambekov, 2018) жұмыстарында бастапқы ақпараттың бір бөлігінің айқынсыздығында технологиялық жүйелер агрегаттарының модельдерін құрып, олардың негізінде агрегаттардың жұмыс режимдерін басқару тәсілдемелері қарастырылған. Бұл тәсілдемелер эксперттік бағалау және айқын емес жиындар теориясы тәсілдері негізінде агрегаттардың шығыс параметрлері айқын емес, ал кіріс параметрлері айқын болғанда  $\alpha$  деңгейлі жиындар көмегімен агрегаттардың айқын емес модельдерін құруға мүмкіндік береді. Алайда мұндай тәсілдемелерде нысанның барлық кіріс, шығыс параметрлері айқын емес болғанда ол олардың лингвистикалық модельдерін құру мәселелері шешілмейді. Ал көптеген өндірістік жүйелердің кіріс және шығыс параметрлері айқынсыздықпен сипатталатындықтан олардың лингвистикалық модельдерін құру қажет болады. Сондай-ақ бұл зерттеу жұмыстарында да (Dzhambekov, 2018) алдыңғы талданған жұмыстардағыдай жұмыстарындағыдау технологиялық жүйелердің өзара байланысқан агрегаттарының модельдері пакеттерін құру сұрақтары қамтылмаған.

Қорыта айтқанда талданған және көптеген басқа жұмытарда технологиялық нысандардың математикалық модельдерін құрудың дәстүрлі тәсілдері (аналитикалық, эксперименталдық-статистикалық) және нысандардың шығыс параметрлері айқынсыз болғанда, олардың айқын емес модельдерін құру зерттелген. Ал айқынсыздықпен сипатталатын нысандардың кіріс, шығыс параметрлері айқын емес болғанда, олардың лингвистикалық модельдерін құру және технологиялық жүйелердің өзара байланысқан агрегаттарын жүйелі модельдеу үшін олардың құрылған модельдерін бір модельдер пакетіне біріктіру мәселелері әлі де шешілмеген. Бұл мәселелерді толықтай шешу үшін жүйелік талдау, эксперттік бағалау және айқын емес жиындар тәсілдеріне негізінде күрделі айқынсыздықпен сипатталатын жүйелердің жұмыс режимдерін модельдеу мен басқарудың жүйелі тәсілін құрып, қолдану қажет. Бұл жұмыс бастапқы ақпараттың тапшылығы мен айқынсызшылығы жағдайларында БҚК мысалында, технологиялық жүйелердің өзара байланысқан агрегаттарының модельдерін қол жетімді түрлі сипаттағы ақпарат негізінде құру әдістемесін әзірлеуге бағытталған.

### **Зерттеу мақсаты мен міндеттері және тәсілдері**

Жұмыстың зерттеу мақсаты айқынсыз ортада технологиялық кешендерді жүйеді модельдеу үшін, олардың өзара байланысқан агрегаттарының модельдері пакетін бастапқы ақпараттың тапшылығы мен айқынсызшылығы жағдайларында құру әдісіемесін әзірлеу болып табылады. Тұжырымдалған

мақсатқа қол жеткізу үшін келесі зерттеу міндеттері қойылып, шешілетін болады:

- технологиялық жүйелердің өзара байланысқан жеке агрегаттарының құрылуы тиімді модельдерінің типін қажетті критерийлер бойынша таңдап, оларды құру;

- жүйе агрегаттары жағдайы мен жұмысы жайындағы қол жетімді ақпараттары мен олардың айқынсыздық деңгейіне байланысты, агрегаттардың модельдерін құру әдістемесін даярлау;

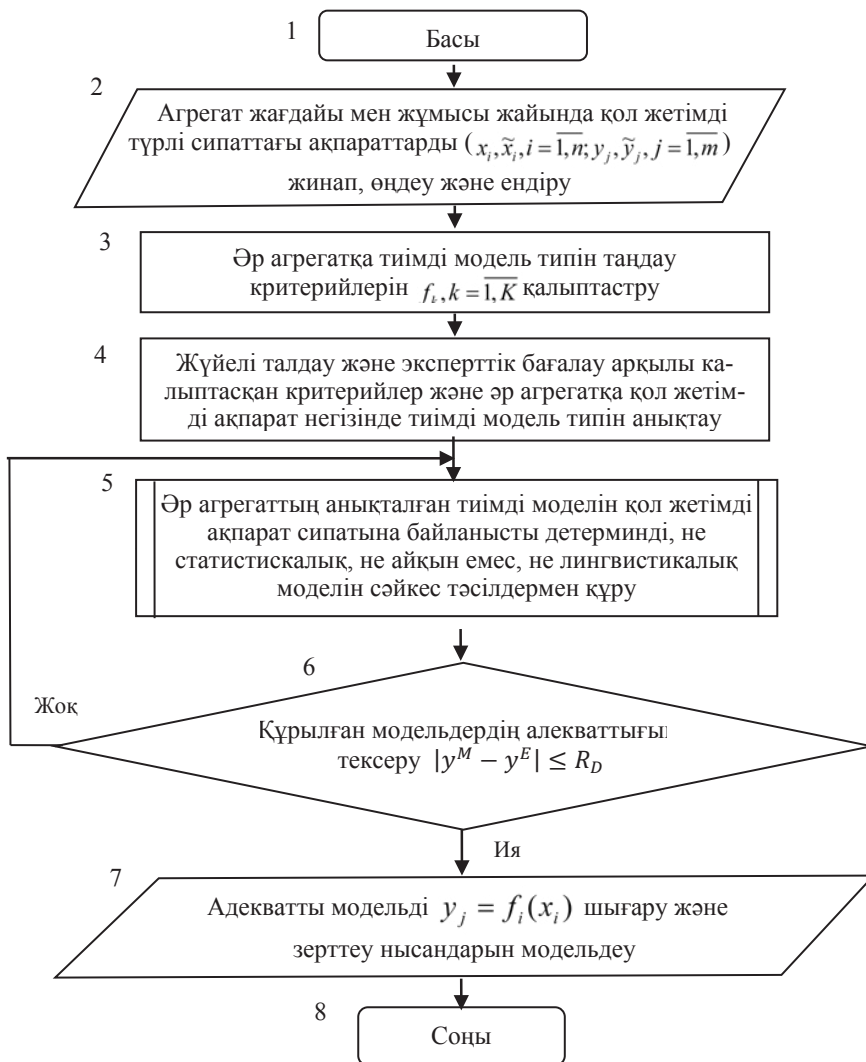
- қол жетімді түрлі сипаттағы ақпарат негізінде құрылатын жеке агрегаттардың модельдерін біріктіру арқылы технологиялық жүйе жұмысын жүйелі модельдеуге қабілетті модельдер пакетін алу.

БҚК мысалында технологиялық жүйелер агрегаттарының модельдерін құруда олардың жұмысын сипаттайтын қол жетімді ақпараттың түріне байланысты модель құрудың түрлі белгілі тәсілдері жүйеленіп, осы жұмыста ұсынылатын жүйелік тәсілдеме қолданылады.

Сонымен бұл зерттеуде технологиялық жүйелер агрегаттары жұмыс жайында теориялық мәліметтер жеткілікті болғанда, аналитикалық тәсілдер негізінде детерминді модельдер, ал сенімді статистикалық деректер жеткілікті болса, онда эксперименталдық-статистикалық тәсілдер (Jorgensen, 2019) негізінде статистикалық модельдер құрылатын болады. Бастапқы ақпараттың айқынсыздығы жағдайында мұндай ақпаратты жинау және өңдеу үшін эксперттік бағалау мен айқын емес жиындар тәсілдері (Orazbayev, 2018; Pavlov, 2016; Reverberi, 2019) қолданылады. Бұл зерттеуле БҚК сияқты технологиялық жүйелер агрегаттарының модельдері пакеті жүйелік талдау тәсілдемесі және гибридтік модель құру тәсілі [25] негізінде құрылады

### **Зерттеу нәтижелері**

Математикалық модельдерді құға қажетті бастапқы ақпараттардың жеткіліксіздігі мен айқын еместегі жағдайында БҚК сияқты күрделі технологиялық жүйелердің өзара байланысқан агрегаттарының модельдері пакетін құруға ұсынылған жүйелік әдістеменің жалпы блогының негізгі кезеңдері 1-суретте келтірілген.



Сурет 1 –Түрлі сипаттағы қол жетімді ақпараттар негізінде технологиялық жүйелердің өзара байланысқан агрегаттарының модельдерін құру әдістемесінің жалпы блок-схемасы (General flowchart of the methodology for constructing models of interconnected aggregates of technological systems based on available information of various nature)

Түрлі сипаттағы қол жетімді ақпараттар негізінде технологиялық жүйелердің өзара байланысқан агрегаттарының модельдерін құру әдістемесінің келтірілген жалпы блок-схемасының негізгі блокторына түсініктеме берейік.

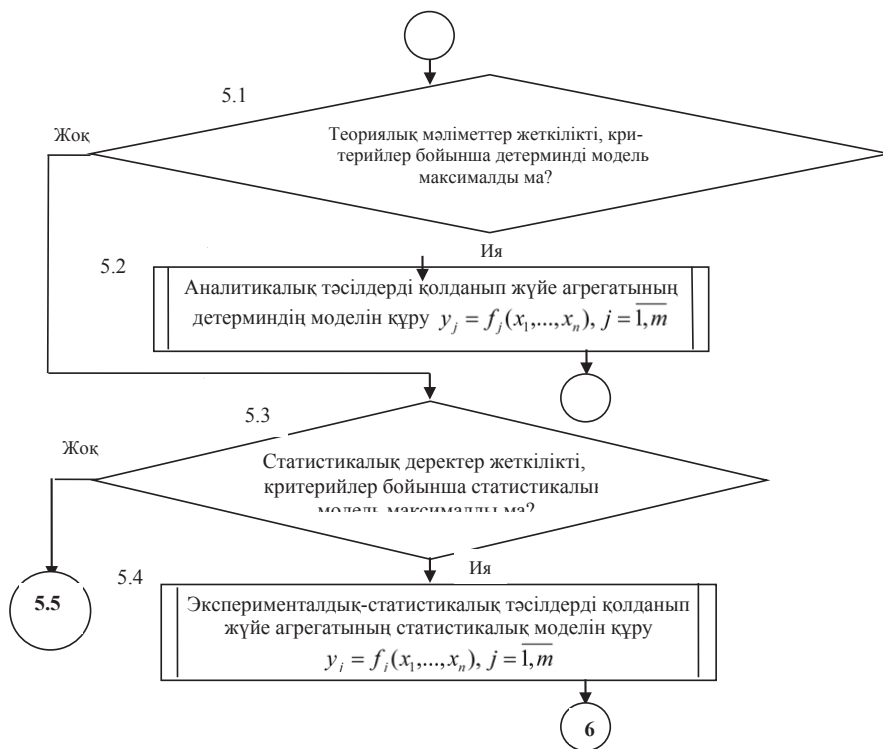
Әдістеменің 2-ші блогында зерттелетін технологиялық жүйе агрегаттарының жағдайы, жұмысы жайлы қол жетімді түрлі сипаттағы ақпараттар (теориялық мәліметтер, эксперименталдық-статистикалық деректер, айқын емес ақпарат) жинақталып, өңделіп, ендіріледі.

3-ші блокта әр агрегатқа тиімді модель типін таңдау үшін қажетті критерийлер  $f_k, k = 1, K$ , мысалы: модель құруға қажетті ақпараттың қол жетімдігі; агрегаттың жұмысын басқаруда қолданылуы; құру мүмкіндігі; адекваттық деңгейі; бір модельдер пакетіне біріктіру мүмкіндігі т.б. қалыптастарылады.

4-ші блокта жүйелі талдау және эксперттік бағалау арқылы қалыптасқан критерийлер және әр агрегатқа қол жетімді ақпарат негізінде тиімді модель типін анықтау жүзеге асырылады.

Ұсынылған әдістеменің бастапқы ақпараттардың сипаты мен тиімді модельді таңдау критерийлері бойынша нысандардың түрлі модельдері құрылатын негізгі 5-блогы төмендегі 2- және 3- суреттертегі әдістеменің жете келтірілген блок-схемаларында ашылған.

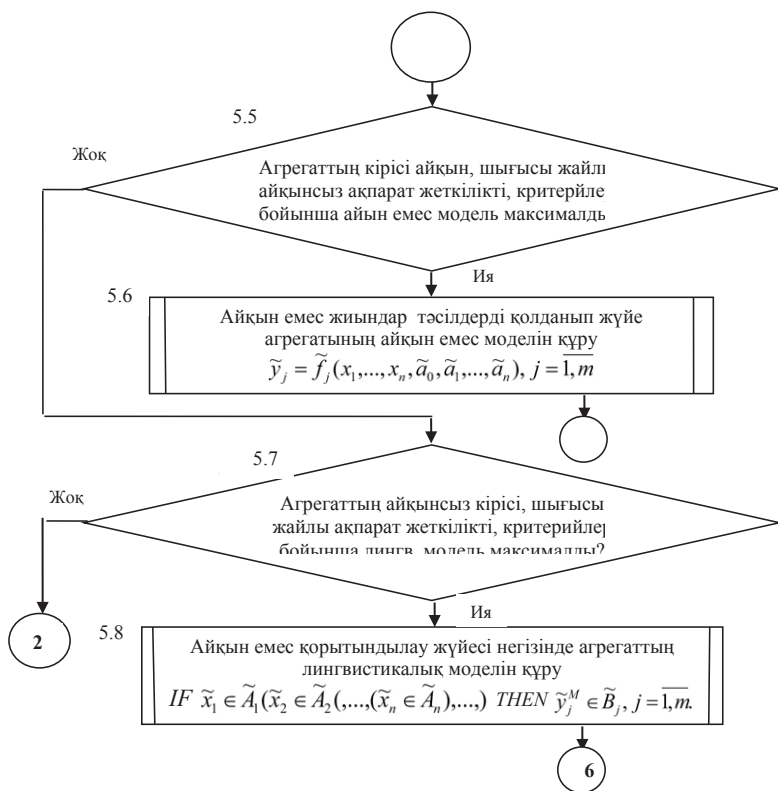
6-шы блокта құрылған модельдердің адекваттығын тексеріледі. Мысалы келесі критерий арқылы:  $|y^M - y^E| \leq R_D$ , мұнда  $y^M, y^E$  сәйкесінше, нысан шығысының модельдер арқылы алынған есептік және эксперименттер арқылы алынған шынайы мәндері;  $R_D$  рұқсат етілген ауытқу мәні.



Сурет 2 –Технологиялық жүйе өзара байланысқан агрегаттарының дәстүрлі тәсілдер негізінде детерминді және статистикалық модельдерін құру блок-схемасы (Block diagram of the construction of deterministic and statistical models of interconnected aggregates of technological systems based on traditional approaches)

Келтірілген дәстүрлі тәсілдер негізінде агрегаттың детерминді және статистикалық модельдерін құру блок-схемасында 5.1 блогында жеке агрегаттың моделін құруға теориялық мәліметтердің жеткіліктігі және критерийлер бағасы бойынша детерминді модель максималды мәнге иелігі тексеріледі. Егер бұл шарт орындалса, онда 5.2 блокта белгілі аналитикалық тәсілдер көмегімен агрегаттың детерминді моделі құрылады да құрылған модельдің адекваттығын тексеру үшін жалпы бло-схеманың 6-блогына өтеді.

Егер жоғарыдағы шарт орындалмаса, онда 5.3-блокқа өтеді. Бұл блокта агрегаттың моделін құруға эксперименталдық-статистикалық деректердің желкілікті және таңдау критерийлері бойынша статистикалық модель мәнінің максималдығы тексеріледі. Бұл шарт орындалған жағдайда 5.4-блокта эксперименталдық-статистикалық тәсілдер негізінде агрегаттың статистикалық моделі құрылады, құрылған модельдің адекваттығын тексеру үшін жалпы бло-схеманың 6-шы блогына өтеді. Ал шарт орындалмаса, онда айқын емес не лингвистикалық модель мүмкіндігін тексеріп, мұндай модельдерді құру үшін келесі блок-схеманың 5.5-блогына өтеді (сурет 3).



Сурет 3 –Технологиялық жүйе өзара байланысқан агрегаттарының ұсынылған айқын емес және лингвистикалық модельдерін құру блок-схемасы  
(Block diagram of the construction of the proposed fuzzy and linguistic models of interconnected aggregates of the technological system)

Сурет 3-те келтірілген агрегаттардың қол жетісміді ақпараттар негізінде оның айқын емес не лингвистикалық модельдерін құру үшін осы жұмыста жүйеленіп, ұсынылған тәсілдемелер қолданлады.

Егер технологиялық жүйе агрегатының кірісі  $\mathbf{x} = (x_1, \dots, x_n)$  айқын, ал күйі мен жұмысын сипаттайтын шығысы  $\tilde{\mathbf{y}} = (\tilde{y}_1, \dots, \tilde{y}_m)$  айқын емес және модельдерді таңдау критерийлері бойынша нысанның айқын емес моделін құру максималды мәнге болса (блок 5.5), онда оның айқын емес моделі келесідей құрылады.

5.6-блокта агрегаттың айқын емес моделін құру үшін эксперттік бағалау және айқын емес жиындар тәсілдері қолданылады. Бұл блокта агрегаттың айқын емес модельдердің құрылымын көпкерегрессиялық теңдеулер түрінде регрессорларды тізбектей қосу әдісі көмегімен анықтауға болады:

$$\tilde{y}_j = \tilde{a}_{0j} + \sum_{i=1}^n a_{ij} x_{ij} + \sum_{i=1}^n \sum_{k=i}^n a_{ikj} x_{ij} x_{kj}, j = \overline{1, m}. \quad (1)$$

Айқын емес модельдің құрылымы (1) оның айқын емес параметрлерін  $\tilde{a}_0, \tilde{a}_1, \dots, \tilde{a}_n$  идентификациялау кірістірілген цикл ретінде, сыртқы циклде модельдің құрылымы, ал ішкі циклінде параметрлері анықтау орындалады. Айқын емес параметрлерді идентификациялау есебін  $\alpha$  деңгейлі жиын негізінде модификацияланған ең кіші квадраттар тәсілі көмегімен шешуге болады.

Егер 5.7-блок шарты орындалса, яғни агрегаттың кірісі және шығыс параметрлері жайлы айқын емес ақпарат жеткілікті және критерийлер бойынша лингвистикалық модель максималды мәнге ие болса, онда 5.4-блокта нысанның лингвистикалық моделі құрылады.

Лингвистикалық модельді құру үшін нысанның айқын емес шығысына  $\tilde{y}_j \in \tilde{B}_j, j = \overline{1, m}$  әсер ететін кіріс параметрлері лингвистикалық айнымалылар арқылы сипатталады,  $\tilde{A}_i \in X, \tilde{B}_j \in Y$  – айқын емес ішкі жиындар,  $X, Y$  – әмбебап жиындар. Одан кейін эксперттік бағалау негізінде айқын емес параметрлерді сипаттайтын терм-жиындар  $T(X, Y)$  анықталып, олардың термдерінің тиістілік функциялары  $\mu_{A_i}(\tilde{x}_i), i = \overline{1, n}, \mu_{B_j}(\tilde{y}_j), j = \overline{1, m}$  тұрғызылады. Тиістілік функцияларды жылдам тұрғызу үшін MATLAB жүйесінің Fuzzy Logic Toolbox инструментарийін пайдалануға болады, немесе ол функциялардың құрылымдарфы мен параметрлерін идентификациялау керек.

Ары қарай нысанның кірісі  $\tilde{x}_i$  мен шығысын  $\tilde{y}_j$  сипаттайтын лингвистикалық айнымалылар арасындағы байланыстарын анықтайтын айқын емес бейнелеулер  $\tilde{R}_{ij}$  құру қажет.  $\tilde{R}_{ij}$  есептеулерде ыңғайлығы болу үшін оларды келесі тиістілік функциялары көмегімен байланысатар матрицасы ретінде келтірген дұрыс:

$$\tilde{\mu}_{\tilde{R}_{ij}}(\tilde{x}_i, \tilde{y}_j) = \min[\mu_{\tilde{A}_i}(\tilde{x}_i), \mu_{\tilde{B}_j}(\tilde{y}_j)], i = \overline{1, n}, j = \overline{1, m}. \quad (2)$$

Логикалық қорытындылау ережелері негізінде лингвистикалық модельді жалпы түрде келесідей құруға болады (блок 5.8):

$$IF \tilde{x}_1 \in \tilde{A}_1 (\tilde{x}_2 \in \tilde{A}_2 (\dots, (\tilde{x}_n \in \tilde{A}_n), \dots)) THEN \tilde{y}_j^M \in \tilde{B}_j, j = \overline{1, m}. \quad (3)$$

Технологиялық жүйе агрегатының айқын емес шығыс параметрлерінің мәндері айқын емес теориясындағы  $\tilde{B}_j = \tilde{A}_i \circ R_{ij}$  композициялық қорытындылау ережелері көмегімен анықтауға болады. Бұл ережелерді қолдану ыңғайлығы үшін агрегаттың айқын емес шығыс параметрлерінің мәндері максимінді көбейту негізінде келесідей тиістілік функциялармен жазылады:

$$\mu_{\tilde{B}_j}(\tilde{y}_j^*) = \max_{x_i \in X_i} \{ \min[\mu_{\tilde{A}_i}(\tilde{x}_i^*), \mu_{R_{ij}}((x_i^*, \tilde{y}_j^*))] \}. \quad (4)$$

Келтірілген (4) -рнектен агрегаттың шығыс параметрлерінің сандық шамалары  $y_j^c, j = \overline{1, m}$  ол параметрлердің тиістілік функцияларының максималды мәндерінің аргументімен анықталады яғни:

$$y_j^c = \arg \max_{\tilde{y}_j^*} \mu_{\tilde{B}_j}(\tilde{y}_j^*). \quad (5)$$

Технологиялық жүйе агрегаттарының күйі мен жұмысын сипаттайтын теориялық мәліметтер, статистикалық деректер және айқын емес ақпарат қол жетімсіз боған жағдайда, ол агрегаттың моделін қарастырылған түрлі тәсілдерді бірге қолдана отырып нысанның гибриді моделін құруға болады [25].

### **Нәтижелерді талқылау**

Ұсынылған түрлі сипаттағы қол жетімді ақпараттар негізінде технологиялық жүйелердің өзара байланысқан агрегаттары модельдерін құру әдістемесі баяу кокстеу қондырғысының кокстеу реакторлары, негізгі ректификациондық колоннасы және қыздыру пештерінің модельдері пакетін құруға бағытталған. Әдістеме жүйелік тәсілдеме көмегімен әр түрлі формалды және формалсыз тәсілдерді жүйелі қолдануға негізделген. Келтірілген 2-, 3-суреттердегі блок-схемада және олардың сипаттамаларында нысанның кіріс, шығыс параметрлерінің айқынсыздығына байланысты оның айқын емес немесе лингвистикалық модельдерін құру процедуралары ұсынылып, сипатталған. Өзірленіп, сипатталған әдістемеді түрлі белгілі және зерттеу барысында алынған (айқын емес, лингвистикалық модельдерді құру) тәсілдері біртұтас жүйе ретінде қолданылады. Сондықтан ұсынылған жүйелік әдістеме негізінде құрылған модельдер жүйесі эмерджендік қасиетке және синергизм эффектісіне ие. Ал бұл жүйелдік қасиет пен эффектілер әзірленген жүйелік әдістеменің басқа тәсілдемелердің артықшылығын көрсетеді.

Технологиялық жүйе, мысалы БКҚ агрегаттарының өзара байланысқан агрегаттарының модельдер пакетін құрып, оны жүйелік модельдеуге қабілетті



әзірленген жүйелік әдістеме нысанның айқын емес шығыс параметрлері максимінді көбейту нешінде ол параметрлердің тиістілік функциялары арқылы (4) өрнекпен анықталады. Ал нысанның шығыс параметрлерінің сандық мәндері  $y_j^c, j = 1, m$  олардың анықталған айқын емес шығыс параметрлерінің тиістілік функцияларының максималды аргументі ретінде анықталады. сандық мәндері  $y_j^c, j = 1, m$ , бұл (3) формула бойынша анықталған тиістілік функциясының максималды мәндерін қамтамасыз ететін аргументі ретінде (5) формуламен анықталады. Технологиялық жүйелердің өзара байланысқан агрегаттарының тиімді модельдерін құрып, біріккен модельдер пакетіне әзірлеуге ұсынылған бұл тәсіл бастапқы ақпараттың жетіспеушілігі, айқынсыздығы жағдайларында қол жетімді түрлі сипаттағы ақпараттар негізінде нысандардың түрлі модельдерін құратын жүйелі әдістеме болып табылады. Әдістеме модель құруға қажетті қол жетімді ақпараттардың сипаты мен модельдерді таңдау критерийлері мәндеріне байланысты нысандардың детерминдік, стохастикалық, айқын емес немесе лингвистикалық модельдерін құруға мүмкіндік береді. Содан кейін құрылған модельдерді технологиялық жүйеде өтетін процесс және агрегаттар арасындағы ағындарды ескере отырып, бір модельдер пакетіне біріктіру қажет.

БКҚ сияқты бастапқы ақпараттың тапшылығы және айқынсыздығымен сипатталатын технологиялық жүйелердің агрегаттарының модельдерін құрудың белгілі тәсілдерінен әзірленген жүйелік әдістемесінің айырмашылығы, ол қол жетімді түрлі сипаттағы ақпараттар негізінде жүйелердің өзара байланысқан модельдер пакетін құруға мүмкіндік береді. Сондай-ақ ұсынылған жүйелік әдістеме жүйе агрегаттарының дәстүрлі детерминді, статистикалық модельдерімен қатар айқын емес, лингвистикалық модельдерін де құрып, алынған модельдерді бір модельдер пакетіне біріктіруге мүмкіндік береді.

### **Қорытынды**

Жұмыста БКҚ сияқты бастапқы ақпараттың жетіспеушілігі және айқынсыздығымен сипатталатын технологиялық жүйелердің өзара байланысқан агрегаттарының бір пакетке біріктірілген модельдерін құруға арналған жүйелік әдістеме әзірленіп, сипатталған. Әзірленген әдістеме жүйелік тәсілдемеге, түрлі модельдер құрудың дәстүрлі тәсілдері мен айқынсыздықта модель құрудың ұсынылған тәсілдерін жүйелеп, кешенді қолдануға негізделген.

Зерттеу нәтижелері бойынша келесідей негізгі қорытындылар жасауға болады:

1) БКҚ сияқты технологиялық жүйелердің өзара байланысқан агрегаттарының бастапқы қол жетімді ақпарат сипатына байланысты құруға болатын модельдерінің түрін жүйелік талдау мен эксперттік бағалау тәсілдері арқылы олардың тиімді модельдерін таңдау және таңдалған модельді құру процедурасы ұсынылып, сипатталған;

2) Зерттеу нысанының күй мен жұмысы жайында, оның кіріс, шығыс параметрлерінің айқынсыздығына байланысты, нысандардың айқын емес немесе лингвистикалық модельдерін құру тәсілі ұсынылған (сурет 3). Бұл тәсіл ақпараттың тапшылығы мен айқынсыздығында агрегаттардың түрлі модельдерін құру үшін әзірленген жүйелік әдістеме құрамына енеді;

3) Түрлі қол жетімді ақпарат негізінде өзара байланысқан агрегаттардың модельдерін бір модельдер пакетіне, технологиялық жүйеде жүретін процесстің өтуін ескере отырып біріктіру, принципі сипатталған. Ұсынылған принцип нысандағы технологиялық процесстің жүруі бағғытына ескеріле отырып, агрегат модельдерінің біреуінің шығысы басқа модельдердің кіріс ретінде пайдаланылады.

#### REFERENCES

Aliyev R.A., Tserkovny A.E., Mammadova G.A. (2018). Production management with fuzzy initial information." -M.: Energoatomizdat. 2nd ed. 2018. -253 p.

Tuleuov Zh.N., Suleimenov D. (2018). Technological regulations for the installation of delayed coking of the UZK 21-10/6 Atyrau refinery: - Atyrau: publishing house, 2018. 217 p.

Valyavin G.G., Zaporin V.P., Kalimullin T.I. (2020). The process of delayed coking and the production of petroleum cokes, specialized in application. -Tomsk: Publishing House of Tomsk Polytechnic University, 2020. 135 p.

Kafarov V.V., Dorokhov I.N. (2018). System analysis of chemical and technological processes. Fundamentals of Strategy. -M.: Nauka, 3rd ed. 2018, 517 p.

Zhorov M. (2015). Calculations and modeling of thermocatalytic processes in oil refining". -Moscow: Energoatomizdat, 2nd ed. 2015, 307 p.

Kabibullin M.D., Orazbayev B.B., Orazbayeva K.N., Iskakova S.Sh., Amanbayeva Zh.Sh. (2023). Kurdeli himiyalyk-technologiyalyk zhuyeler aggregattaryn modelderin bastapky akparattyn zhetispeushiligi men aykynsyzdygy zhagdayynda kuru // Proceedings of the National Academy of Sciences of the Republic of Kazakhstan Kazakh National- Al-Farabi National University. Series physics and information technology. 2 (346) April – June 2023. -Pp.154–171. ISSN 1991-346X

Mohaddecy S.R., Zahedi S., Sadighi S., Bonyad H. (2006). Reactor modeling and simulation of catalytic reforming process. // Petroleum & Coal. 2006. V. 48. № 3. -Pp. 28–35.

Sharikov Yu.V., Petrov P.A. (2018). Universal model for catalytic reforming, Chemical and Petroleum Engineering, 2018. V. 25, №4. -Pp. 33–47.

Pinheiro C.C., Fernandes J.L., Domingues L. (2018). Catalytic Cracking (FCC) Process Modeling, Simulation, and Control // Industrial Engineering Chemistry Research., 2018. V 51, № 1. -Pp. 1–29.

Zhi-Wen Zhao, De-Hui Wang. Statistical inference for generalized random coefficient autoregressive model // Mathematical and Computer Modelling, 2018, V.56. № 3. -Pp.152–166.

Coleman B., Babu J. (2018). Techniques of Model Based Control. UpperSaddle River: Prentice Hall PTT, 2018, p. 576.

Поспелов Д.А. Ситуационное управление: теория и практика. -М.: Наука, 3-е изд. 2018, стр. 305.

Keller T., Gorak A. (2019). Modelling of homogeneously catalysed reactive distillation processes in packed columns: Experimental model validation. Comput. Chem. Eng. 2019. V. 48, -P. 74–93.

Dzhambekov A.M., Fyodorova O.V. (2018). Optimal control of the process of catalytic reforming of petrol fractions”, Bulletin of the Astrakhan State Tech. University Ser. control, computing tech. inform., 2018. V. 329. №. 2. -Pp.34–42.

Карманов Ф.И., Острейковский В.А. (2019). Статистические методы обработки экспериментальных данных с использованием пакета MathCad, -М.: Радио и связь. 2019. -187 с. ISBN: 978-5-905554-96-4.

Шакирзянова Г.И., Сладовская О.Ю., Сладовский А.Г. (2019). Замедленное коксование как эффективная технология углубления переработки нефти. -М.: Химия, 2-ен изд. 2019. 347 с.

Douglas A.M. and Danny A.M. (2021). Statistical Methods in Experimental Pathology: A Review and Primer. The American Journal of Pathology. 2021. Vol. 191. No 5. 784–794 <https://doi.org/10.1016/j.ajpath.2021.02.009>

Jorgensen M. (2019). A Review of Studies on Expert Estimation of Software Development Effort. Journal of Systems and Software. 2019. V.70. -Pp.37–60.

Гуцыкова С.В. (2017). Метод экспертных оценок. Теория и практика, -М., ИП РАН. 2017. -278 p. ISBN: 978-5-9270-0209-2.

Zimmermann H.-J. Fuzzy Set Theory – and Its Applications. Springer Science+Business Media, LLC. Fifth Edition 2018. p.525. ISBN: 978-94-010-3870-6.

Ryzhov A.P. (2017). Theory of fuzzy sets and its applications. -Moscow: MSU, 2017. -278. ISBN: 978-3-540-70777-6Dubois D. The role of fuzzy sets indecision sciences: old techniques and new directions. *Fuzzy Sets and Systems*. 2011. V. 184. 3–17.

Reverberi A.P., Kuznetsov N.T., Meshalkin V.P., Salerno M., Fabiano B. Systematical Analysis of Chemical Methods in Metal Nanoparticles Synthesis//Theoretical Foundations of Chemical Engineering. 2019. V. 50. №1. -Pp. 63–75.

Pavlov S.Yu., Kulov N.N., Kerimov R.M. Improvement of Chemical Engineering Processes Using Systems Analysis.//Theoretical Foundations of Chemical Engineering. 2016. V. 53. № 2. -Pp.117–133.

Orazbayev B.B., Ospanov E.A., Orazbayeva K.N., Kurmangazieva L.T. (2018). A Hybrid Method for the Development of Mathematical Models of a Chemical Engineering System in Ambiguous Condition // Mathematical Models and Computer Simulations, ISSN 2070-0482, 2018, V. 10, № 6, -Pp.748–758.

UDC 004.94

© **G.B. Bahadirova<sup>1\*</sup>, H. Tasbolatuly<sup>1</sup>, A.S. Mukanova<sup>1</sup>, Sh. Turaev<sup>2</sup>, 2023**

<sup>1</sup>Astana International University, Astana, Kazakhstan;

<sup>2</sup>College of Information Technology, United Arab Emirates University,  
Al Ain, United Arab Emirates.

E-mail: [gulnaz.bahadirova.84@mail.ru](mailto:gulnaz.bahadirova.84@mail.ru)

## **DESIGNING LINEAR FEEDBACK CONTROL FOR A NONLINEAR SYSTEM IN MATLAB SIMULINK**

**Bahadirova Gulnaz Bekualykyzy** — PhD student. Astana International University. Astana, Kazakhstan

E-mail: [gulnaz.bahadirova.84@mail.ru](mailto:gulnaz.bahadirova.84@mail.ru). <https://orcid.org/0000-0002-9464-9809>;

**Tasbolatuly Nurbolat** — associate professor, Deputy Dean of the Higher School of Information Technologies and Engineering, Astana International University. Astana, Kazakhstan

E-mail: [tasbolatuly@gmail.com](mailto:tasbolatuly@gmail.com). <https://orcid.org/0000-0002-0511-7000>;

**Mukanova Asel Serikovna** — associate professor, dean of the Higher School of Information Technologies and engineering, Astana International University. Astana, Kazakhstan

E-mail: [assel.mukanova@aiu.edu.kz](mailto:assel.mukanova@aiu.edu.kz). <https://orcid.org/0000-0002-8964-3891>;

**Turaev Sherzod** — PhD, assoc. Professor. College of Information Technology, United Arab mirates University Al Ain, United Arab Emirates

E-mail: [sherzod@uaeu.ac.ae](mailto:sherzod@uaeu.ac.ae). <https://orcid.org/0000-0001-6661-8469>.

**Abstract.** Changes in equations in nonlinearized systems of existence are characterized by nonlinearized differential equations. The tasks of monitoring nonlinearized systems have been of great interest to the scientific community for several recent years. Significant positive changes are observed in the development of various methods of nonlinearized control theory and obtaining results by modeling them on a computer. Such success is associated with the rapid development of computing equipment and increased requirements for the quality of control, which allows you to fully and quickly study nonlinearized trends. The purpose of the article is to build a computer model of linearized feedback control for a nonlinearized system with one input and one output for a pendulum system, using various blocks of MATLAB Simulink libraries. In a computer model for linearized feedback control for a nonlinearized system, a reference signal block, a pendulum block Tracking control, was created and the input signal or control, the output signal, and the error results were displayed as graphs. To construct the law of control of nonlinearized systems, there are several approaches, including

the method of constructing the law of control using linearization, the method of linearization of the system by feedback, Lyapunov methods. In addition, the presented article provides and describes brief definitions, lemmas about feedback control, state control by feedback, stabilization, tracking. Using various blocks of MATLAB Simulink libraries for the pendulum system, a nonlinearized system, a scheme and a result of the feedback control law were given.

**Keywords:** Nonlinearized systems, large-scale stabilization, output feedback, state feedback, tracking

© Г.Б. Бахадирова<sup>1\*</sup>, Н. Тасболатұлы<sup>1</sup>, А.С. Муканова<sup>1</sup>, Ш. Тураев<sup>2</sup>, 2023

<sup>1</sup>Астана халықаралық университеті, Астана, Қазақстан;

<sup>2</sup>Ақпараттық технологиялар колледжі, Біріккен Араб Әмірліктері университеті, Әл-Айн, Біріккен Араб Әмірліктері.

E-mail: [gulnaz.bahadirova.84@mail.ru](mailto:gulnaz.bahadirova.84@mail.ru)

## МATLAB SIMULINK-ТЕ СЫЗЫҚТЫҚ ЕМЕС ЖҮЙЕ ҮШІН КЕРІ БАЙЛАНЫСТЫ СЫЗЫҚТЫҚ БАСҚАРУДЫ ЖОБАЛАУ

**Бахадирова Гулназ Бекуалықызы** — PhD студент. Астана халықаралық университеті. Астана, Қазақстан

E-mail: [gulnaz.bahadirova.84@mail.ru](mailto:gulnaz.bahadirova.84@mail.ru). <https://orcid.org/0000-0002-9464-9809>;

**Тасболатұлы Нұрболат** — доцент, Ақпараттық технологиялар және инженерия жоғары мектебінің декан орынбасары, Астана халықаралық университеті. Астана, Қазақстан

E-mail: [tasbolatuly@gmail.com](mailto:tasbolatuly@gmail.com). <https://orcid.org/0000-0002-0511-7000>;

**Муканова Асель Сериковна** — доцент, Ақпараттық технологиялар және инженерия жоғары мектебінің деканы. Астана халықаралық университеті. Астана, Қазақстан

E-mail: [assel.mukanova@aiu.edu.kz](mailto:assel.mukanova@aiu.edu.kz). <https://orcid.org/0000-0002-8964-3891>;

**Тураев Шерзод** — PhD, қауымдастырылған профессор. Ақпараттық технологиялар колледжі, Біріккен Араб Әмірліктері университеті, Әл-Айн, Біріккен Араб Әмірліктері

E-mail: [shezod@uaeu.ac.ae](mailto:shezod@uaeu.ac.ae). <https://orcid.org/0000-0001-6661-8469>.

**Аннотация.** Сызықтық емес болмыстық жүйелердегі тендеулер өзгерісі сызықтық емес дифференциалдық тендеулермен сипатталады. Сызықтық емес жүйелерді бақылау міндеттері бірнеше соңғы жылдарды ғылыми қауымдастық тарапынан үлкен қызығушылыққа ие болып келеді. Сызықтық емес басқару теориясының әр түрлі әдіс-тәсілдерін әзірлеу мен оларды компьютерде модельдеу арқылы нәтиже алуда айтарлықтай оң өзгерістер байқалуда. Мұндай жетістік сызықтық емес үрдістерді толық және тез зерттеуге мүмкіндік беретін есептеуіш техниканың қарқынды дамуымен және бақылау сапасына қойылатын талаптардың артуымен байланысты. Мақаланың мақсаты маятниктік жүйеге арналған бір кірісі және бір шығысы бар сызықтық емес жүйе үшін кері байланысты сызықтық басқаруды, MATLAB Simulink кітапханаларының әр түрлі блоктарын қолдану арқылы компьютерлік моделін құру болып табылады. Сызықтық емес жүйе үшін кері байланысты сызықтық басқаруға арналған компьютерлік модельде тірек сигналының блогы, маятник блогының ізіне түсіруді басқару,

сызықтандыруды басқару жасалып, кіріс сигналы немесе басқару, шығыс сигналы және қателіктің нәтижелері график түрінде көрсетілді. Сызықтық емес жүйелерді басқару заңын құру үшін бірнеше тәсілдер бар, оның ішінде сызықтандыруды пайдаланып басқару заңын құру әдісі, жүйені кері байланыс арқылы сызықтандыру әдісі, Ляпунов әдістері. Сонымен қатар берілген мақалада кері байланысты басқару, күй кері байланыс арқылы басқару, орнықтандыру, ізге түсіру туралы қысқаша анықтамалар, леммалар беріліп, сипатталды. Маятниктің жүйе үшін MATLAB Simulink кітапханаларының әртүрлі блоктарын қолдану арқылы сызықтық емес жүйе, кері байланысты басқару заңының схемасы және нәтижесі келтірілді.

**Түйін сөздер:** сызықтық емес жүйелер, кең ауқымды орнықтандыру, шығыс кері байланыс, күй кері байланысы, ізге түсіру

© Г.Б. Бахадирова<sup>1\*</sup>, Н. Тасболатұлы<sup>1</sup>, А.С. Муканова<sup>1</sup>, Ш.Тураев<sup>2</sup>, 2023

<sup>1</sup>Международный университет Астана;

<sup>2</sup>Колледж информационных технологий, Университет Объединенных Арабских Эмиратов, Эль-Айн, Объединенные Арабские Эмираты.

E-mail: gulnaz.bahadirova.84@mail.ru

## **ПРОЕКТИРОВАНИЕ ЛИНЕЙНОГО УПРАВЛЕНИЯ С ОБРАТНОЙ СВЯЗЬЮ ДЛЯ НЕЛИНЕЙНОЙ СИСТЕМЫ В MATLAB SIMULINK**

**Бахадирова Гулназ Бекуалықызы** — PhD студент. Международный университет Астана.

E-mail: gulnaz.bahadirova.84@mail.ru. <https://orcid.org/0000-0002-9464-9809>;

**Тасболатұлы Нұрболат** — доцент, заместитель декана Высшей школы информационных технологий и инженерии, Международный университет Астана, Астана, Казахстан.

E-mail: tasbolatuly@gmail.com, <https://orcid.org/0000-0002-0511-7000>

**Муканова Асель Сериковна** — доцент, декан Высшей школы информационных технологий и инженерии, Международный университет Астана, Астана, Казахстан

E-mail: assel.mukanova@aiu.edu.kz. ORCID:0000-0002-8964-3891;

**Тураев Шерзод** — PhD, ассоциированный профессор. Колледж информационных технологий. Университет Объединенных Арабских Эмиратов, Аль-Айн, Объединенные Арабские Эмираты

E-mail: sherzod@uaeu.ac.ae. <https://orcid.org/0000-0001-6661-8469>.

**Аннотация.** Изменение уравнений характеризуются нелинейными дифференциальными уравнениями. Проблемы управления нелинейными системами в последние годы вызывают большой интерес в научном сообществе. Наблюдаются существенные изменения в развитии различных методов теории нелинейного управления и их компьютерного моделирования. Эти достижения связаны как с бурным развитием компьютерных технологий, позволяющих полно и быстро исследовать нелинейные процессы, так и с повышением требований к качеству управления. Целью данной работы является создание компьютерной модели управления с линейной обратной связью для нелинейной маятниковой системы с одним входом и одним

выходом с использованием различных блоков библиотеки MATLAB Simulink. В компьютерной модели управления с линейной обратной связью для нелинейной системы были созданы блоки опорного сигнала, следящего управления маятником и линеаризованного управления, построены графики входного сигнала или управления, выходного сигнала и результатов ошибки. Методы создания законов управления для нелинейных систем включают использование линеаризации для создания законов управления, линеаризацию системы с помощью обратной связи и метод Ляпунова. В работе также даны и объяснены простые определения и леммы по управлению с обратной связью, управлению с обратной связью по состоянию, стабилизации и слежению. Представлена нелинейная система, схема закона управления с обратной связью и результаты для маятниковой системы с использованием различных блоков библиотеки MATLAB Simulink.

**Ключевые слова:** нелинейные системы, крупномасштабная стабилизация, обратная связь по выходу, обратная связь по состоянию, отслеживание

### Кіріспе

Кіріс және шығыс айнымалылар арасындағы сызықтандыруға келмейтін жүйелерді немесе сызықтық емес жүйелерді қазіргі заманғы технологияларда зерттеудің тиімділігі артып келеді. Басқару сызықтық емес жүйелерді жүйелерді түсінуге және оларды нақты әлемдік мәселелерді шешуге манипуляциялауға арналған басқару теориясының бөлімі болып табылады. Бұл жүйелер табиғатта және техникада жиі кездеседі және оларды басқару қиын болуы мүмкін. Сызықтық емес жүйелер - бұл кіріс пен шығысты сызықтық тәсілдері арқылы сипаттауға болмайтын жүйелер. Математика мен жаратылыстану ғылымдарында сызықтық емес жүйе - шығыстың өзгеруі кірістің өзгеруіне пропорционалды емес жүйе. Сызықтық емес жүйе деп сызықтық шарттарды қанағаттандырмайтын жүйені айтамыз. Сызықтық емес жүйелердегі теңдеулер өзгерісі сызықтық емес дифференциалдық теңдеулермен сипатталады. Қазіргі кезде сызықтық емес жүйелерді басқарудың бірнеше әдістері бар.

Алғаш рет сызықтық емес жүйелер үшін тірек сигналдары уақытқа байланысты өзгертін күйі А. Айсидори және С.И. Бирнс (Burnes et al., 1997) еңбектерінде жазылған.

Ал ХХІ ғасырдың басында анықталмайтын сызықтандыруы бар белгісіз сызықтық емес жүйелер үшін динамикалық шығыс кері байланысы бар робосты практикалық шығыс бақылауды Алимхан және Инаба еңбектерінде жазылған (Alimhan et al., 2008).

Сонымен қатар уақыты кешіккен жоғары ретті сызықтық емес жүйелер класы үшін шығыс кері байланысты бақылау контроллерінің жобалық тәсілі ұсынылған. Уақыттың кешігуіне арналған Ляпунов–Красовский тәсілі ұсынылады, ал шығыс кері байланысты бақылау контроллерін біртекті үстемдік әдісін қолдана отырып жасалғандығы (Alimhan et al., 2021) мақалада қарастырылған.

Қазіргі күнге дейін сызықтық емес болған бақылаушыны құрудың жалпыланған және тиімді әдісі жоқ болуына байланысты күйі бойынша кері байланысты бақылауға қарағанда, шығыс кері байланысты бақылау баяу дамығанын байқауға болады.

Қазіргі зерттеулерде уақыт кешігу параметрінің өзі уақыт бойынша өзгеретін функция деп алынып, айнымалы уақыт кешігу параметрі бар шын мәніндегі сызықтық емес жүйелер класы үшін шығыс сигналын күйі бойынша кері байланыс арқылы берілген тірек сигнал ізіне түсіру мәселелері қарастырылып, зерттелу үстінде.

Маятникке арналған жүйенің кері байланысты сызықтық басқаруды пайдалана барысында жүйе орнықтылығы, бақылау және ауытқуларды азайту, оларды компьютерде моделдеу мәселелері қарастырылды. Компьютерлік модельдеуге анықтама беретін болсақ, онда компьютерлік модельдеу - бұл нақты немесе физикалық жүйенің күйін немесе нәтижелерін болжауға арналған және компьютерде орындалатын математикалық модельдеу процесі. Сызықтық емес жүйенің арнайы класын кері байланыс жәрдемінде бақылауды теориялық зерттеу, дәлелдеу нәтижесінде қол жеткізген жетістіктерді компьютерлік модельдеудің бірден бір тәсілі-MATLAB/Simulink сияқты модельдеу құралын пайдалану болып табылады.

### **Әдістер мен материалдар**

Сызықтық емес процестерді басқару үшін сенімді кері байланысты сызықтандыру әдісі зерттеледі және келесідей сипатталады: 1) егер сызықтық басқарылатын процесс тұрақты болса, онда сызықты емес процесс күйлері асимптотикалық тұрақты болады, ол қолданбаларда қанағаттандырылмайды, өйткені кейбір күйлер шағын мәндерге жақындайды; сондықтан, егер сызықтық басқарылатын процесс тұрақты болса, онда сызықты емес процесс күйлері біркелкі тұрақты болатынын дәлелдеу үшін Ляпунов теориясына негізделген теорема ұсынылады. 2) сызықтық емес процестерді реттеу үшін барлық негізгі және қиылысатын күйлердің кері байланысын ескеру керек екені теорияларда айтылады, бұл контроллердің кірістерін табуды қиындатады; демек, қолданбаларда қанағаттанарлық нәтиже алу үшін тек негізгі күйлердің кері байланыстары пайдаланылады. (José de Jesús Rubio, 2018)

Кері байланысты сызықтық басқаруды жобалауды зерттеу барысында келесі Ляпуновтың жүйелерді басқару, кері байланыс әдістері қолданылды. Енді осы әдістерге жеке жеке тоқталатын болсақ, онда төменде келтірілген.

### **1. Ляпуновтың орнықтылық теоремалары**

Ляпунов дифференциалдық теңдеуді анық интегралдаусыз жүйенің орнықтылығын анықтау әдісін ойлап тапты, яғни механикалық қасиеттерді пайдаланды, жүйенің энергетикалық өрісінің оның орнықтылығына әсер етуін зерттеді. Бұл Ляпуновтың тікелей әдісі немесе Ляпуновтың екінші әдісі деп аталды. Классикалық механика бойынша физикалық жүйеде энергияның төмен болуымен салыстырғанда жоғары энергияға ие болған кезде, массаның



орнықты болмағанын айтады. Бөлшектердің орныксыз күйден орнықты күйге ауысқан кезде оның энергиясы үнемі төмендеуі тиіс. Егер энергияны  $E$  арқылы өрнектесек, онда  $E > 0$ ,  $dE/dt < 0$  болады. Механикалық осциллятор жұмысын мысал ретінде, яғни осциллятордың жылдамдығы азайған кезде, жүйенің жалпы энергиясы да төмендеп, соңында тепе-теңдік нүктеде нөлге айналып орнықты болады.

Жоғарыда айтылған принциптердің негізінде Ляпунов тек қана күй энергиясы арқылы сипатталатын  $V(t, x(t))$  энергия функциясын құрды.

$$\text{Егер } V(t, x(t)) \begin{cases} > 0, \text{ егер } x \neq 0, \\ = 0, \text{ егер } x = 0, \end{cases}$$

$\dot{V}(t, x(t)) \leq 0$  болса, тепе-теңдік нүктедегі орнықтылық – жүйенің қозғалыс теңдеуінің шешімдері туралы ақпаратсыз дәлелденеді. Мұндағы:  $V(t, x(t))$ -*Ляпунов функциясы* деп аталады.

*Теорема 1.1.* Дискретті уақыттағы жүйелерге арналған үшін Ляпуновтың орнықтылық теоремасы:  $x(k+1) = f(k, x(k))$ ,  $x(k_0) = x_0$  жүйе берілсін, мұнда  $f(0, k) = 0$  болсын. Жүйенің тепе теңдік нүктесі  $x_e = 0$  тең;

1) Оң функция  $V(k, x(k))$  бар және  $\Delta V(k, x(k)) := V(k+1, x(k+1)) - V(k, x(k)) \leq 0$ ,  $\forall x \neq 0$  шарты қанағаттандыратын болса, келесі берілген жүйе  $x_e = 0$  нүктесінде орнықты болып табылады.

2) Егер функция  $V(k, x(k))$  бар және  $\Delta V(k, x(k)) := V(k+1, x(k+1)) - V(k, x(k)) \leq 0$ , шарты қанағаттандыратын болса, келесі берілген жүйе  $x_e = 0$  нүктесінде асимптотикалық орнықты болып табылады.

Берілген жүйе  $x_e = 0$  тепе-теңдік нүктесінде асимптотикалық орнықты және  $\|x\| \rightarrow \infty$  ұмтылғанда,  $V(k, x(k)) \rightarrow \infty$  болса, көрсетілген жүйеде  $x_e = 0$  нүктесінде кең ауқымды асимптотикалық орнықты болып табылады (Wu M et al., 2010).

*Теорема 1.2.* Үздіксіз уақыт жүйелеріне арналған Ляпуновтың орнықтылық теоремасы  $x(t) = f(t, x(t))$ ,  $x(t_0) = x_0$  берілген жүйеде  $f(t, 0) = 0$  болса, яғни жүйенің тепе-теңдік нүктесі  $x_e = 0$  болып табылады.

1) Егер  $V(k, x(t))$  оң анықталған және оң туындысы  $V(t, x(t)) := \frac{d}{dt} V(t, x(t))$  жартылай теріс анықталған болса, келесі берілген жүйе  $x_e = 0$  тепе теңдік нүктесінде орнықты болып табылады.

2) Егер  $V(k, x(t))$  оң анықталған және оң туындысы  $V(t, x(t)) := \frac{d}{dt} V(t, x(t))$  жартылай теріс анықталған болса, келесі берілген жүйе  $x_e = 0$  тепе теңдік нүктесінде асимптотикалық орнықты болып табылады

3) Егер берілген жүйе  $x_e = 0$  нүктеде асимптотикалық орнықты және  $\|x\| \rightarrow \infty$  ұмтылғанда,  $V(k, x(k)) \rightarrow \infty$  ұмтылатын болса, көрсетілген жүйеде  $x_e = 0$  нүктесінде кең ауқымды асимптотикалық орнықты болып табылады (Liao X. X. et al., 2007).

## 2. Кері байланысты басқару

Басқару жүйелері  $x=f(t,x,u)$ ,  $x \in \square^n, u \in \square^m$  (2.1)

динамикалық кері байланыс (немесе динамикалық компенсатор) (2.1) формула бойынша кеңейтілген басқару жүйесі болып табылады.

$$\begin{aligned} \dot{x} &= f(t, x, \beta(t, x, y, w)), \\ \dot{y} &= g(t, x, y, w), \\ u &= \beta(t, x, y, w), \end{aligned} \quad (2.2)$$

мұндағы  $x \in \square^n, y \in \square^k, w \in \square^{m'}$   $m' \geq m$ .

Динамикалық кері байланыс (2.2) тұрақты, егер  $m' = m$  және ол шешімнің сәйкестік шартын қанағаттандырады: Әрбір тегіс шешім үшін  $(x(t), u(t))$  бастапқы жүйе (2.1), біркелкі функциялар бар (міндетті түрде бірегей емес).  $(y(t), w(t))$  берілген функциялармен бірге  $(x(t), u(t))$  жүйені (2.2) бірдей қанағаттандырады (Clelland J. et al., 2021).

Күйі бойынша кері байланыс арқылы жүйені тұрақтандыру проблемасы келесідей жүйе түрде берілген болса  $\dot{x} = f(t, x, u)$ , кері байланысты басқару заңын әзірлеу проблемасы  $u = \gamma(t, x)$ , осылайша, координаттардың басталуы  $x = 0$  бұл тұйық жүйенің біркелкі асимптотикалық тұрақты тепе-теңдік нүктесі  $\dot{x} = f(t, x, \gamma(t, x))$  болып табылады.

Кері байланысты басқару заңы  $u = \gamma(t, x)$  әдетте «статикалық кері байланыс» деп аталады, себебі бұл  $x$ -тің жадсыз функциясы болып қарастырылады. Кейде біз жағдай бойынша динамикалық күйі кері байланысты басқаруды  $u = \gamma(t, x, z)$  қолданамыз. мұндағы  $z$  -  $x$  басқаратын динамикалық жүйенің шешімі; яғни  $\dot{z} = g(t, x, z)$  болып табылады.

Шығыс кері байланыс арқылы жүйені орнықтандыру проблемасы

$$\begin{aligned} \dot{x} &= f(t, x, u) \\ y &= h(t, x, u) \end{aligned}$$

шығыс кері байланысы бар статикалық басқару заңын әзірлеу проблемасы болып табылады  $u = \gamma(t, y)$  немесе шығыс кері байланысы бар динамикалық басқару заңы  $u = \gamma(t, y, z)$  осылайша, координаттардың басталуы  $x = 0$  бұл тұйық жүйенің біркелкі асимптотикалық тұрақты тепе-теңдік нүктесі болып табылады.

Динамикалық кері байланысты басқару жағдайында тұрақтандыруды қажет ететін координаттардың басталуы  $x = 0, z = 0$ ) көбінесе шығыс кері байланыс тізбектерінде кездеседі, өйткені кейбір күй айнаымалыларын өлшеудің болмауы әдетте кері байланыс контроллеріне «бақылаушыларды» немесе «бақылаушы тәрізді» компоненттерді қосу арқылы өтеледі.

Стандартты тұрақтандыру мәселесі бастапқы координаттардағы тепе-теңдік нүктесін тұрақтандыру ретінде анықталғанымен, біз жүйені ерікті нүктеге қатысты тұрақтандыру үшін бірдей тұжырымдаманы қолдана аламыз.

### 3. Күй кері байланыс арқылы басқару

#### 3.1 Орнықтандыру

Кері байланыс арқылы бар сызықтық жүйенің түрін қарастырсақ:

$$\dot{\eta} = f_0(\eta, \xi) \quad (3.1)$$

$$\dot{\xi} = A\xi + B\gamma(x)[u - \alpha(x)] \quad (3.2)$$

$$\text{мұндағы } z = \begin{bmatrix} \eta \\ \xi \end{bmatrix} = T(x) = \begin{bmatrix} T_1(x) \\ T_2(x) \end{bmatrix}$$

$T(x)$  бұл аймақтағы диффеоморфизм  $D \subset R^n, D_z = T(D)$  координаттардың басталуын қамтиды,  $(A, B)$  басқарылатын болып табылады, барлығы үшін сингулярлы емес  $x \in D, f_0(0, 0) = 0$ , және  $f_0(\eta, \xi), \alpha(x),$  және  $\gamma(x)$  үздіксіз дифференциалданады. Біздің мақсатымыз-координаттардың басталуы  $z = 0$  орнықтандыруға арналған кері байланысты басқару заңын әзірлеу болып табылады. ((3.1)-(3.2) формуласы (1)-(3) сызықтық кіріс және шығыс жүйелерінің қалыпты формуласымен айқын дәлелденген.

$$\dot{\eta} = f_0(\eta, \xi) \quad (1)$$

$$\dot{\xi} = A_c \xi + B_c \gamma(x)[u - \alpha(x)] \quad (2)$$

$$y = C_c \xi \quad (3)$$

Алайда (3) есептелмейді, өйткені  $u$  шығыс кері байланысты тұрақтандыру мәселесінде ешқандай рөл атқармайды. (3.1)-(3.2) жүйесіне (3.1) теңдеуді алып тастағанда кері байланысы бар сызықтық жүйелер де кіреді. Біз басқарылатын канондық түрде бір кіріс немесе жұп  $(A, B)$  жүйелерін қарастырумен шектелмейміз, жалпы жүйені талқылауға көшеміз (3.1)-(3.8) және біздің қорытындыларымыз қалыпты формаға (1)-(3) немесе ерекше жағдайлар ретінде сызықтық кері байланыс жүйелеріне қолданылады.

Күй кері байланыс арқылы басқару  $u = \alpha(x) + \beta(x)v$  мұндағы  $\beta(x) = \gamma^{-1}(x)$ , (3.1)-(3.2) "үшбұрышты" жүйеге дейін азайтады

$$\dot{\eta} = f_0(\eta, \xi) \quad (3.3)$$

$$\dot{\xi} = A\xi + B\vartheta \quad (3.4)$$

(13.40) теңдеу  $v = K\xi$  болса, мұндағы  $K$   $(ABC)$  Гурвиц болатындай етіп таңдап, оңай тұрақтандыруға болады. Толық тұйық жүйенің асимптотикалық тұрақтылығы.

$$\dot{\eta} = f_0(\eta, \xi) \quad (3.5)$$

$$\dot{\xi} = (A - BK)\xi \quad (3.6)$$

Координата басының асимптотикалық орнықтылығы  $\dot{\eta} = f_0(\eta, 0)$  келесі леммада көрсетіледі.

*Лемма 3.1* *Егер.  $\dot{\eta} = f_0(\eta, 0)$  координаттарының басталуы асимптотикалық тұрақты болса (3.5) - (3.6) координаттарының басталуы асимптотикалық тұрақты болып табылады.*

Дәлел: (кері Ляпунов) теоремасы бойынша Ляпуновтың үздіксіз дифференциалданатын функциясы  $V_1(\eta)$  бар  $\frac{\partial V_1}{\partial \eta} f_0(\eta, 0) \leq -\alpha_3(\|\eta\|)$  кейбір ауданда

$\eta = 0$ , мұндағы  $\alpha_3$  бұл  $K$  класының функциясы. Сол жақта  $P = P^T > 0$  Ляпунов теңдеуінің шешімі болады  $P(A - BK) + (A - BK)^T P = -I$  және  $V(\eta, \xi) = V_1(\eta) + k\sqrt{\xi^T P \xi}$  пайдалансақ,  $k > 0$  бірге, Ляпунов-кандидат функциясы ретінде (3.5)-(3.6). Туынды  $\dot{V}$  төменде келтірілген формуламен берілген.

$$\begin{aligned} \dot{V} &= \frac{\partial V_1}{\partial \eta} f_0(\eta, \xi) + \frac{k}{2\sqrt{\xi^T P \xi}} \xi^T [P(A - BK) + (A - BK)^T P] \xi \\ &= \frac{\partial V_1}{\partial \eta} f_0(\eta, 0) + \frac{\partial V_1}{\partial \eta} [f_0(\eta, \xi) - f_0(\eta, 0)] - \frac{k\xi^T \xi}{2\sqrt{\xi^T P \xi}} \end{aligned}$$

Кез-келген шектеулі ауданда координаталар басына үздіксіз дифференциалдануға  $V_1$  және  $f_0$  кейбір оң  $k_1$  және  $k_2$  тұрақтыларын қолданамыз,

$$\dot{V} \leq -\alpha_3 (\|\eta\|) + k_1 \|\xi\| - k k_2 \|\xi\|$$

$k > k_1 / k_2$  таңдау  $\dot{V}$  теріс анықталған болып табылады. Байқағанымыздай координаттардың басталуы асимптотикалық тұрақты.

Жоғарыда айтылғандардан минималды фазалық кіріс-шығысы бар сызықтық жүйені күй бойынша кері байланысты басқару арқылы тұрақтандыруға болады

$$u = \alpha(x) - \beta(x)KT_2(x) \quad (3.7)$$

(3.7) формулада басқару  $T_1(x)$  тәуелді емес. Сондықтан ол ішінара дифференциалдық теңдеуді  $\frac{\partial \phi}{\partial x} g(x) = 0$ ,  $1 \leq i \leq n - \rho$ , үшін  $\forall x \in D_0$  қанағаттандыратын  $\phi$  функцияға тәуелді емес.

Лемма 3.1 дәлелі тек шекті жиындарда жарамды. Демек, оны глобальды асимптотикалық тұрақтылықты көрсету үшін кеңейту мүмкін емес. Біз глобальды асимптотикалық тұрақтылықты  $\xi$  кіріс ретінде қарастырған кезде жүйенің  $\dot{\eta} = f_0(\eta, 0)$  кірістен күйге тұрақты болуын талап ету арқылы көрсете аламыз.

Лемма 3.2 (3.5)-(3.6) формулалары *глобальдық асимптотикалық тұрақты, егер жүйе  $\dot{\eta} = f_0(\eta, \xi)$  кіріс күйіне тұрақты болса*

Дәлелдеу: Лемманы қолдану

Кірістің күйге тұрақтылығы  $\dot{\eta} = f_0(\eta, \xi)$  4.10-бөлімде көргеніміздей, формуласы глобальды асимптотикалық, тіпті экспоненциалды тұрақтылығынан туындамайды. Демек, кіріс-шығыс сызықтық жүйенің «глобальдық» минималды фаза екенін білу басқарудың (3.7) жүйені глобальдық тұрақтандыруға автоматты түрде кепілдік бермейді.

Егер глобалды тұрақтылығы  $\dot{\eta} = f_0(\eta, 0)$  және глобалды экспоненциалды тұрақты  $f_0(\eta, \xi)$  болса, Липшиц  $(\eta, \xi)$ , бұл жағдайда 4.6 Лемма  $\dot{\eta} = f_0(\eta, 0)$  жүйенің күйге ену кезінде тұрақты болатындығын көрсетеді. Әйтпесе, біз одан әрі талдау арқылы кірістен күйге тұрақтылықты орнатуымыз керек. Ғаламдық Липшиц шарттары кейде сызықтық өсу шарттары деп аталады.

## 13.4.2 Ізге түсіру

Қалыпты түрде ұсынылған жалғыз кіріс, жалғыз шығыс, сызықтық жүйені қарастырыңыз

$$\dot{\eta} = f_0(\eta, \xi)$$

$$\dot{\xi} = A_c \xi + B_c \gamma(x) [u - \alpha(x)]$$

$$y = C_c \xi$$

біз жалпылықты жоғалтпай,  $f_0(0,0) = 0$  деп есептейміз.  $Y$  шығыс сигналын кері байланысты сызықтық жүйелердің күй сигналын асимптотикалық түрде тірек сигнал  $r(t)$  ізіне түсіру басқару заңын жасағымыз келеді. Жүйе салыстырмалы  $P = n$ , дәрежесіне ие болған кезде, оның тривиальды емес нөлдік динамикасы болмайды. Жүйенің салыстырмалы дәрежесі болған кезде де оның тривиальды емес нөлдік динамикасы болмайды. Бұл жағдайда айнымалы және оның теңдеуі алынып тасталады, бірақ қалған жағы  $\eta$  өзгеріссіз қалады, төменде келтірілгендей болжауға болады:

$r(t)$  дейін және оның туындылары  $r^{(p)}(t)$  барлығына шектелген  $t \geq 0$  және  $\rho$  туындысы  $r^{(\rho)}(t)$ ,  $t$  - дан үздіксіз функция болып табылады;

Сигналдар  $r, \dots, r^{(\rho)}$  онлайн режимде тиімді;

Көбінесе сандық есептеулерде бақылау процесінің өз деңгейінде орындалуын анықтау барысында тірек сигналы (*опорный сигнал – reference signal*) пайдаланылады. Тірек сигналы – жүйенің барлық негізгі элементтерін синхрондау іске асырылатын мерзімді сигнал. Бұрынғы жарық көрген еңбектерде (Huang J. et al., 1990), (Herburn J et al., 1984) тірек сигналы орнына тұрақты шама алынған.

Тірек сигнал  $r(t)$  өзінің туындыларымен бірге белгілі бір уақыт функциясы ретінде берілуі мүмкін немесе бұл кейбір кіріс сигналымен басқарылатын тірек модельдің шығыс сигналы  $\omega(t)$  болуы мүмкін. Соңғы жағдайда  $r$  туралы болжамдар тірек сигналдың моделін дұрыс таңдау арқылы орындалуы мүмкін.

Соңғы жағдай  $r$  бойынша жорамалдарды анықтамалық үлгіні дұрыс таңдау арқылы орындауға болады. Мысалы, салыстырмалы екінші дәрежелі жүйе үшін тірек сигналдың моделін тасымалдау функциясымен ұсынылған екінші ретті сызықтық уақыт - инварианттық жүйе болуы мүмкін.

$$\frac{\omega_n^2}{s^2 + 2\zeta\omega_n s + \omega_n^2}$$

мұнда оң тұрақтылар  $\zeta$  берілген кіріс сигналымен  $w(t)$  тірек сигналды қалыптастыру үшін таңдалады. Тірек сигнал  $r(t)$  нақты уақыт режимінде күй моделін қолдана отырып жасалуы мүмкін

$$\dot{y}_1 = y_2$$

$$\dot{y}_2 = -\omega_n^2 y_1 - 2\zeta\omega_n y_2 + \omega_n^2 w$$

$$r = y_1$$

Сондықтан,  $r(t)$ ,  $\dot{r}(t)$ , және  $\ddot{r}(t)$  онлайн режимде тиімді. Егер  $w(t)$  үздіксіз шектелген функция,  $t$  болғаннан кейін  $r(t)$ ,  $\dot{r}(t)$ , және  $\ddot{r}(t)$  қажетті болжамдарды қанағаттандырады.

$$\mathfrak{R} = \begin{bmatrix} r \\ \vdots \\ r^{(p-1)} \end{bmatrix}, \quad e = \begin{bmatrix} \xi_{1-r} \\ \vdots \\ \xi_{p-r^{(p-1)}} \end{bmatrix} = \xi - \mathfrak{R}$$

Айнымалыларды ауыстырғанда  $e = \xi - \mathfrak{R}$

$$\dot{\eta} = f_0(\eta, e + \mathfrak{R})$$

$$\dot{e} = A_c e + B_c \{ \gamma(x)[u - \alpha(x)] - r^{(p)} \}$$

Күй кері байланыс арқылы басқару  $u = \alpha(x) + \beta(x)[v + r^{(p)}]$  мұндағы  $\beta(x) = 1/\gamma(x)$ , қалыпты форманы каскадтық жүйеге дейін азайтады  $\dot{\eta} = f_0(\eta, e + \mathfrak{R})$

$$\dot{e} = A_c e + B_c \mathfrak{G}$$

Біздің мақсатымыз екінші теңдеуді тұрақтандыратын және бәріне шектеулі  $t \geq 0$  болатынына көз жеткізу.  $\mathfrak{G} = -Ke$ , мұнда  $A_c - B_c K$  Гурвиц болса, күйдің толық кері байланысын басқару келесі формула арқылы сипатталады:

$$u = \alpha(x) + \beta(x) \{ -K[T_2(x) - \mathfrak{R} + r^{(p)}] \} \quad (3.7)$$

ал тұйықталған жүйе мынадай түрде

$$\dot{\eta} = f_0(\eta, e + \mathfrak{R}) \quad (3.8)$$

$$\dot{e} = (A_c e + B_c K)e \quad (3.9)$$

Минималды фазалық жүйелер арналған бастапқы  $\dot{\eta} = f_0(\eta, 0)$  асимптотикалық болады. Бұл (кері Ляпунов функциясы) теоремасынан шығатыны жеткілікті  $e(0), \eta(0)$ , және  $\mathfrak{R}(t)$ , күй  $\eta(t)$  барлығына шектелетін болады  $t \geq 0$ . Осылайша, күй кері байланыс басқару (3.7) формулада локальды ізге түсіру мәселесін шешеді. Басқарудың жарамдылығын глобалды бақылауға дейін кеңейту үшін, мұнда кез келген шектелген  $\mathfrak{R}(t)$  функциясы болуы мүмкін, біз глобалды тұрақтандыру кезінде кездесетін мәселелерге тап боламыз. Глобалды ізге түсіруді қамтамасыз етудің жеткілікті шарты жүйенің күй тұрақтылығын  $\dot{\eta} = f_0(\eta, \xi)$  енгізу болып табылады. (Khalil H.K., 2002).

#### 4. Кері байланысты сызықтық басқаруды жобалау

Екінші ретті сызықтық емес жүйенің күйі:

$$\begin{cases} \dot{x}_1 = x_2 \\ \dot{x}_2 = f(x) + g(x)u \end{cases} \quad (4.1)$$

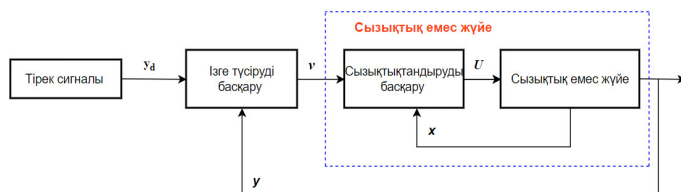
мұндағы  $x = [x_1, x_2]^T$  - күй векторы;  $u$  - басқару сигналы,  $y = x_1$  - шығыс сигналы,  $f(x)$  - тегіс сызықты емес функция,  $g(x)$  бұл тегіс сызықты емес және нөлдік емес функция.

Басқарудың мақсаты  $e(t) = y(t) - y_d(t)$ , нөлге жақындаған  $\lim_{t \rightarrow +\infty} e(t) = 0$  шығыс ізіне түсіру қателігін анықтау болып табылады.

Кері байланысты сызықтық басқару схемасы сызықтық басқаруды және тірек сигнал ізге түсіруді басқаруды қамтиды.

Сызықтықтандыруды басқару сызықтық емес жүйені сызықтық жүйеге түрлендіру үшін қолданылады.

Ізге түсіруді басқару теориялық сызықтық басқару негізінде жасалады.



1 сурет-Кері байланысты басқару құрылымының схемасы (Fig. 1. Feedback control structure scheme)

1-қadam: Шығыс сигналын тірек ізіне түсіру қатесінен туындыларды уақыт бойынша көшіру, ізіне түсу қателігі:  $e = y_d - y$  (4.2)

Ізіне түсу қателігінен уақыт бойынша 1-ші ретті туынды:

$$\dot{e} = \dot{y}_d - \dot{y} = \dot{y}_d - x_2 \quad (4.3)$$

Ізіне түсу қателігінен уақыт бойынша 2-ші ретті туынды:

$$\ddot{e} = \ddot{y}_d - \dot{x}_2 = \ddot{y}_d - f(x) - g(x)u \quad (4.4)$$

2-қadam: Сызықтық басқаруды жобалау:  $u = \frac{1}{g(x)}(-f(x) + v)$  (4.5)

(5) - ті (4) - ке ауыстырғандағы нәтиже келесідей болады:  $\ddot{e} = \ddot{y}_d - v$  (4.6)

3-қadam: Ізіне түсіруді басқаруды жобалау:  $v = \ddot{y}_d + k_1\dot{e} + k_2e$  (4.7)

4-қadam: Басқару параметрлерін таңдасак:

(6) орнына (7) қойсак, келесі түрде көрсетіледі:  $\ddot{e} = -k_1\dot{e} - k_2e$  (4.8)

Сипаттамалық теңдеу арқылы өрнектеуге болады  $s^2 + k_1s + k_2 = 0$  (4.9)

Қайта реттеу, белгіленген режим уақыты немесе тұрақты режим қатесі негізінде қажетті сипаттамалық теңдеу алынуы мүмкін. Екі сипаттамалық теңдеуді салыстыру арқылы басқару параметрлерін анықтауға болады.

5-қadam: Тірек сигналға арналған төмен жиілік сүзгісін жасау төменде келтірілген:

$$G_{Lf}(s) = \frac{1}{(0.1s + 1)^2} \quad (4.10)$$

Маятникті қарастырайық:  $ml^2\ddot{q} + B\dot{q} + mgl\sin(q) = u$  (4.11)

Мұндағы  $m=0.1$  (kg),  $l=1$  (m),  $B=0.01$  (Nms/rad)

Маятник үшін кері байланысты сызықтық басқаруды жобалауда  $\text{РОТ} < 10\%$  және  $t_{\text{qd}} < 0.3$  шығыс сипаттамаларына қол жеткізу қамтамасыз етіледі. Тірек сигнал импульстік сигнал болып табылады.

**Шешім:**

Маятник жүйесінің айнымалы күйін анықтасак:  $x_1 = q, x_2 = \dot{q}$  және

шығыс  $y = q = x_1$  болса, (4.1-4.4) формулаларын колдансақ, мұндағы

$$f(x) = -\frac{g}{l} \sin(x_1) - \frac{B}{ml^2} x_2, g(x) = \frac{1}{ml^2}$$

Өзірленген талаптарға сүйене отырып:  $POT = \exp\left(\frac{-\xi\pi}{\sqrt{1-\xi^2}}\right) < 0.1$

$$\rightarrow \xi > 0.59 \rightarrow \xi = 0.7$$

$$t_{qd} = \frac{4}{\xi w_n} < 0.3$$

$$\rightarrow w_n > 19.05 \rightarrow w_n = 25$$

Қажетті сипаттаманың тұжырымдамасы келесідей  $s^2 + 2\xi w_n s + w_n^2 = 0$

$$s^2 + 35s + 625 = 0$$

$$(4.12)$$

(4.8) және (4.9) формулаларды есептеу арқылы басқару параметрлері келесідей таңдалады  $k_1 = 35, k_2 = 625$ .

5-қадам: төмен жиілікті фильтр жасасақ:  $G_{LF}(s) = \frac{1}{(0.1s + 1)^2}$  (4.13)

**Нәтиже**

Matlab ортасының Simulink ортасында бекітілген қадам өлшемі ретінде 0,01 секунд алынды, ал модельдеу уақыты: 20 секунд болып табылады.

Келесі маятниктің сызықтық емес моделін

$f(x) = -\frac{g}{l} \sin(q) - \frac{B}{ml^2} \dot{q}, g(x) = \frac{1}{ml^2}$  колданып, сызықтық басқаруды жобалауға арналған  $u = g^{-1}(x)(-f(x) + v)$ , мұндағы

$x_1 - \theta$  бұрыштық позиция [rad];

$\dot{x}_2 - \dot{\theta}$  бұрыштық позиция [rad/s];

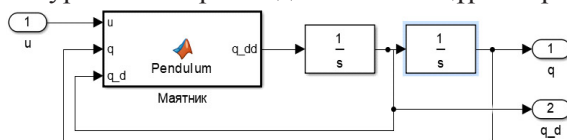
$m$ -маятник массасы [kg];  $m=0,1$ ;

$l$ -маятник ұзындығы [m];  $l=1$ ;

$k$ -үйкеліс коэффициенті [Nm/rad/s];

$g$ -гравитациялық үдеу [m/c<sup>2</sup>].  $g=10$ ;

$B=0.01$  болса, 3- суретте келтірілгендей маятник құрастырылды.



2 сурет. Simulink-те маятник блогының құрастырылған схемасы (Fig. 2. Built-in scheme of the pendulum block in Simulink)

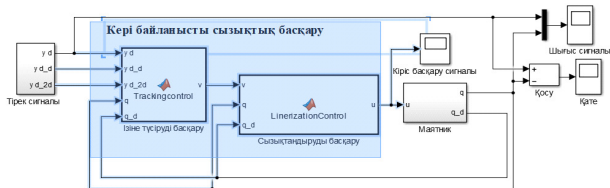
Ізіне түсіруді басқаруды жобалау:  $v = \ddot{y}_d + k_1 \dot{e} + k_2 e$  болса, онда  $k_1=35, k_2=625$  тең болса, онда 3 суретте Simulink-те тірек блогын құрастырамыз.



3 сурет. Simulink-те тірек блогының құрастырылған схемасы (Fig. 3. Compiled scheme of the support block in Simulink)

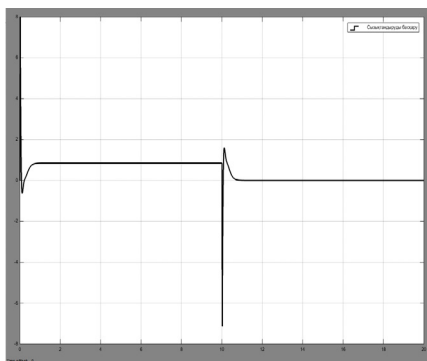


Динамикалық жүйе контроллерлерінің шығысын қолданып, кері байланысты сызықтандырудың мысалы тұйық жүйе контурын имитациялайтын Simulink Library Browser көмегімен маятник, тірек сигналы блоктары жалғанды. Ол төмендегі 4-суретте көрсетілген.



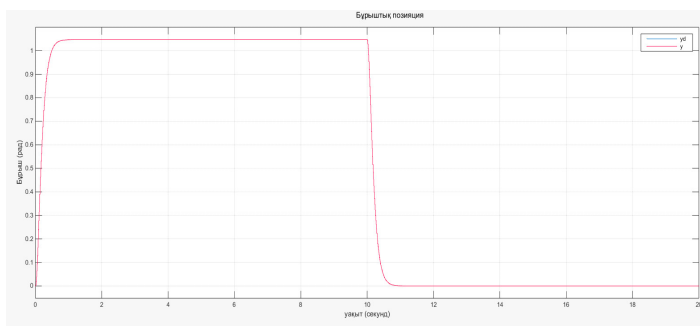
4 сурет. MATLAB Simulink-те жалпы модельдеу құрылымы  
(Fig. 4. General structure of modeling in MATLAB Simulink)

U сызықтандыруды басқару сигналынан кейін төменде 4-суретте кіріс сигналы немесе басқару моделі төмендегі суретте келтірілген.



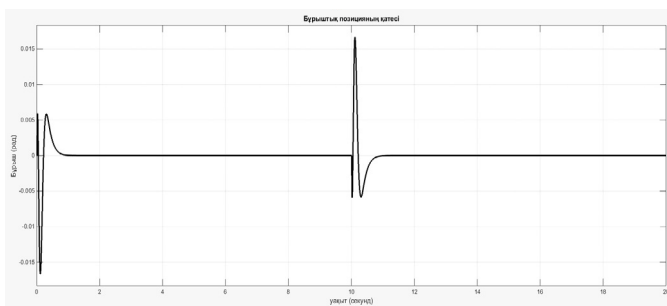
5 сурет. Кіріс сигналы немесе басқару траекториясы  
(Fig. 5. Input signal or control trajectory)

Келесі 6-суретте шығыс сигналының жауабы қызғылт түс болса, көк түс тірек сигналы келтірілген.



6 сурет. Шығыс сигналының траекториясы  
(Fig. 6. Output signal trajectory)

Келесі кездейсоқ қателік 7-суретте көрсетілген.



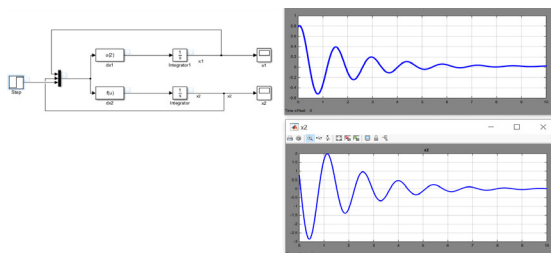
7 сурет. Кездейсоқ қателік траекториясы  
(Fig. 7. Trajectory of random error)

### Талқылау

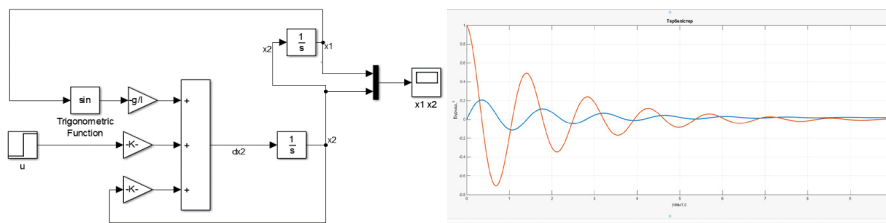
Маятниктің сызықтық емес моделі

$$\begin{cases} \dot{x} = x_2 \\ \ddot{x} = -\frac{g}{l} \sin(x_1) - \frac{k}{ml^2} + \frac{1}{ml^2} u \end{cases} \quad (1)$$

базалық, функционалды блоктар мен интеграторлар арқылы MATLAB Simulink-тің негізгі блоктары арқылы модельденуі төменде келтірілген суреттерде келтірілген (Kwadzogah et al., 2014).



8 сурет. Функционалды блоктар мен интеграторлар арқылы құрылған схемасы және нәтижесі  
(Fig. 8. Scheme and result created by functional blocks and integrators)



9 сурет. Базалық блоктар мен интеграторлар арқылы құрылған схемасы және нәтижесі  
(Fig. 9. Scheme and result created by functional blocks and integrators)

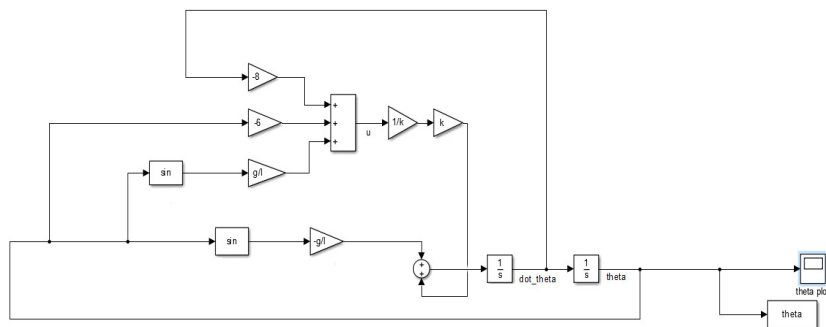
Кері байланысты сызықтандыру және бұл әдісті динамикалық жүйелердің контроллерлерін анық түрде алу үшін қалай пайдалануға болатынының және тұйық жүйені модельдейтін Simulink блок-схемалары 10 суретте берілген. Әрі

қарай, біз кері байланысты сызықтық басқару заңын модельдейміз. Басқару параметрлері  $c_0 = 8$ ,  $c_1 = 6$  келесі басқару заңын төмендегідей анықталады.

$$u = \frac{1}{k} \left( \frac{g}{l} \sin(x_1) - 8x_1 - 6\dot{x}_2 \right)$$

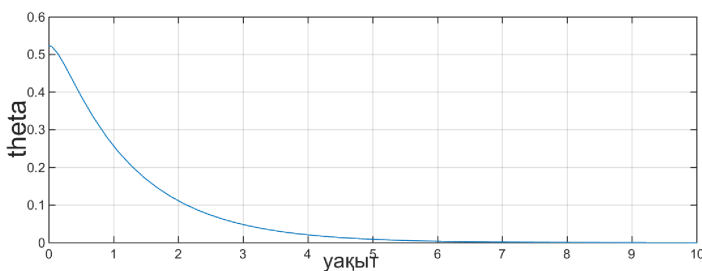
$$u = \frac{1}{k} \left( \frac{g}{l} \sin(\theta) - 8\theta - 6\dot{\theta} \right) \tag{2}$$

(2) формуланы, яғни басқару заңына сәйкес келетін кері байланысты сызықтық басқару заңына сәйкес келетін құрылымдық сызбасы төменде 9 суретте келтірілген.



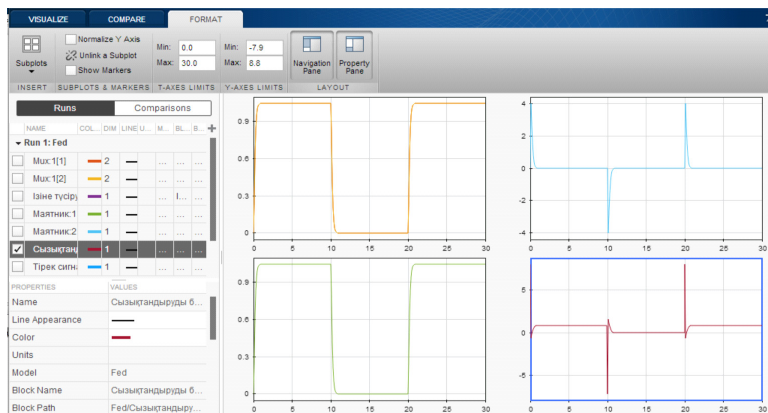
10 сурет. Тұйықталған жүйе контурының құрылымдық сызбасы (Fig. 10. Structural diagram of a closed system)

Жүйенің реакциясы бұл жұмыста тұйық сызықтық емес қарапайым жүйені асимптотикалық орнықтандыруды ең негізгі әдіс кері байланыс әдісі арқылы орнықтандыру мысалы компьютерде модельдеу есебі (theta айнымалысы) төменде 10-суретте келтірілген (Bahadirova G. et al.,2023).



11 сурет.Тұйық жүйеде асимптотикалық түрде кері байланыспен сызықтық емес басқару заңымен орнықтандырылу моделі (Fig. 11. Stability model with nonlinearized control law with feedback asymptotically in a closed system)

11 суретте жоғарыда мақалада келтірілген сызықтық емес жүйе үшін кері байланысты сызықтық басқаруды жобалау процесі Simulation Data Inspector арқылы ашқанда көрінісі келтірілген.



12 сурет. Simulation Data Inspector терезесі  
(Fig. 12. Simulation Data Inspector window)

## Қорытынды

Екінші ретті сызықтық емес жүйе үшін жалпы тұжырымдамада кері байланыспен сызықтық басқару жасалды. Маятниктік жүйе үшін MATLAB Simulink-те компьютерлік модельдеу жүргізілді. Көрсетілген тәсіл бір кірісі және бір шығысы бар сызықтық емес жүйе үшін жүзеге асырылады. Болашақта басқару жүйелерін жобалау кезінде кейбір көп кіріс және көп шығысы бар сызықтық емес жүйелері қарастырылады.

## ӘДЕБИЕТТЕР

Алимхан А., Инаба Н. (2008). Анықтамлаған сызықты емес жүйелер класы үшін шығыс компенсаторы арқылы сенімді практикалық шығыс басқару. Модельдеу, анықтау және басқару, халықаралық журналы. 4: 304–314. DOI:10.1504/IJMIC.2008.021470 (in Eng.).

Алимхан К., Тасболатұлы Н., Ерденова А. (2021). Анықталмаған сызықтық емес жүйелердің шығыс мәліметтерін кері байланыс күйі арқылы бақылау. Теориялық және қолданбалы ақпараттық технологиялар журналы, 99(13): 3337–3352 <https://doi.org/10.15199/48.2019.05.22> (in Eng.).

Бахадирова Г.Б., Тасболатұлы Н. (2023). Simulink-те кері байланыс жәрдемінде сызықтық емес жүйені басқаруды модельдеу. АХУ-2023: Талдау, Жаңалық енгізу. Қолдану атты конференция. 1970–1977 б. (in Kaz.).

Бирнс С.И., Делли Присколи Ф., Исидори А. (1997). Анықталмаған сызықты емес жүйелердің шығысын басқару, Биркхаузер. ISBN 978-1-4612-7384-4. DOI 10.1007/978-1-4612-2020-6 (in Eng.).

Буфаден М. (2018). MATLAB көмегімен сызықтық емес жүйелерді басқару. СРС Пресс, Нью-Йорк. ISBN: 978-1-138-35955-0. (in Eng.).

Клелланд Дж., Клотц Т., Василиу П. (2021). Динамикалық кері байланыстың басқару жүйелерін симметриялық түрде сызықтандыру. Алдын ала басып шығару Архив:2103.05078. <https://doi.org/10.48550/arXiv.2103.05078> (in Eng.).

Гонг Q., Цянь С. (2005). Шығыс кері байланысын пайдалана отырып, сызықты емес жүйелер класының шығысын глобалдық практикалық басқару. IEEE-нің 44-ші конференциясының шешімдері мен бақылауы бойынша іс жүргізуі және Еуропалық бақылау конференциясы. 7278–7283 б.

Хуан Дж., Руг В.Дж. (1990). Сызықты емес көп айнымалы сервомеханизм мәселесі бойынша. Автоматика. 26 (6): 963–972. (in Eng.).

Хепберн Дж., Вонхам У.А. (1984). Кері байланыс және дифференциалданатын коллекторлардағы ішкі модель және қателік. *IEEE Транс. Автоматты. Бақылау*. 29 (5): 397–403. (in Eng.).

Хосе де Хесус Рубио (2018). Сызықты емес процестерді басқару үшін кері байланысты сенімді сызықтандыру. *ISA транзакциялары*. 74: 155–164. DOI:10.1016/j.isatra.2018.01.017 (in Eng.).

Халил Х.К. (2002). Сызықтық емес жүйелер. Нью-Джерси: Жоғарғы Садль, 750 б, ISBN:0-13-067389-7. (in Eng.).

Квадзога, Роджер Кобла (2014). Күйге тәуелді тасымалдау функциялары бар -аффиндік сызықты емес жүйелерді басқаруды модельдеу, басқару және модельдеу. *Диссертациялар*. 164. (in Eng.).

Ляо Х.Х., Ван Л.К., Ю.П. (2007). Динамикалық жүйелердің тұрақтылығы. Эльзевир, ISBN: 9780080550619. (in Eng.).

Тасболатұлы Н. (2020). Жоғары ретті анықталмаған сызықтық емес жүйелерге кең ауқымды практикалық бақылау және ол үшін бағдарламалық кешен құру: филос.докт.(PhD) ... дис. – Алматы (in Kaz.).

Бу М., Хэ Ю., Ше Ж.Х. (2010). Уақытты кешіккен жүйелерінің тұрақтылығын талдау және сенімді басқару. – Пекин: Ғылыми баспасөз, 335 б. DOI:10.1007/978-3-642-03037-6 (in Eng.).

## REFERENCES

Alimhan K., Inaba H. (2008). International Journal of Modelling, Identification and Control Robust practical output tracking by output compensator for a class of uncertain inherently non-linear systems. 4. – Pp. 304–314. DOI:10.1504/IJMIC.2008.021470 (in Eng.).

Alimhan K., Tasbolatuly N., Yerdenova A. (2021). Global output tracking control for high-order non-linear systems with time-varying delays, *Journal of Theoretical and Applied Information Technology*, 99(13): 3337–3352 <https://doi.org/10.15199/48.2019.05.22> (in Eng.).

Bahadirova G.B., Tasbolatuly N. (2023). Modeling control of a nonlinear system using feedback in Simulink. *AIU-2023: Analyze, Innovate. Conference called Use*. -Pp 1970–1977. (in Kaz.).

Byrnes C.I., Delli Priscoli F., Isidori A. (1997). Output regulation of uncertain nonlinear systems. Birkhauser, ISBN 978-1-4612-7384-4. DOI 10.1007/978-1-4612-2020-6 (in Eng.).

Boufadene M. (2018). *Nonlinear Control Systems Using MATLAB®*. – CRC Press, New York. ISBN: 978-1-138-35955-0. (in Eng.).

Clelland J., Klotz T., Vassiliou P. (2021). Dynamic Feedback Linearization of Control Systems with Symmetry. *arXiv preprint arXiv:2103.05078*. <https://doi.org/10.48550/arXiv.2103.05078> (in Eng.).

Gong Q., Qian C. (2005). Global Practical Output Regulation of a Class of Nonlinear Systems by Output Feedback, *Proceedings of the 44th IEEE Conference on Decision and Control, and the European Control Conference*. -Pp. 7278–7283. (in Eng.).

Huang J., Rugh W.J. (1990). On a non-linear multivariable servomechanism problem. *Automatica*. 26 (6). –Pp. 963–972. (in Eng.).

Hepburn J., Wonham W.A. (1984). Error feedback and internal model on differentiable manifolds. *IEEE Trans. Automat. Control*. 29 (5): –Pp. 397–403. (in Eng.).

Jose de Jesus Rubio (2018). Robust feedback linearization for nonlinear processes control. *ISA transactions* 74: 155–164. DOI:10.1016/j.isatra.2018.01.017 (in Eng.).

Khalil H.K. (2002). *Nonlinear systems*. New Jersey.: Upper Saddle River. -P. 750. ISBN:0-13-067389-7. (in Eng.).

Kwadzogah, Roger Kobla (2014). Modeling, control and simulation of control-affine nonlinear systems with state-dependent transfer functions. *Dissertations*. 164. (in Eng.).

Liao X.X., Wang L.Q., Yu P. (2007). *Stability of Dynamical Systems*. London: Elsevier, ISBN: 9780080550619. (in Eng.).

Tasbolatuly N. (2020). Large-scale practical control of indeterminate nonlinear systems of higher order and the creation of a software complex for it. *Almaty. Dissertations*. (in Kaz.).

Wu M., He Y., She J.H. (2010). *Stability Analysis and Robust Control of Time- Delay Systems*. – Beijing: Science Press. -P. 335. DOI:10.1007/978-3-642-03037-6 (in Eng.).

NEWS OF THE NATIONAL ACADEMY OF SCIENCES OF THE REPUBLIC OF  
KAZAKHSTAN  
PHYSICO-MATHEMATICAL SERIES  
ISSN 1991-346X  
Volume 3. Number 347 (2023). 62–75  
<https://doi.org/10.32014/2023.2518-1726.204>

UDC 004.057A

© **Y.S. Golenko, A.A. Ismailova, 2023**

NCJSC «S. Seifullin Kazakh AgroTechnical Research University»,  
Astana, Kazakhstan.

E-mail: [golenko.katerina@gmail.com](mailto:golenko.katerina@gmail.com)

## **PROTEIN FUNCTION PREDICTION USING THE COMBINATION OF BILSTM AND SELF-ATTENTION ALGORITHM**

**Golenko Yekaterina Sergeevna** — Master of Science, PhD candidate, NCJSC «S. Seifullin Kazakh AgroTechnical Research University», Astana, Kazakhstan

E-mail: [golenko.katerina@gmail.com](mailto:golenko.katerina@gmail.com); ORCID ID: <https://orcid.org/0000-0002-4643-4571>;

**Ismailova Aisulu Abzhapparovna** — PhD, associate professor, NCJSC «S. Seifullin Kazakh AgroTechnical Research University», Astana, Kazakhstan

E-mail: [a.ismailova@mail.ru](mailto:a.ismailova@mail.ru); ORCID ID: <https://orcid.org/0000-0002-8958-1846>.

**Abstract.** With the development of genome sequencing technology, the use of computational technologies to predict the function of proteins has become one of the important tasks of bioinformatics. Early research in this area was based on sequence similarity and assumed that proteins with similar amino acid sequences had similar functions. However, previously proposed methods for predicting functions often failed to reveal hidden patterns between proteins and gene ontology terms, which reduced the accuracy of functional annotation. Deep machine learning, as many studies show, copes with this task at a higher level. First, deep learning methods can be trained on large amounts of protein sequence data without considering additional information about protein properties. Secondly, deep learning approaches solve such side problems as data noisiness, their redundancy and high dimensionality. The combination of a self-attention mechanism and a bidirectional network with long-term short-term memory can be used to solve the problem of protein functional annotation. Bidirectional LSTM is used to obtain both global and local information about the properties of protein sequences, as well as to store the information obtained. The self-attention algorithm is applied to make optimal use of the relationship of the sequence and information about the functions of different positions of the sequence, which will increase the reliability of the prediction. The python language was chosen as a tool for implementing the algorithms, the model was trained for 50 epochs and tested on an experimental dataset of the Indica protein obtained from open sources. The results of the experiment show that the algorithm for combining

the self-attention mechanism and the bidirectional network with long-term short-term memory outperforms other traditional neural network algorithms and can more accurately predict protein function, which shows the possible applicability of the algorithm in functional annotation of protein sequences.

**Keywords:** bidirectional LSTM, self-attention, function prediction, proteins, machine learning

© **Е.С. Голенко, А.А. Исмаилова, 2023**

«Сәкен Сейфуллин атындағы Қазақ агротехникалық зерттеу университеті»  
КеАҚ, Астана, Қазақстан.

E-mail: [golenko.katerina@gmail.com](mailto:golenko.katerina@gmail.com)

## ПРЕДСКАЗАНИЕ ФУНКЦИЙ БЕЛКА С ИСПОЛЬЗОВАНИЕМ КОМБИНАЦИИ VILSTM И АЛГОРИТМА САМОВНИМАНИЯ

**Голенко Екатерина Сергеевна** — техника ғылымдарының магистрі, PhD докторанты, «Сәкен Сейфуллин атындағы Қазақ агротехникалық зерттеу университеті» КеАҚ, Астана, Қазақстан Республикасы

E-mail: [golenko.katerina@gmail.com](mailto:golenko.katerina@gmail.com); ORCID ID: <https://orcid.org/0000-0002-4643-4571>;

**Исмаилова Айсулу Абжаппаровна** — PhD, ассоциированный профессор, «Сәкен Сейфуллин атындағы Қазақ агротехникалық зерттеу университеті» КеАҚ, Астана, Қазақстан Республикасы

E-mail: [a.ismailova@mail.ru](mailto:a.ismailova@mail.ru); ORCID ID: <https://orcid.org/0000-0002-8958-1846>.

**Аннотация.** Геномды секвенирлеу технологиясының дамуымен белоктардың қызметін болжау үшін есептеу технологияларын қолдану биоинформатиканың маңызды міндеттерінің біріне айналды. Бұл саладағы алғашқы зерттеулер дәйектілік ұқсастығына негізделген және ұқсас аминқышқылдары тізбегі бар белоктардың ұқсас функциялары бар деп болжаған. Дегенмен, функцияларды болжау үшін бұрын ұсынылған әдістер көбінесе белоктар мен гендік онтология терминдері арасындағы жасырын заңдылықтарды аша алмады, бұл функционалдық аннотацияның дәлдігін төмендетті. Терең машиналық оқыту, көптеген зерттеулер көрсеткендей, бұл тапсырманы жоғары деңгейде женеді. Біріншіден, белок қасиеттері туралы қосымша ақпаратты есепке алмай-ақ, ақуыз тізбегі деректерінің үлкен көлемі бойынша терең оқыту әдістерін үйретуге болады. Екіншіден, терең оқыту тәсілдері деректердің шулылығы, олардың артықтығы және жоғары өлшемділігі сияқты жанама мәселелерді шешеді. Ұзақ мерзімді қысқа мерзімді жады бар өзіне-өзі назар аудару механизмі мен қос бағытты желі комбинациясы белоктың функционалдық аннотациясының мәселесін шешу үшін пайдаланылуы мүмкін. Екі бағытты LSTM протеин тізбегінің қасиеттері туралы ғаламдық және жергілікті ақпаратты алу үшін, сондай-ақ алынған ақпаратты сақтау үшін қолданылады. Өзіндік назар аудару алгоритмі реттілік қатынасын және реттіліктің әртүрлі позицияларының функциялары туралы ақпаратты оңтайлы пайдалану үшін қолданылады, бұл болжамның

сенімділігін арттырады. Алгоритмдерді енгізу құралы ретінде питон тілі таңдалды, модель 50 дәуір бойы оқытылды және ашық көздерден алынған Indica протеинінің тәжірибелік деректер жинағында сынақтан өтті. Тәжірибе нәтижелері көрсеткендей, өзіне-өзі назар аудару механизмі мен екі бағытты желіні ұзақ мерзімді қысқа мерзімді жадымен біріктіру алгоритмі басқа дәстүрлі нейрондық желі алгоритмдерінен асып түседі және алгоритмнің мүмкін болатын қолдану мүмкіндігін көрсететін ақуыз функциясын дәлірек болжауға болады. ақуыз тізбектерінің функционалдық аннотациясында.

**Түйін сөздер:** екі бағытты LSTM, өзіндік зейін, мүмкіндіктерді болжау, ақуыздар, машиналық оқыту

© **Е.С. Голенко, А.А. Исмаилова, 2023**

НАО «Каззахский агротехнический исследовательский университет имени Сакена Сейфуллина», Астана, Казахстан.

E-mail: [golenko.katerina@gmail.com](mailto:golenko.katerina@gmail.com)

## **ПРЕДСКАЗАНИЕ ФУНКЦИЙ БЕЛКА С ИСПОЛЬЗОВАНИЕМ КОМБИНАЦИИ VILSTM И АЛГОРИТМА САМОВНИМАНИЯ**

**Голенко Екатерина Сергеевна** — магистр технических наук, выпускник докторантуры, НАО «Каззахский агротехнический исследовательский университет имени Сакена Сейфуллина», Астана, Казахстан

E-mail: [golenko.katerina@gmail.com](mailto:golenko.katerina@gmail.com); ORCID ID: <https://orcid.org/0000-0002-4643-4571>;

**Исмаилова Айсулу Абжаппаровна** — PhD, ассоциированный профессор, НАО «Каззахский агротехнический исследовательский университет имени Сакена Сейфуллина», Астана, Казахстан

E-mail: [a.ismailova@mail.ru](mailto:a.ismailova@mail.ru); ORCID ID: <https://orcid.org/0000-0002-8958-1846>.

**Аннотация.** С развитием технологии секвенирования генома использование вычислительных технологий для прогнозирования функции белков стало одной из важных задач биоинформатики. Ранние исследования в этой области основывались на сходстве последовательностей и предполагали, что белки со схожими аминокислотными последовательностями имеют схожие функции. Однако предложенные ранее методы прогнозирования функций часто не могли выявлять скрытые закономерности между белками и терминами геной онтологии, что понижало точность функционального аннотирования. Глубинное машинное обучение, как показывает множество исследований, справляется с этой задачей на более высоком уровне. Во-первых, методы глубинного обучения могут обучаться на больших объемах данных белковых последовательностей, не принимая во внимание дополнительную информацию о свойствах белков. Во-вторых, подходы глубинного обучения решают такие побочные задачи как зашумленность данных, их избыточность и высокая размерность. Комбинирование механизма самовнимания и двунаправленной сети с долговременной краткосрочной памятью может быть использовано для решения проблемы функционального



аннотирования белка. Двухнаправленная LSTM используется для получения как глобальной, так и локальной информации о свойствах белковых последовательностей, а также для сохранения полученной информации. Алгоритм самовнимания применяется для оптимального использования взаимосвязи последовательности и информации о функциях различных позиций последовательности, что повысит надежность прогнозирования. В качестве инструмента для реализации алгоритмов был выбран язык python, модель обучена в течение 50 эпох и протестирована на экспериментальном наборе данных белка Indica, полученного из открытых источников. Результаты эксперимента показывают, что алгоритм комбинирования механизма самовнимания и двухнаправленной сети с долговременной краткосрочной памятью превосходит другие традиционные алгоритмы нейронных сетей и может более точно прогнозировать функцию белка, что показывает возможную применимость алгоритма в функциональном аннотировании белковых последовательностей.

**Ключевые слова:** двухнаправленная LSTM, самовнимание, предсказание функций, белки, машинное обучение

### **Введение**

Предсказание функций белка является серьезной проблемой в области биоинформатики. С развитием технологии секвенирования основным методом прогнозирования функций белка являлся метод биологических экспериментов, который требовал большого количества материальных ресурсов и времени. Увеличивающаяся скорость роста данных о последовательностях белка сделала ручную аннотацию неконкурентоспособной (Саидния и др., 2015), а вычислительные методы аннотации стали основными в области прогнозирования функций белка (Цзян и др., 2017).

Ранние вычислительные методы использовали BLAST, PSI-BLAST, FASTA и другое программное обеспечение для поиска похожих последовательностей каждого белка в обучающем наборе, а затем предполагали, что аналогичные последовательности имеют схожие функции (Гиллис и др., 2013), и мигрировали аннотации функций белков. С развитием искусственного интеллекта многие методы машинного обучения стали широко применяться для прогнозирования белковых функций. Например, SVM-Prot (Цай и др., 2003) использовал состав и трансформацию белка, особенности распределения и алгоритм SVM для прогнозирования функций белка. ProMK (Ю и др., 2015) объединил алгоритм KN с пятью различными методами измерения расстояний между характеристическими значениями для прогнозирования функций белка в различных наборах данных. Многие другие исследователи использовали различные методы машинного обучения для предсказания функций белка и добились хороших результатов, таких как совместное обучение (Нам и др., 2005), наивная байесовская модель (Юсеф и др., 2008), случайный лес (Чен и др., 2005) и другие.

Однако неглубокие методы прогнозирования функций белков часто затрудняют выявление глубоких (нелинейных) взаимосвязей между белками и функциональными терминами Gene Ontology (GO). По сравнению с традиционными методами машинного обучения, методы глубокого обучения могут обучаться на массивных данных о последовательностях белков без разработки признаков. Пока данные аминокислотной последовательности просто обрабатываются, их можно напрямую вводить в нейронную сеть для обучения. Методы глубокого обучения решают проблемы, которые трудно было решить с помощью традиционных алгоритмов машинного обучения в прошлом, такие как высокая размерность, избыточность и высокий уровень шума, вызванные массивными данными о последовательностях белков.

DeepGO (Кулманов и др., 2018) как одна из первых моделей глубокого обучения использовала алгоритм сверточной нейронной сети (CNN) для прогнозирования функции белка с использованием различных наборов данных. Это был алгоритм предсказания функций белков по последовательностям phstein и сетям PPI. На основе алгоритма DeepGO был разработан DeepGOPlus (Кулманов и др., 2019) для прогнозирования функции белка только по аминокислотным последовательностям, в котором модель CNN сочетается с методом BLAST, основанным на сходстве. Он объединил прогнозы нейронной сети с методами, основанными на сходстве последовательностей, для сбора информации о взаимодействии.

В ProtConv (Сара и др., 2021) алгоритм CNN был представлен и обучен для задачи прогнозирования функции белка. Он преобразовал векторное представление последовательности белка или пептида в двумерное изображение с одним каналом, которое подается в CNN.

Несмотря на то, что предложенные выше модели обеспечивают относительно хорошие результаты прогнозирования при решении задачи прогнозирования функции белка, все же существуют некоторые проблемы. С одной стороны, сетевая структура не может эффективно фиксировать долгосрочную зависимость между одной и той же последовательностью белка и не может полностью извлекать информацию о последовательности аминокислот. Долгосрочная зависимость относится к отношениям зависимости на большом расстоянии между каждой аминокислотой в последовательности белка. Установив эту взаимосвязь, можно лучше усвоить общую информацию о последовательности. С другой стороны, трудно эффективно отличить достоверную информацию от недействительной информации о последовательности белка. Трудно уловить аминокислотную последовательность, которая оказывает большее влияние на функцию белка. Достоверная информация относится к информации о последовательности белка, которая оказывает большое влияние на функцию белка. Соответственно, недействительная информация относится к информации о последовательности белка, которая оказывает меньшее влияние на функцию белка.

Метод комбинации механизма самовнимания и ViLSTM может быть

использован для решения проблемы прогнозирования функции белка. Во-первых, для информации о последовательности аминокислот, которая не может быть полностью извлечена, двунаправленная сеть долговременной кратковременной памяти (BiLSTM) (Грейвс и др., 2005) используется для извлечения глобальной и локальной информации о свойствах белков. В то же время связь последовательности между информацией об объектах может быть эффективно сохранена, так что модель может получить лучший эффект прогнозирования. Во-вторых, чтобы лучше использовать взаимосвязь последовательности между информацией о функциях и отражать важность различных позиций последовательности, в этом эксперименте используется механизм самовнимания (Ченг и др., 2016), чтобы заставить модель уделять больше внимания важным функциям в последовательности, тем самым повышая надежность и универсальность модели предсказания функции белка.

Таким образом, в качестве объекта исследования в данной статье используются различные белки, последовательности которых находятся в открытом доступе. Экспериментальные результаты показывают, что, по сравнению с классическим алгоритмом нейронной сети (алгоритм CNN и алгоритм LSTM) и комбинированной версией алгоритма CNN и алгоритма BiLSTM (алгоритм CNN-BiLSTM) алгоритм самовнимания-BiLSTM, использованный в этой статье, позволяет достичь лучших результатов прогнозирования при прогнозировании функции белка.

#### **Материалы и методы**

Чтобы использовать нейронную сеть для прогнозирования функции белка на основе аминокислотных последовательностей, первая задача состоит в том, чтобы найти лучший способ представления входных данных, чтобы последовательности белков могли быть распознаны программой. Популярные на сегодняшний день методы кодирования включают метод «горячего» кодирования, «изученные вложения» и «вложения BLOSUM62». По сравнению с методом «изученные вложения» метод «горячего» кодирования позволяет не только уменьшить количество параметров модели, но и избежать проблемы переобучения. Метод BLOSUM62 является одним из популярных методов кодирования. Он представляет каждую аминокислоту соответствующей строкой в матрице BLOSUM62. Вместо того, чтобы рассматривать все аминокислоты независимо друг от друга, матрица BLOSUM62 хранит эволюционную информацию о том, какие пары аминокислот легко взаимозаменяемы в ходе эволюции. Исследование показало, что метод «горячего» кодирования обеспечивает меньшую ошибку модели по сравнению с встраиванием BLOSUM62. Поэтому в данном исследовании последовательности кодов аминокислот кодируется методом «горячего» кодирования. Этот метод сопоставляет каждой букве аминокислоты конкретное действительное число от 1 до 20.

Затем каждому терму n-граммы ставится в соответствие вектор, состоящий из всех нулей, кроме единицы в позиции, зарезервированной для этого терма.

Например, действительное число, соответствующее букве D, равно 3, а это значит, что третьей позиции ее вектора присваивается единица, а остальные позиции равны нулю.

Стоит отметить, что длины белковых последовательностей в основном неодинаковы и сильно варьируют. Чтобы унифицировать формат входных данных и сократить время расчета модели, в этом эксперименте длина каждой белковой последовательности унифицирована до 1002. Несмотря на ограничение, состоящее в том, что длина последовательности составляет 1002 и она не содержит неоднозначных кодов аминокислот, около 90% белковых последовательностей в UniProt удовлетворяют этим условиям (Кулманов и др., 2019).

Другими словами, белковые последовательности длиной более 1002 отфильтровываются. Если последовательности белков с начальной длиной меньше 1002, они дополняются нулями слева до тех пор, пока длина последовательности не станет равной 1002. Наконец, все последовательности белков с неоднозначными кодами аминокислот (B, J, O, U, X, Z) удаляются.

Пример последовательности:

*PESRIRLSTRRDAHGMPPIRIESRLGPDAFARLRFMARTCRILAAAGCAAP  
FEFSSADAFSSTHVFGTCRMGHDPMRNVVDGWGRSHRWPNLNFVADAS  
LFPSSGGGESPGLTIQALALRT*

Для каждой последовательности один буквенный код заменяется на число. После кодирования вышеупомянутая последовательность будет иметь такой вид:

*[13, 7, 13, 4, 16, 15, 8, 15, 10, 16, 17, 15, 15, 3, 1, 7, 6, 11, 13, 8, 13, 15, 8, 4,  
16, 15, 10, 6, 13, 3, 1, 5, 1, 15, 10, 15, 5, 11, 1, 15, 17, 2, 15, 1, 8, 10, 1, 1, 1, 6, 2,  
1, 1, 13, 5, 4, 4, 5, 16, 16, 1, 3, 1, 5, 16, 16, 17, 7, 18, 5, 6, 17, 2, 15, 11, 6, 7, 3, 13,  
11, 15, 12, 18, 18, 3, 6, 19, 6, 15, 16, 7, 15, 19, 13, 12, 10, 5, 18, 1, 3, 1, 16, 10, 5,  
13, 16, 16, 6, 6, 6, 4, 16, 13, 6, 10, 17, 8, 14, 1, 10, 1, 10, 15, 17]*

В используемом методе, во-первых, для информации о последовательности аминокислот, которая не может быть полностью извлечена, используется двунаправленная сеть долговременной кратковременной памяти (BiLSTM) для извлечения глобальной и локальной информации о свойствах белков (Грейвс и др., 2005). В то же время связь последовательности между информацией об объектах может быть эффективно сохранена, так что модель может получить лучший эффект прогнозирования.

Во-вторых, чтобы лучше использовать взаимосвязь последовательности между информацией об особенностях и отражать важность различных положений последовательности, в этом эксперименте используется механизм self-attention, чтобы заставить модель уделять больше внимания важным функциям в последовательности, тем самым повышая эффективность, надежность и универсальность алгоритма предсказания функции белка.

BiLSTM — это один из типов рекуррентных нейронных сетей, который обрабатывает данные последовательности как в прямом, так и в обратном

направлении с двумя отдельными скрытыми слоями. BiLSTM основан на вентилях ввода, забывания и вывода. Для расчета прогнозных значений используются следующие формулы (1) (Абдулджаббар и др., 2021):

$$\begin{aligned} \text{input gate}(i_t) &= \sigma_g(W_i X_t + R_i h_{t-1} + b_i), \\ \text{forget gate}(f_t) &= \sigma_g(W_f X_t + R_f h_{t-1} + b_f), \\ \text{cell candidate}(c_t) &= \sigma_g(W_c X_t + R_c h_{t-1} + b_c), \\ \text{output gate}(o_t) &= \sigma_g(W_o X_t + R_o h_{t-1} + b_o), \end{aligned} \quad (1)$$

где  $\sigma_g$  — функция активации вентиля, а  $W_i$ ,  $W_f$ ,  $W_c$  и  $W_o$  входные весовые матрицы, тогда как  $R_i$ ,  $R_f$ ,  $R_c$  и  $R_o$  — весовые матрицы, соединяющие предыдущее выходное состояние ячейки с тремя вентилями и входное состояние ячейки.  $X_t$  — вход, и  $h_{t-1}$  выход в предыдущий момент времени ( $t-1$ ).  $b_i$ ,  $b_f$ ,  $b_c$  и  $b_o$  — векторы смещения.

На каждой временной итерации  $t$  состояние выхода ячейки  $C_t$  и выход слоя  $h_t$  можно рассчитать следующим образом (Куртукова и др., 2019):

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t, \quad (2)$$

$$h_t = o_t * \tanh(C_t) \quad (3)$$

Архитектура двунаправленной модели LSTM представлена на Рисунке 1.

Использование двунаправленного LSTM запускает ввод двумя способами, позволяя сохранять контекстную информацию из прошлого и будущего в любой момент времени. Алгоритм BiLSTM может собирать важную информацию об аминокислотных последовательностях в двух направлениях, полностью учитывать информацию о контекстуальной корреляции текущих аминокислотных последовательностей и может более глубоко изучать особенности белковых последовательностей.

Однако из-за длинной аминокислотной последовательности модель BiLSTM не может уловить самую прямую связь между вектором признаков и меткой результата. Добавление в модель механизма self-attention (Васвани и др., 2017) может решить эту проблему. Он может взвешивать входные функции и измерять важность каждой функции для экспериментального объекта. Механизм self-attention широко используется в области классификации текстов и изображений машинного перевода и биоинформатики. В этой экспериментальной модели взаимосвязь вычислений в механизме self-attention показана на Рисунке 2.

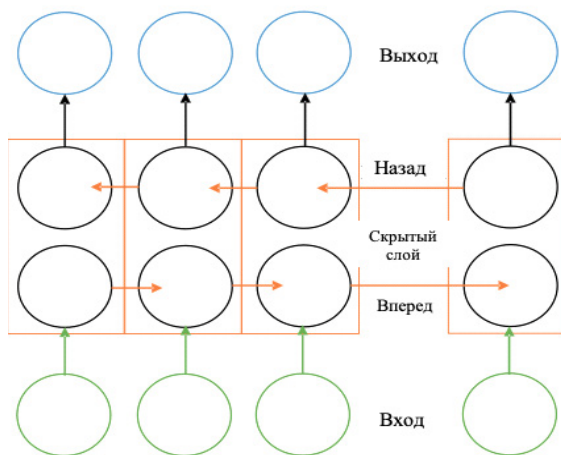


Рисунок 1 – BiLSTM-архитектура

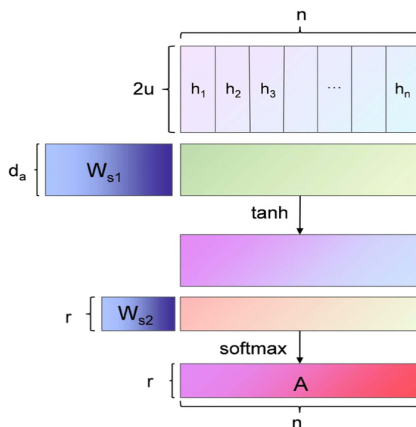


Рисунок 2 - Взаимосвязь вычислений в механизме self-attention

Для признаков, находящихся далеко друг от друга и взаимозависимых, требуется определенное количество времени и шагов, чтобы накопить достаточно информации, чтобы связать их. Чем дальше они друг от друга, тем меньше вероятность того, что сеть BiLSTM захватит эффективную информацию. Это означает, что, когда аминокислота может быть связана с окружающими ее аминокислотами или более отдаленными аминокислотами, использование алгоритма BiLSTM учитывает только информацию до и после последовательности белка в определенном диапазоне и не может решить проблему корреляции между прерывистыми аминокислотами. Стоит отметить, что одна аминокислота или несколько аминокислот могут иметь большое влияние на функцию белка. В процессе расчета механизм self-attention может напрямую связать корреляцию между любыми двумя функциями в последовательности за один шаг расчета, что значительно

сокращает расстояние между зависимыми функциями на большом расстоянии. Следовательно, комбинация алгоритма BiLSTM и self-attention уделяет больше внимания аминокислотам, которые могут иметь большое влияние на функцию белка, так что аминокислотная последовательность вносит большой вклад в точное предсказание функции белка.

Техническая реализация были исполнена на языке Python с использованием библиотек Num.py, TensorFlow и Keras и запущена на облачной платформе Colab (<https://colab.research.google.com/>).

При реализации модели были использованы следующие инструменты:

- В качестве алгоритма оптимизации был выбран алгоритм Adam;
- показатель точности Accuracy как целевая функция;
- Бинарно-кроссэнтропийная функция, возвращающая ошибку классификации как функцию логистических потерь Loss:

$$Loss = -\frac{1}{N} \sum_{i=1}^N y_i * \log(\hat{y}_i) + (1 - \hat{y}_i) * \log(1 - \hat{y}_i), \quad (4)$$

где  $y_i$  – истинная метка класса;  $\hat{y}_i$  ответ классификатора (вычисляемая метка класса) на  $i$ -й объект;  $N$  – количество классов.

Эффективность алгоритмов прогнозирования функции белка оценивается по четырем популярным показателям (Cheng., 2016): чувствительности (sensitivity - SE), специфичности (specificity - SP), точности (accuracy - ACC) и коэффициенту корреляции Мэттьюса (Matthews correlation coefficient – MCC). Все четыре показателя широко используются для оценки эффективности предикторов функции белков и основываются на четырех составляющих  $TP$ ,  $TN$ ,  $FP$  и  $FN$ , представляющих собой истинно положительные, истинно отрицательные, ложноположительные и ложноотрицательные результаты соответственно.

В частности, SE определяется процентом истинно положительных образцов, правильно идентифицированных как «положительные»:

$$SE = \frac{TP}{TP+FN} \quad (5)$$

SP указывает долю истинно отрицательных образцов, которые были правильно предсказаны как «отрицательные»:

$$SP = \frac{TN}{TN+FP} \quad (6)$$

ACC относится к количеству истинных образцов (положительных плюс отрицательных), деленному на количество всех изученных образцов:

$$ACC = \frac{TP+TN}{TP+TN+FP+FN} \quad (7)$$

MCC является важным показателем, отражающим стабильность предиктора функции белка, который описывает корреляцию между прогностическим значением и фактическим значением. Он считается одним из наиболее полных параметров в любой категории предикторов из-за полного учета всех четырех результатов. В частности, MCC (F1) можно рассчитать по следующей формуле:

$$MCC = \frac{(TP*TN-FP*FN)}{\sqrt{(TP+FN)*(TP+FP)*(TN+FP)*(TN+FN)}} \quad (8)$$

### Результаты и обсуждение

Для обучения модели были запущены 50 эпох обучения. Для предотвращения переобучения модели были выставлены следующие параметры: `earlystopping = EarlyStopping(monitor='val_loss', patience=3, verbose=1)`. В Таблице 1 представлены результаты обучения сети, полученные в течение первых 39 эпох обучения.

Таблица 1. Результаты обучения сетевой модели

Epoch	Loss	Accuracy	Val_loss	Val_accuracy
1	5.4675	0.1226	3.4304	0.3755
2	2.8003	0.4818	1.9909	0.6735
3	1.8967	0.6634	1.3769	0.7867
4	1.4488	0.7507	1.0554	0.8479
5	1.2040	0.7947	0.8728	0.8772
6	1.0520	0.8230	0.7566	0.8987
7	0.9574	0.8404	0.7134	0.9029
...				
33	0.5505	0.9144	0.3783	0.9610
34	0.5439	0.9147	0.3818	0.9609
35	0.5398	0.9163	0.3673	0.9635
36	0.5344	0.9167	0.3582	0.9648
37	0.5306	0.9174	0.3720	0.9616
38	0.5259	0.9186	0.3598	0.9638
39	0.5209	0.9189	0.3611	0.9639

Оценка производительности модели показала, что наилучшая точность BiLSTM составила 0.9699 для обучающего набора, 0.9638 для набора данных и 0.9632 для тестового набора.

На Рисунке 3 показаны значения всех четырёх показателей для разработанной модели на примере тестового набора данных белков *Indica*, также приведены результаты запуска нескольких классических моделей на тех же наборах данных.

Экспериментальные результаты алгоритма SA-BiLSTM сравниваются с алгоритмами CNN, LSTM и CNN-BLSTM.



Набор данных	Подтип онтологии	Алгоритм	Accuracy (%)	Precision (%)	Recall (%)	F1 (%)
Indica	BP	CNN	73.856	50.237	88.228	64.020
		LSTM	89.140	81.669	77.071	79.294
		CNN-BiLSTM	87.973	78.994	74.139	76.461
		SA-BiLSTM	90.551	83.113	79.443	81.186
	MF	CNN	84.454	73.636	51.654	60.673
		LSTM	86.190	66.831	61.324	63.911
		CNN-BiLSTM	84.796	63.966	63.482	63.660
		SA-BiLSTM	87.274	75.071	63.248	68.608
	CC	CNN	81.506	78.081	77.655	77.800
		LSTM	82.186	80.101	77.618	78.825
		CNN-BiLSTM	81.691	78.400	77.655	78.006
		SA-BiLSTM	85.299	81.409	83.919	82.434

Рисунок 3 - Результаты предсказания функций белков на тестовых данных

Набор данных	Подтип онтологии	Белок	Функция	Предсказанная функция
Indica	BP	Q01N44	GO:0016042	GO:0016042
			GO:0006629	GO:0006629
				GO:0006807
	MF	E0ZS48	GO:0009039	GO:0009039
			GO:0016787	GO:0016787
			GO:0046872	GO:0046872
			GO:0016810	
	CC	Q01N44	GO:0005783	GO:0005783
			GO:0005886	GO:0005886
			GO:0016020	GO:0016020
			GO:0005789	GO:0005789
				GO:0009536

Рисунок 4 - примеры предсказания функций белков Indica

Белок E0ZS48 (UREA\_ORYSI), как было обнаружено вручную, имеет функции белка GO: 0009039, GO: 0016787, GO: 0016810 и GO: 0046872. Среди них функция GO:0009039 представляет активность уреазы. Функция GO:0016787 указывает на гидролазную активность, которая может катализировать гидролиз различных связей. Функция GO:0016810 означает активность гидролазы, которая катализирует гидролиз любой углерод-азотной связи C-N, кроме пептидных связей. GO:0046872 означает, что белок выполняет функцию связывания с ионом металла. К сожалению, экспериментальные результаты показывают, что функция GO:0016810 не может быть успешно предсказана, что может быть связано с тем, что функция GO:0016910 и функция GO:0016787 имеют схожие функции. Обе функции катализируют гидролиз определенных связей. Поэтому для экспериментов трудно сделать абсолютно точное предсказание функции по этой проблеме.

### Заключение

Таким образом, предложенная модель не обязательно точно предсказывает полную функцию белка с очень похожими функциями. Но функции большинства белков можно точно предсказать. Согласно Рисунку 3 и Рисунку 4, эффект предсказания белка с помощью алгоритма SA-BiLSTM в целом

хороший. Более того, в дополнение к полному предсказанию аннотации функции белка эксперимент также может предсказать аннотации функции GO, которые не показаны в базе данных Swiss-Prot. Это дает новое направление для последующих экспериментальных исследований.

#### ЛИТЕРАТУРА

Абдулжаббар Р.Л., Диа Х., Цай П. Однонаправленные и двунаправленные LSTM-модели для краткосрочного прогнозирования трафика // *Journal of Advanced Transportation*. – 2021. – Том 4. – С. 1-16.

Грейвс А., Шмидхубер Дж. Показательная классификация фонем с использованием двунаправленного LSTM и других архитектур нейронных сетей // *Нейронные сети*. – 2005. – С. 602-610.

Гиллис Дж., Павлидис П. Характеристика уровня техники в вычислительном присвоении функции гена: уроки первой критической оценки функциональной аннотации (CAFA) // *BMC Bioinformatics*. – 2013.

Кулманов М., Хендорф Р. DeepGO: прогнозирование функций белка по последовательности и взаимодействиям с использованием классификатора, основанного на глубокой онтологии // *Биоинформатика*. – 2018. – Том 34. – С. 660-668.

Кулманов М., Хендорф Р. DeepGOPlus: Улучшенное предсказание функции белка по последовательности // *Биоинформатика*. – 2019.

Куртукова А.В., Романов А.С. Моделирование архитектуры нейронной сети для идентификации автора исходного кода // *Известия Томского государственного университета систем управления и радиоэлектроники*. – 2019. – Том 22. – С. 37-42.

Сара С., Хасан М., Ахмад А., Шатабда С. Сверточные нейронные сети с графическим представлением аминокислотных последовательностей для прогнозирования функций белков // *Вычислительная биология и химия*. – 2021.

Ченг Дж., Донг Л., Лапата М. Долговременная кратковременная память-сети для машинного чтения // *Конференция по эмпирическим методам обработки естественного языка*. – 2016. – С. 551-561.

Цай С.З., Хан Л.Ю., Ци З.Л., Чен Х., Чен Ю.З. SVM-Prot: веб-программное обеспечение для обработки опорных векторов для функциональной классификации белка по его первичной последовательности. Исследование нуклеиновых кислот. – 2003. – С. 3692-3697.

Чен Х.В., Лю М. Прогнозирование белок-белковых взаимодействий с использованием структуры леса случайных решений // *Биоинформатика*. – 2005. – С. 4394-4400.

Саидния С., Манайи А., Абдоллахи М. От экспериментов *in vitro* к *in vivo* и клиническим исследованиям; плюсы и минусы // *Современные технологии разработки лекарственных средств*. – 2015. – С. 218-224.

Нам Дж.У., Шин К.Р., Хан Дж., Ли Ю., Ким В.Н., Чжан Б.Т. Предсказание микроРНК человека с помощью вероятностной модели совместного обучения последовательности и структуры // *Исследования нуклеиновых кислот*. – 2005. – С. 3570-3581.

Васвани А., Шахир Н., Пармар Н. и др. Внимание - это все, что вам нужно // *NIPS'17: Материалы 31-й международной конференции по нейронным системам обработки информации*. – 2017. – С. 6000-6010.

Цзян Ю., Орон Т., Кларк У. и др. Расширенная оценка методов прогнозирования функций белков показывает повышение точности // *Биология генома*. – 2017.

Ю. Г., Рангвала Х., Доменикони С., Чжан З. Чжан З. Предсказание функции белка с использованием нескольких ядер // *Транзакции IEEE/ACM по вычислительной биологии и биоинформатике*. – 2015. – С. 219-233.

Юсеф М., Юнг С., Шоу Л.С. и др. Обучение на положительных примерах, когда класс отрицательных не определен - идентификация генов микроРНК // *Алгоритмы молекулярной биологии*. – 2008.

## REFERENCES

- Abduljabbar R.L., Dia H., Tsai P. Unidirectional and Bidirectional LSTM Models for Short-Term Traffic Prediction // *Journal of Advanced Transportation*. – 2021. – Vol. 4. – P. 1-16.
- Graves A., Schmidhuber J. Framewise phoneme classification with bidirectional LSTM and other neural network architectures // *Neural Networks*. – 2005. – P. 602-610.
- Gillis J., Pavlidis P. Characterizing the state of the art in the computational assignment of gene function: lessons from the first critical assessment of functional annotation (CAFA) // *BMC Bioinformatics*. – 2013.
- Kulmanov M., Hoehndorf R. DeepGO: predicting protein functions from sequence and interactions using a deep ontology-aware classifier // *Bioinformatics*. – 2018. – Vol. 34. – P. 660-668.
- Kulmanov M., Hoehndorf R. DeepGOPlus: Improved protein function prediction from sequence // *Bioinformatics*. – 2019.
- Kurtukova A.V., Romanov A.S. Modeling the neural network architecture to identify the author of the source code // *Proceedings of Tomsk State University of Control Systems and Radioelectronics*. – 2019. – Vol. 22. – P. 37-42.
- Sara S., Hasan M., Ahmad A., Shatabda S. Convolutional neural networks with image representation of amino acid sequences for protein function prediction // *Computational Biology and Chemistry*. – 2021.
- Cheng J., Dong L., Lapata M. Long Short-Term Memory-Networks for Machine Reading // *Conference on Empirical Methods in Natural Language Processing*. – 2016. – P. 551-561.
- Cai C.Z., Han L.Y., Ji Z.L., Chen X., Chen Y.Z. SVM-Prot: Web-based support vector machine software for functional classification of a protein from its primary sequence. *Nucleic Acids Research*. – 2003. – P. 3692-3697.
- Chen X.W., Liu M. Prediction of protein-protein interactions using random decision forest framework // *Bioinformatics*. – 2005. – P. 4394-4400.
- Saeidnia S., Manayi A., Abdollahi M. From in vitro Experiments to in vivo and Clinical Studies; Pros and Cons // *Current Drug Discovery Technologies*. – 2015. – P. 218-224.
- Nam J.W., Shin K.R., Han J., Lee Y., Kim V.N., Zhang B.T. Human microRNA prediction through a probabilistic co-learning model of sequence and structure // *Nucleic Acids Research*. – 2005. – P. 3570-3581.
- Vaswani A., Shazeer N., Parmar N., et al. Attention is all you need // *NIPS'17: Proceedings of the 31st International Conference on Neural Information Processing Systems*. – 2017. – P. 6000-6010.
- Jiang Y., Oron T., Clark W. et al. An expanded evaluation of protein function prediction methods shows an improvement in accuracy // *Genome Biology*. – 2017.
- Yu G., Rangwala H., Domeniconi C., Zhang G., Zhang Z. Predicting Protein Function Using Multiple Kernels // *IEEE/ACM Transactions on Computational Biology and Bioinformatics*. – 2015. – P. 219-233.
- Yousef M., Jung S., Showe L.C. et al. Learning from positive examples when the negative class is undetermined- microRNA gene identification // *Algorithms for Molecular Biology*. – 2008.

NEWS OF THE NATIONAL ACADEMY OF SCIENCES OF THE REPUBLIC OF KAZAKHSTAN  
PHYSICO-MATHEMATICAL SERIES

ISSN 1991-346X

Volume 3. Number 347 (2023). 76–87

<https://doi.org/10.32014/2023.2518-1726.205>

UDC 004-93

© **L. Zholshiyeva**<sup>1\*</sup>, **T. Zhukabayeva**<sup>1</sup>, **Sh. Turaev**<sup>2</sup>, **M. Berdieva**<sup>3</sup>, 2023

<sup>1</sup>Astana International University, Eurasian National University named after  
L.N. Gumilyov, Astana, Kazakhstan;

<sup>2</sup>College of Information Technology, United Arab Emirates University,  
Al Ain United Arab Emirates;

<sup>3</sup>South Kazakhstan Medical Academy, Shymkent, Kazakhstan;  
E-mail: [lazzat.zhol.81@gmail.com](mailto:lazzat.zhol.81@gmail.com)

### **KAZAKH SIGN LANGUAGE RECOGNITION BASED ON CNN**

**Zholshiyeva Lazzat Zulpuharkyzy** — PhD student. Astana International University. Astana, Kazakhstan

E-mail: [lazzat.zhol.81@gmail.com](mailto:lazzat.zhol.81@gmail.com). ORCID: 0000-0002-2526-8471;

**Zhukabayeva Tamara Kokenovna** — PhD, assoc. Professor. Astana International University, Eurasian National University named after L.N. Gumilyov. Astana, Kazakhstan

E-mail: [tamara\\_kokenovna@mail.ru](mailto:tamara_kokenovna@mail.ru). ORCID: 0000-0001-6345-5211;

**Turaev Sherzod** — PhD, assoc. Professor. College of Information Technology, United Arab Emirates University Al Ain, United Arab Emirates

E-mail: [sherzod@uaeu.ac.ae](mailto:sherzod@uaeu.ac.ae). ORCID: 0000-0001-6661-8469;

**Berdieva Meruert Aimambetovna** — PhD. South Kazakhstan Medical Academy, Shymkent, Kazakhstan

E-mail: [meruert\\_berdieva@mail.ru](mailto:meruert_berdieva@mail.ru). ORCID: 0009-0008-5259-7618.

**Abstract.** Technology offers to deaf and dumb people more opportunity for interaction with other people because the most effective way for people to interact is speech and hand gestures. Therefore, the interaction between man and technology is important, and it has been developing rapidly in recent times. This requires a method or application that allows a person to communicate with technology, that is, recognize sign language and hand gestures. There are many methods for Sign Language Recognition. Deep learning techniques using artificial neural networks are often recommended among them. In this study, the Convolutional Neural Network (CNN) was studied. The study presents a machine learning system independent of humans for translating Kazakh gestures into text and recognizing them. Dataset consisting of 42 Kazakh Dactyls that divided into 80% training and 20 % test data, and then trained using the CNN classifier. VGG16 and ResNet50 architectures have been used to improve gesture recognition performance. According to the study,

the recognition rates were 98.867 % (test) and 91.323 % (estimate) with VGG16; 78.612 % (test) and 62.69 % (estimate) with ResNet50, respectively.

**Keywords:** Kazakh Sign Language, Sign Language Recognition (SLR), Hand Gesture Recognition (HGR), Convolutional Neural Network (CNN)

© Л.З. Жолшиева<sup>1\*</sup>, Т.К. Жукабаева<sup>1</sup>, Ш. Тураев<sup>2</sup>, М.А. Бердиева<sup>3</sup>, 2023

<sup>1</sup>Астана халықаралық университеті, Л.Н. Гумилев атындағы Еуразия ұлттық университеті, Астана, Қазақстан;

<sup>2</sup>Ақпараттық технологиялар колледжі, Біріккен Араб Әмірліктері университеті, Әл-Айн, Біріккен Араб Әмірліктері;

<sup>3</sup>Оңтүстік Қазақстан медициналық академиясы, Шымкент, Қазақстан.  
E-mail: lazzat.zhol.81@gmail.com

### CNN НЕГІЗІНДЕ ҚАЗАҚ ҰМ ТІЛІН ТАҢУ

**Жолшиева Лаззат Зулпухарқызы** — PhD студент. Астана халықаралық университеті. Астана, Қазақстан

E-mail: lazzat.zhol.81@gmail.com. ORCID: 0000-0002-2526-8471;

**Жукабаева Тамара Кокеновна** — PhD, қауымдастырылған профессор. Астана халықаралық университеті. Л.Н. Гумилев атындағы Еуразия ұлттық университеті. Астана, Қазақстан

E-mail: tamara\_kokenovna@mail.ru. ORCID: 0000-0001-6345-5211;

**Тураев Шерзод** — PhD, қауымдастырылған профессор. Ақпараттық технологиялар колледжі, Біріккен Араб Әмірліктері университеті, Әл-Айн, Біріккен Араб Әмірліктері

E-mail: sherzod@uaeu.ac.ae. ORCID: 0000-0001-6661-8469;

**Бердиева Меруерт Аймамбетовна** — PhD. Оңтүстік Қазақстан медициналық академиясы, Шымкент, Қазақстан

E-mail: meruert\_berdieva@mail.ru. ORCID: 0009-0008-5259-7618;

**Аннотация.** Адамдар арасындағы өзара әрекеттесудің ең тиімді жолы сөйлесу және қол қимылдары болғандықтан, сөйлеуге, естуге мүмкіндігі жоқ адамдар үшін технологияның үлесі көп. Демек, бұл жағдайда адам мен технологияның арасындағы байланыс маңызды және ол соңғы уақытта қарқынды дамуда. Осы мақсатта адам мен технологияның қарым-қатынасын жүзеге асыруға мүмкіндік беретін, яғни ұм тілі мен қол қимылдарын тани алатын әдіс немесе қолданбаның қажеттілігі орынды. Қол қимылдарын тануды орындайтын әдістер көп. Солардың ішінен жасанды нейрондық желілерді қолдана отырып, терең оқыту әдістері жиі ұсынылады. Бұл жұмыста конволюциялық нейрондық желі (CNN) зерттеуге алынды. Зерттеуде адамнан тәуелсіз, машиналық оқытуға негізделген қазақ ұм тілін мәтінге аудару және оларды тану жүйесі қарастырылған. Дайын дербес 42 қазақ дактилінен құралған деректер алдымен 80 % оқыту мен 20 % тестілеу деректеріне бөлініп, содан соң CNN классификаторы арқылы оқытылды. Ұмдарды тану өнімділігін жақсарту үшін VGG16, ResNet50 архитектуралары қолданылды. Зерттеу нәтижесі бойынша тану көрсеткіштері сәйкесінше VGG16 бойынша 98,867 % (сынақ) және 91,323 % (бағалау), ResNet50 бойынша 78,612% (сынақ) және 62,69 % (бағалау) дәлдіктері алынды.

**Түйін сөздер:** қазақ ым тілі, ым тілін тану, қол қимылын тану, конволюциялық нейрондық желі

© **Л.З. Жолшиева<sup>1\*</sup>, Т.К. Жукабаева<sup>1</sup>, Ш. Тураев<sup>2</sup>, М.А. Бердиева<sup>3</sup>, 2023**

<sup>1</sup>Международный университет Астана, Евразийский национальный университет имени Л.Н. Гумилева, Астана, Казахстан;

<sup>2</sup>Колледж информационных технологии, Университет Объединенных Арабских Эмиратов, Эль-Айн, Объединенные Арабские Эмираты;

<sup>3</sup>Южно-Казахстанская медицинская академия, Шымкент, Казахстан.  
E-mail: lazzat.zhol.81@gmail.com

## **РАСПОЗНАВАНИЕ КАЗАХСКОГО ЖЕСТОВОГО ЯЗЫКА НА ОСНОВЕ CNN**

**Жолшиева Лаззат Зулпухаркызы** — PhD студент. Международный университет Астана. Астана, Казахстан

E-mail: lazzat.zhol.81@gmail.com. ORCID: 0000-0002-2526-8471;

**Жукабаева Тамара Кокеновна** — PhD, ассоциированный профессор. Международный университет Астана. Евразийский национальный университет им. Л.Н. Гумилева. Астана, Казахстан

E-mail: tamara\_kokenovna@mail.ru. ORCID: 0000-0001-6345-5211;

**Тураев Шерзод** — PhD, ассоциированный профессор. Колледж информационных технологий. Университет Объединенных Арабских Эмиратов, Аль-Айн, Объединенные Арабские Эмираты

E-mail: sherzod@uaeu.ac.ae. ORCID: 0000-0001-6661-8469;

**Бердиева Меруерт Аймамбетовна** — PhD. Южно-Казахстанская медицинская академия, Шымкент, Казахстан

E-mail: meruert\_berdieva@mail.ru. ORCID: 0009-0008-5259-7618.

**Аннотация.** Поскольку наиболее эффективным способом человеческого взаимодействия друг с другом является речь и жесты рук, технологии могут многое предложить людям, которые не могут говорить или слышать. Поэтому в данном случае важна связь человека и техники, которая в последнее время быстро развивается. Для этой цели необходим метод или приложение, которое обеспечивает общение человека с технологией. Есть много методов, выполняющих распознавания жестов рук. Среди них часто рекомендуются методы глубокого обучения с использованием искусственных нейронных сетей. В этой работе изучалась сверточная нейронная сеть (CNN). В исследовании представлена независимая от человека система машинного обучения перевода казахских жестов в текст и их распознавание. Данные, состоящие из 42 независимых казахских дактилей, сначала были разделены на 80% обучающих и 20% тестовых данных, а затем обучены с использованием классификатора CNN. Архитектуры VGG16 и ResNet50 использовались для улучшения производительности распознавания жестов. По результатам исследования показатели распознавания оказались точными на 98,867 % (тест) и 91,323 % (оценка) с VGG16; 78,612 % (тест) и 62,69 % (оценка) с ResNet50 соответственно.

**Ключевые слова:** казахский жестовый язык, распознавание жестового языка, распознавание жестов рук, сверточная нейронная сеть

### **Кіріспе**

Терең оқыту мен компьютерлік көру саласындағы бүгінгі жетістіктер қозғалыс пен қол, дене және бет қимылдарын танудың дамып келе жатқандығын көрсетеді. Ым тілдерінің үш түрі бар: әр дактильді саусақпен көрсету; қол және дене қозғалысы арқылы жасалған сөздер және тіркестер сөздігі; мимика мен ерін қимылдары. Бұл жұмыста аталған бірінші ым тілі қарастырылады. Қол қимылдарын танудың барлық әдістерін көру қабілетіне негізделген және қолғапты сенсорлар шығаратын өлшемдерге негізделген деп жіктеуге болады. Көру әдісіне негізделген әдіс адам мен компьютердің өзара әрекеттесуін қамтиды. Сенсорға негізделген тануда арнайы сенсорлық құрылғы қажет болса, ал көру әдісіне тек камера қажет. Сондықтан, бұл жүйе камера мен ноутбук интерфейсінен тұрады.

Ым тілдерін тану алгоритмдерге, белгілерді тану әдістері мен оларды мәтін немесе сөйлеу түрінде көрсетіп, түсіндіруге бағытталған және де ол көру қабілетіне негізделгендіктен, ым тілдерін танудың барлық процесін компьютерлік көру арқылы автоматтандыру актуалды болып табылады.

Зерттеудің мақсаты – конволюциялық нейрондық желі арқылы қазақ ым тілін жазбаша қазақ тіліне аудару.

### **Материалдар мен әдістер**

Көру әдісіне негізделген ым тілін тану жүйесін дәстүрлі және конволюциялық нейрондық желі деп екі санатқа бөлуге болады.

Қимылдарды тану үшін қолданылатын әртүрлі әдістер мен алгоритмдер бар. Компьютерлік көру жүйесінде ымдау тілі кескін және бейне түрінде қолданылады (Dongxu Li et al., 2020). Бірқатар шектеулерге байланысты көру арқылы басқарылатын нақты уақыт режимінде қол қимылдарын танудың тиімді жүйесін құру қиын міндет болғандықтан, камералар, кескіндерді сегменттеу және қадағалау, белгілерді бөлектеу және бақылау арқылы кескіндерді алу кезінде кездесетін шектеулер анықталған (Farid et al., 2022). Табиғи тілді өңдеу арқылы ымдау тілін танитын бірнеше зерттеулер де бар (Miah et al., 2023; Subburaj et al., 2022). Көруге негізделген (Sharma et al., 2023) әдістер автоматтандырылған ым-ишара жүйелерін дамытудың ең кең тараған бағыттары болып табылады. Кодтау үшін терең оқытуға негізделген әр түрлі әдістерді зерттеген жұмыстар да бар (Ananthanarayana et al., 2021). Соңғы онжылдықта конволюциялық желі арқылы статистикалық немесе динамикалық қозғалыстарды жазу үшін ымдау тілін тану саны артты. Кейбіреулерін атап айтақ, (Krizhevsky et al., 2022) 1,2 миллион жоғары ажыратымдылықтағы кескінді 1000 түрлі сыныпқа жіктеу үшін үлкен терең конволюциялық нейрондық желіні жаттықтырған. Ал BenSignNet деп аталатын CNN негізіндегі модель арқылы үлгі дәлдігі BdSL Alphabet, KU-BdSL және Ishara-Lipi деректер жиындары үшін сәйкесінше 94,00 %, 99,60 %

және 99,60 % мақсатқа жету үшін үш қадам орындаған: сегменттеу, кеңейту және конволюциялық нейрондық желі (CNN) негізінде жіктеу (Miah et al., 2022; Barbhuiya et al., 2021) қол қимылдарын тану үшін пайдаланылатын конволюциялық нейронды желінің (CNN) артықшылықтары мен кемшіліктері атап өтіп, AlexNet және VGG16 үлгілерін зерттеп, жоғары 99,82 % тану дәлдігін алды. (Eid et al., 2023) зерттеуінде деректерді ұлғайту және теріні сегменттеу арқылы модельдің дәлдігін айтарлықтай арттыруға қол жеткізілді. Келесі мақалада (Kenshimov et al., 2021) конволюциялық нейрондық желілерді зерттеу және талдау LeNet, AlexNet, ResNet және EffectiveNet - EfficientNetB7 әдістерін салыстыру, тестілеу, нәтижелері сипатталған. Бұл алгоритмдердің архитектурасы мен жұмыс істеу принципі олардың ым тілін танудағы қолдану тиімділігін көрсетеді. Аталған зерттеу қазақ ым тілін тануға жасалған. Қазақ ым тілдерін 42 дактилін тану үшін SVM, LTSM, GRU әдістері қолданылған және олар арқылы танудың жоғары дәлдігіне қол жеткізген жұмыстар да бар (Zholshiyeva et al., 2023).

### **Нәтиже мен дискуссия**

#### *Конволюциялық нейрондық желі*

Конволюциялық нейронды желі көрнекі объектілерді өңдеу және жіктеу салаларында қолдануға болатын терең нейрондық желі (CNN) және ол компьютерлік көруде, атап айтқанда, объектілерді тануда, суреттерді класка бөлуде, суреттерді сегменттеуде кеңінен қолданылады. Конволюциялық нейронды желі архитектурасының басты компоненттері келесі қабаттардан тұрады:

1. Convolutional Layer: Бұл қабат кескіннің әртүрлі мүмкіндіктерін шығару үшін кіріс деректеріне сүзгілерді (ядроларды) қолдану арқылы конволюция операциясын орындайды. Әрбір сүзгі жиектер, бұрыштар және текстуралар сияқты нақты сипаттамаларды береді.

2. Pooling Layer: Біріктіру деңгейі деректер өлшемін азайту және есептеулерді жеңілдету үшін пайдаланылады. Ол әдетте кескіннің берілген аймағында максималды немесе орташа біріктіру әрекетін орындайды.

3. Hidden Layers: Шығарылған мүмкіндіктер негізінде жіктеуді немесе регрессияны орындайтын бір немесе бірнеше Fully Connected Layers қабаты біріктіру қабаттарынан кейін болуы мүмкін.

4. Activation Functions: ReLU (Rectified Linear Unit) секілді активтендіру функциясы желіге бейсызықтылықты қосу үшін жабық қабаттардың шығысында қолданылады.

5. Output Layer: Шығару деңгейі мәселеге байланысты қандай кластарды немесе мәндерді болжау керектігін анықтайды. Мысалы, кескінді жіктеу тапсырмасында бұл softmax белсендірілген қабат болуы мүмкін.

CNN архитектурасында қабаттардың әртүрлі комбинациялары пайдаланылады. Үлгіні жалпылауды жақсарту үшін мұнда Batch Normalization немесе Dropout сияқты қосымша элементтер болады. CNN архитектураларына LeNet, AlexNet, VGGNet, GoogLeNet, ResNet және басқалары кіреді, олардың

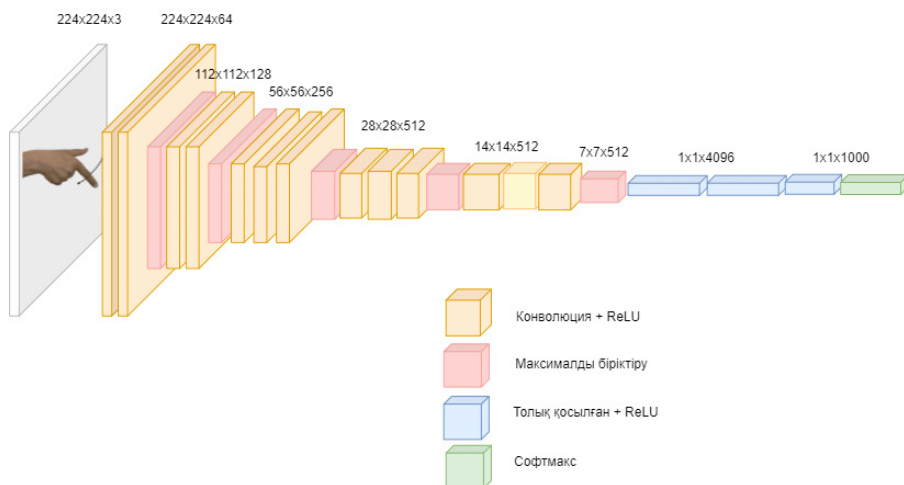


әрқайсысында компьютерлік көрудің әртүрлі тапсырмаларында өзіндік бірегей мүмкіндіктері мен қолданбалары бар.

Кескіндерді классификациялауда қолдануға болатын трансферттік оқыту моделінің бірнеше түрлері бар, бұл зерттеуде осылардың арасынан модельді оқыту үшін VGG16 және ResNet50 таңдап алынды. Қазақ ым тілін тану үшін осы архитектуралар орындалды.

*VGG16 архитектурасы*

Бұл желілік архитектура (сурет 1) қарапайым және оны кодтауда қолдану оңай болып есептеледі, өйткені өте кішкентай (3×3) конволюциялық сүзгілері бар архитектураны пайдалана отырып, тереңдігі жоғары желілерді мұқият бағалауға болады (Simonyan et al., 2015).



1- сурет. VGG16 архитектурасы

*Деректер жүктеу*

Зерттеуде қазақ ым тілінің 42 дактилінен тұратын дербес (Zholshiyeva et al., 2023) деректер жинағы пайдаланылды. Бұл деректер жинағында қазақ ым тілінің 42 класының 4200 кескіні жинақталған. Бұл деректер жинағы оқыту 80 % және тестілеу 20% деп екі жинаққа бөлінді (кесте 1).

Кесте 1 – 42 қазақ дактилінің деректер жинағының бөлінуі

Sets	Number
Training images	2821
Testing images	706
Evaluation images	922
Total	4449

Оқыту мен тестілеуді бөлгенде жапсырмалардағы стратификация аргументі қолданылды. Бұл аргумент деректердің барлық белгілер бойынша біркелкі бөлінуін қамтамасыз етеді.

*Алдын ала өңдеу*

Бұл түрлендіру белгілердің бір өлшемді массивін екі өлшемді массивке айналдырады. Қалыпқа келтіру кескіндегі жарық пен көлеңкелерден туындаған бұрмалауларды жоюға көмектеседі.

*Модель нәтижелері*

VGG16 архитектурасы ең жақсы машиналық көру моделі болып есептеледі, оның ерекшелігі - үлкен гиперпараметрлердің орнына 1 қадамнан тұратын 3x3 конволюциялық сүзгі қабаттарының болуы және әрқашан 2 қадамнан тұратын сүзгісінің бірдей толтырғышты және максималды қабатын пайдалануы. Өйткені, оның 2 FC (толық қосылған қабаттар), одан кейін шығыс үшін softmax бар. VGG16-дағы 16 саны оның салмағы бар 16 қабаты бар екенін білдіреді. Бұл желі айтарлықтай үлкен және шамамен 138 миллион (шамамен) параметрлері бар.

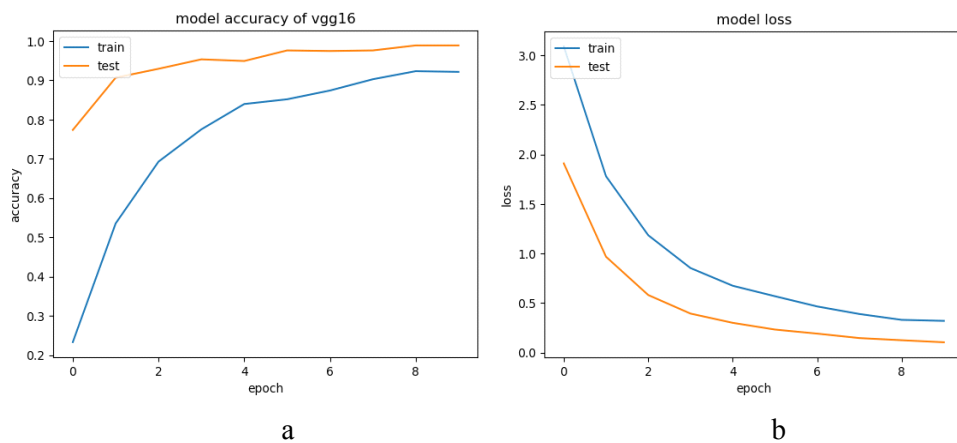
ResNet көмегімен көптеген мәселелерді шешуге болады, мысалы: ResNet-ті оңтайландыру салыстырмалы түрде оңай: «қарапайым» желілер тереңдік артқан сайын оқытудың үлкен қатесін көрсетеді. ResNet тереңдікті ұлғайту арқылы дәлдікті арттыруды салыстырмалы түрде жеңілдетеді, мұндай нәтижеге басқа желілерде қол жеткізу қиынырақ. Әрбір 3 қабатты блок 34 қабатты желіде осы 3 қабатты бөгетпен ауыстырылады, нәтижесінде 50 қабатты ResNet пайда болады. Олар өлшемдерді үлкейту үшін 2 нұсқаны пайдаланады. Бұл модельде 3,8 миллиард FLOP бар.

Дәл нәтиже беретін терең оқыту әдістемесі анықталды, жұмыстың псевдокоды:

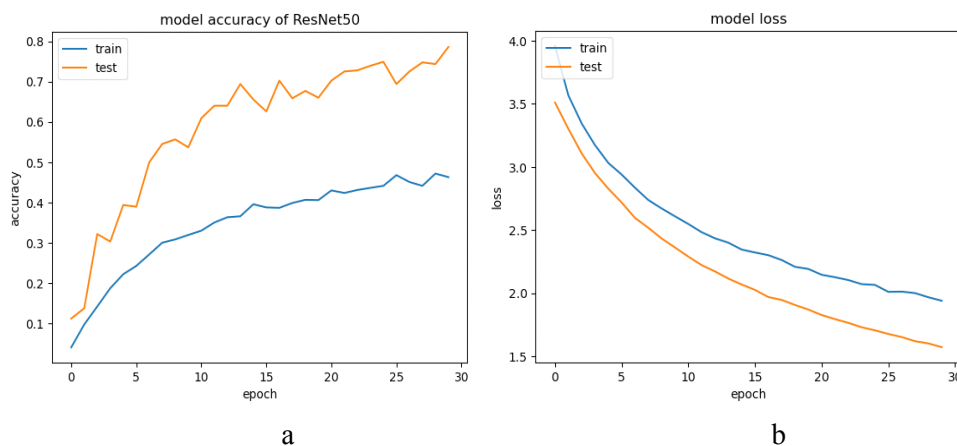
```
Begin
  For all images:
    Image resize
    Image filter
    Image segmentation
    Labeling and data splitting
  End for
  Build CNN models: VGG16, ResNet50
  Train two models
  Test and evaluate two models
End
```

### Модельдердің дәлдігі

2,3 суреттерде модельдердің дәлдігі мен шығыны көрсетілген.



2-сурет. VGG16 моделінің дәлдігі (a) мен шығыны (b)

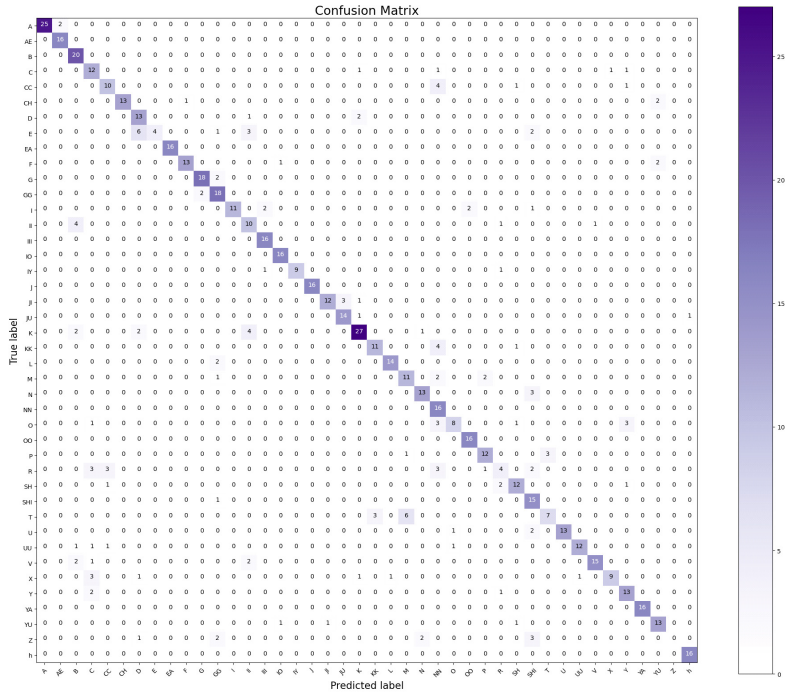


3-сурет. Resnet50 моделінің дәлдігі (a) мен шығыны (b)

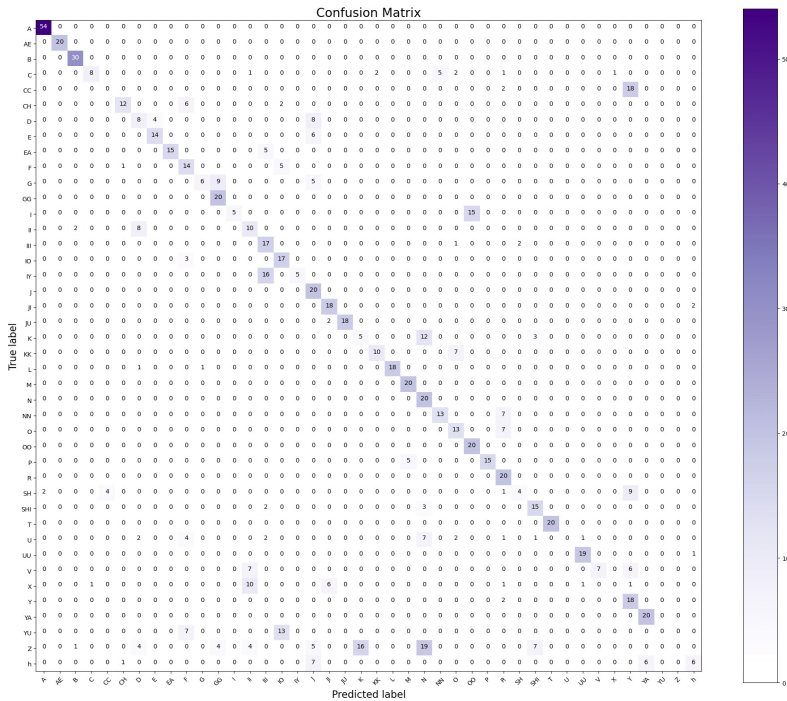
### VGG16 және Resnet50 шатасуы матрицалары

VGG16 және Resnet50 шатасуы матрицалары жоғары дәлдік беретін деректерді сынауға арналған матрица. Диагональды элементтер үлкен мәндерге ие болады және кейбір мәндер диагональдан тыс элементтерге таралады (4 - 7 сурет).



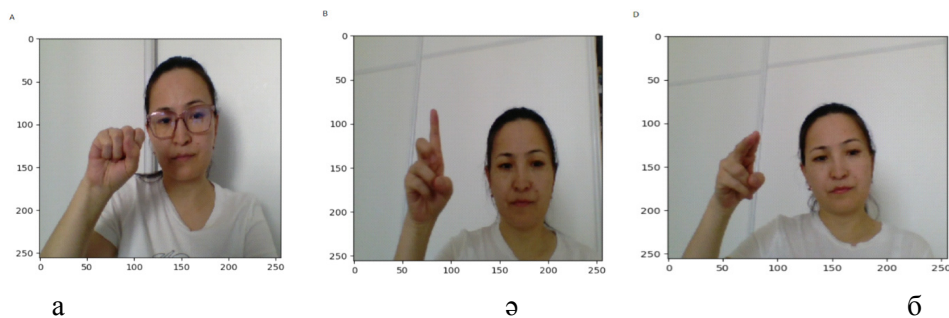


6-сурет. ResNet50 моделінің сынақ нәтижесінің шатасу матрицасы



7-сурет. ResNet50 моделінің болжау нәтижесінің шатасу матрицасы

Алынған нәтижелер қазақ ым тілін тану үшін ResNet50 архитектурасына қарағанда VGG16 архитектурасы дәлірек екенін көрсетеді. Аталған архитектуралар арқылы танылған кейбір қазақ қол-қимылдары 8-суретте көрсетілген.



8-сурет. А (а), Б (ә), Д (б) қазақ дактильдері

### **Қорытынды**

Зерттеу жұмысы қазақ ым тілін жазбаша қазақ тіліне аударудың жоғары дәлдігіне жету үшін терең оқыту әдістері пайдаланылу арқылы қойылған мақсатқа жетті. Конволюциялық нейрондық желінің (CNN) VGG16 және ResNet50 архитектуралары оқытуға қолданылды. Алынған нәтижелер VGG16 архитектурасы ResNet50 архитектурасына қарағанда жоғары көрсеткішке ие екенін көрсетті. Алдағы уақытта бұл әдіспен шектелмей, бұдан жоғары дәлдікті алу, қолданыстағы модельді жетілдіру, сондай-ақ, мүмкіндігінше көп деректерді жинауды жалғастырып, басқа алдын ала дайындалған адамдармен тәжірибе жасау және нәтижелерді салыстыру, әртүрлі әдістермен жұмыс істеу, жиналған деректер қорын түрлі әдістер арқылы салыстыру жоспарлануда.

### **ӘДЕБИЕТТЕР**

Al Farid F., Hashim N., Abdullah J., Bhuiyan M.R., Shahida Mohd Isa W.N., Uddin J., Haque M.A., Husen M.N. A Structured and Methodological Review on Vision-Based Hand Gesture Recognition System. *J. Imaging* 2022, 8, 153. <https://doi.org/10.3390/jimaging8060153> (in Eng.)

Barbhuiya A.A., Karsh R.K., Dutta S. AlexNet-CNN Based Feature Extraction and Classification of Multiclass ASL Hand Gestures. *Tools Appl.* 2021, 80, 3051–3069. (in Eng.)

D. Li, C. R. Opazo, X. Yu, and H. Li, "Word-level deep sign language recognition from video: A new large-scale dataset and methods comparison," *Proc. - 2020 IEEE Winter Conf. Appl. Comput. Vision, WACV 2020*. Pp. 1448–1458, 2020. Doi: 10.1109/WACV45572.2020.9093512. (in Eng.)

Eid A., Schwenker F. Visual Static Hand Gesture Recognition Using Convolutional Neural Network. *Algorithms* 2023, 16, 361. <https://doi.org/10.3390/a16080361> (in Eng.)

G.G. Abdullayeva, N.O. Alishzade (2022). Transfer learning for Azerbaijani Sign Language Recognition / *Informatics and Control Problems* 42. Issue 2. Pp. 65–72. (in Eng.)

Kenshimov C., Mukhanov S., Merembayev T., Yedilkhan D. (2021). A comparison of convolutional neural networks for Kazakh sign language recognition. *Eastern-European Journal of Enterprise Technologies*, 5 (2 (113)), 44–54. Doi: <https://doi.org/10.15587/1729-4061.2021.241535> (in Eng.)

Krizhevsky A., Sutskever I., Hinton G.E. Imagenet classification with deep convolutional neural networks. *Commun. ACM* 2017, 60, 84–90. (in Eng.)

Miah A.S.M., Hasan M.A.M., Shin J., Okuyama Y., Tomioka Y. Multistage Spatial Attention-Based Neural Network for Hand Gesture Recognition. *Computers* 2023, 12, 13. <https://doi.org/10.3390/computers12010013> (in Eng.)

Miah A.S.M., Shin J., Hasan M.A.M., Rahim M.A. BenSignNet: Bengali Sign Language Alphabet Recognition Using Concatenated Segmentation and Convolutional Neural Network. *Appl. Sci.* 2022, 12, 3933. (in Eng.)

S. Subburaj, S. Murugavalli, Survey on sign language recognition in context of vision-based and deep learning, *Measurement: Sensors* 23 (2022) 100385. <https://doi.org/10.1016/j.measen.2022.100385> (in Eng.)

Sharma S., Singh S. Vision-based sign language recognition system: A Comprehensive Review. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Seattle, WA, USA, 14–19. June 2020. Pp. 10023–10033. (in Eng.)

Simonyan K. and Zisserman A. (2015). Very Deep Convolutional Networks for Large-Scale Image Recognition. *The 3rd International Conference on Learning Representations (ICLR2015)*. <https://arxiv.org/abs/1409.1556> (in Eng.)

Tejaswini Ananthanarayana, Priyanshu Srivastava, Akash Chintla, Akhil Santha, Brian Landy, Joseph Panaro, 32 Andre Webster, Nikunj Kotecha, Shagan Sah, Thomastine Sarchet, Raymond Ptucha, and Ifeoma Nwogu. 33 2021. Deep Learning Methods for Sign Language Translation. *ACM Trans. Access. Comput.* 14, 4, Article 22 34 (September 2021). 30 p. 35 <https://doi.org/10.1145/3477498> (in Eng.)

Zholshiyeva L., Zhukabayeva T., Turaev Sh., Berdiyeva M., Sengirbaeva R. (2023). Real-time Kazakh Sign Language Recognition Using Media Pipe and SVM. *News of the National Academy of Sciences of the Republic of Kazakhstan Physico-mathematical Series. ISSN 1991–346X. Volume 1. Number 345. Pp. 82–93.* <https://doi.org/10.32014/2022.2518-1726.170> (in Eng.).

Zholshiyeva L., Zhukabayeva T., Turaev Sh., Berdiyeva M., Sengirbaeva R. (2023). Real-time Kazakh Sign Language Recognition Using Media Pipe and SVM. *News of the National Academy of Sciences of the Republic of Kazakhstan Physico-mathematical Series. ISSN 1991–346X. Volume 1. Number 345. Pp. 82–93.* <https://doi.org/10.32014/2022.2518-1726.170> (in Eng.).

NEWS OF THE NATIONAL ACADEMY OF SCIENCES OF THE REPUBLIC OF  
KAZAKHSTAN  
PHYSICO-MATHEMATICAL SERIES  
ISSN 1991-346X  
Volume 3. Number 347 (2023). 88–98  
<https://doi.org/10.32014/2023.2518-1726.206>

UDC 004.056

© **K. Kadirkulov, A. Ismailova, A. Beissegul, 2023**  
NAO «S. Seifullin Kazakh Agrotechnical Research University», Astana,  
Kazakhstan.  
E-mail: [kkuanyshe@gmail.com](mailto:kkuanyshe@gmail.com)

### SELECTION OF A MACHINE LEARNING MODEL FOR INTERPRETING LABORATORY RESULTS

**Kuanyshe Kadirkulov** — Doctoral student of NAO «S. Seifullin Kazakh Agrotechnical Research University», Astana, Kazakhstan.

E-mail: [kkuanyshe@gmail.com](mailto:kkuanyshe@gmail.com). ORCID ID: <https://orcid.org/0000-0003-0506-4890>;

**Aisulu Ismailova** — PhD Doctor, Associate Professor of the Department of Information Systems, NAO «S. Seifullin Kazakh Agrotechnical Research University», Astana, Kazakhstan.

E-mail: [a.ismailova@mail.ru](mailto:a.ismailova@mail.ru). ORCID ID: <https://orcid.org/0000-0002-8958-1846>;

**Aliya Beissegul** — DBA doctoral student, NAO «Almaty Management University», Almaty, Kazakhstan.

E-mail: [beissegul@gmail.com](mailto:beissegul@gmail.com). ORCID ID: <https://orcid.org/0000-0002-0053-2539>.

**Abstract.** The purpose of the presented work is the process of automating the interpretation of the results of laboratory tests. The system of modern health care puts before clinical laboratory diagnostics the task of rapid and qualitative determination of results. The solution to such an urgent problem lies in the automation of processes at all stages of diagnostic studies, which will reduce the time of obtaining results, minimize human participation and obtain reliable results. The presented work describes the process of selection and formation of a stack of machine learning models for interpretation of laboratory results. Machine learning is used to generate a knowledge base that is further applied by artificial intelligence (AI). This approach aims to streamline the process of interpreting laboratory results and provide more accurate, consistent and timely results to healthcare providers. Automating the interpretation of laboratory results has the potential to improve the efficiency and accuracy of interpreting laboratory results, leading to better patient outcomes and better clinical decisions. However, it is important to note that AI models are not perfect and can still make errors, and that healthcare providers should always review the results of automated interpretation before making a diagnosis or treatment decision. The scientific and practical significance of the work lies in the approach to the selection of a stack of machine learning models,



which can also be applied to solutions of other problems. The paper considers machine learning methods such as neural network, gradient boosting and random forest. The models were validated over 7.6 million anonymized laboratory results. The result of the presented work is a stacking model of machine learning methods, which can be optimally applied in building a knowledge base for identification of various pathologies.

**Keywords:** Automation, machine learning, artificial intelligence, laboratory research, neural network, random forest, gradient boosting

© **К.К. Кадиркулов, А.А. Исмаилова, Ә.Б. Бейсеғұл, 2023**

«С. Сейфуллин атындағы Қазақ агротехникалық зерттеу университеті»

КеАҚ, Астана, Қазақстан.

E-mail: kkuanysh@gmail.com

### **ЛАБОРАТОРИЯЛЫҚ ЗЕРТТЕУ НӘТИЖЕЛЕРІН ТАЛДАУ ҮШІН МАШИНАЛЫҚ ОҚЫТУДЫҢ МОДЕЛІН ТАҢДАУ**

**Кадиркулов Куаныш Кайсарович** — «С. Сейфуллин атындағы Қазақ агротехникалық зерттеу университеті» КеАҚ докторанты, Астана, Қазақстан.

E-mail: kkuanysh@gmail.com. ORCID ID: <https://orcid.org/0000-0003-0506-4890>;

**Исмаилова Айсулу Абжаппаровна** — PhD доктор, «С. Сейфуллин атындағы Қазақ агротехникалық зерттеу университеті» КеАҚ Ақпараттық жүйелер кафедрасының қауымдастырылған профессоры, Астана, Қазақстан.

E-mail: a.ismailova@mail.ru. ORCID ID: <https://orcid.org/0000-0002-8958-1846>;

**Бейсеғұл Әлия Болатбековна** — DBA докторанты, «Алматы Менеджмент Университеті» КеАҚ, Алматы, Қазақстан.

E-mail: beissegul@gmail.com. ORCID ID: <https://orcid.org/0000-0002-0053-2539>.

**Аннотация.** Ұсынылған жұмыстың мақсаты — зертханалық зерттеу нәтижелерін интерпретациялауды автоматтандыру процесі. Заманауи денсаулық сақтау жүйесі клиникалық зертханалық диагностиканың нәтижелерін тез және сапалы анықтау міндетін қояды. Мұндай өзекті мәселені шешу диагностикалық зерттеулердің барлық кезеңдеріндегі процестерді автоматтандыруда жатыр, бұл нәтижелерді алу уақытын қысқартуға, адамның қатысуын азайтуға және сенімді нәтижелерге қол жеткізуге мүмкіндік береді. Ұсынылған жұмыс зертханалық зерттеулердің нәтижелерін интерпретациялау үшін машиналық оқыту үлгілерінің стекін таңдау және қалыптастыру процесін сипаттайды. Машиналық оқытудың көмегімен білім базасы қалыптасады, оны әрі қарай жасанды интеллект (AI) қолданады. Бұл тәсіл зертханалық нәтижелерді интерпретациялау процесін оңтайландыруға және медицина мамандарына дәлірек, дәйекті және уақтылы нәтижелерді ұсынуға бағытталған. Зертханалық нәтижелерді интерпретациялауды автоматтандыру зертханалық нәтижелерді интерпретациялаудың тиімділігі мен дәлдігін жақсартуға мүмкіндік береді, бұл емделуші нәтижелерін жақсартуға және жақсы клиникалық шешім қабылдауға әкеледі. Дегенмен,

AI үлгілері мінсіз емес және әлі де қателесуі мүмкін екенін және денсаулық сақтау мамандары диагноз немесе емдеу туралы шешім қабылдамас бұрын әрқашан автоматтандырылған интерпретация нәтижелерін қарап шығуы керек екенін ескеру маңызды. Жұмыстың ғылыми және практикалық маңыздылығы басқа мәселелерді шешуде де қолданылуы мүмкін машиналық оқыту үлгілерінің стекін таңдау тәсілінде жатыр. Бұл мақалада нейрондық желі, градиентті күшейту және кездейсоқ орман сияқты машиналық оқыту әдістері қарастырылады. Модельдік валидация 7,6 миллион иесіздендірілген зертханалық нәтижелер бойынша орындалды. Ұсынылған жұмыстың нәтижесі әртүрлі патологияларды анықтау үшін білім қорын қалыптастыруда оңтайлы пайдаланылуы мүмкін машиналық оқыту әдістерінің стектік моделі болып табылады.

**Түйін сөздер:** Автоматтандыру, машиналық оқыту, жасанды интеллект, зертханалық зерттеулер, нейрондық желі, кездейсоқ орман, градиентті күшейту

© **К.К. Кадиркулов, А.А. Исмаилова, Ә.Б. Бейсеғұл, 2023**  
НАО «Казахский агротехнический университет имени С. Сейфуллина»,  
Астана, Казахстан.  
E-mail: kkuanysh@gmail.com

## **ВЫБОР МОДЕЛИ МАШИННОГО ОБУЧЕНИЯ ПО ИНТЕРПРЕТАЦИИ РЕЗУЛЬТАТОВ ЛАБОРАТОРНЫХ ИССЛЕДОВАНИЙ**

**Кадиркулов Куаныш Кайсарович** — докторант НАО «Казахский агротехнический исследовательский университет имени С. Сейфуллина», г. Астана, Казахстан.

E-mail: kkuanysh@gmail.com. ORCID ID: <https://orcid.org/0000-0003-0506-4890>;

**Исмаилова Айсулу Абжаппаровна** — PhD доктор, Ассоциированный профессор кафедры Информационных систем НАО «Казахский агротехнический исследовательский университет имени С. Сейфуллина», г. Астана, Казахстан.

E-mail: a.ismailova@mail.ru. ORCID ID: <https://orcid.org/0000-0002-8958-1846>;

**Бейсеғұл Әлия Болатбековна** — DBA докторант, НАО «Алматы Менеджмент Университет», г. Алматы, Казахстан.

E-mail: beissegul@gmail.com. ORCID ID: <https://orcid.org/0000-0002-0053-2539>.

**Аннотация.** Целью представленной работы является процесс автоматизации интерпретации результатов лабораторных исследований. Система современного здравоохранения ставит перед клинической лабораторной диагностикой задачу быстрого и качественного определения результатов. Решение такой актуальной проблемы лежит в автоматизации процессов на всех этапах диагностических исследований, что позволит сократить время получения результатов, минимизировать участие человека и получить достоверные результаты. В представленной работе описывается процесс выбора и формирование стека моделей машинного обучения

по интерпретации результатов лабораторных исследований. С помощью машинного обучения формируется база знаний, которая в дальнейшем применяется искусственным интеллектом (ИИ). Этот подход направлен на рационализацию процесса интерпретации результатов лабораторных исследований и предоставление более точных, последовательных и своевременных результатов медицинским работникам. Автоматизация интерпретации результатов лабораторных исследований потенциально может повысить эффективность и точность интерпретации результатов лабораторных исследований, что приведет к улучшению состояния пациентов и принятию более правильных клинических решений. Однако важно отметить, что модели ИИ не совершенны и все еще могут допускать ошибки, и что медицинские работники должны всегда просматривать результаты автоматизированной интерпретации перед принятием диагноза или решения о лечении. Научная и практическая значимость работы заключается в подходе к выбору стека моделей машинного обучения, которая также может быть применена в решениях других задач. В работе рассматриваются методы машинного обучения такие как нейронная сеть, градиентный бустинг и случайный лес. Проверка моделей производилась над 7,6 млн обезличенных результатов лабораторных исследований. Результатом представленной работы является модель стекинга методов машинного обучения, которая оптимально может применяться при формировании базы знаний для идентификации различных патологий.

**Ключевые слова:** Автоматизация, машинное обучения, искусственный интеллект, лабораторные исследования, нейронная сеть, случайный лес, градиентный бустинг

### **Введение**

Президент Республики Казахстан Касым-Жомарт Токаев 4 марта 2020 года на совещании по реализации Государственной программы «Цифровой Казахстан», отметил, что цифровизация направлена не для развития одного сектора, а всей экономики государства и в преобразовании общества в целом. Президент подчеркнул, что в Республике Казахстан имеется зависимость от зарубежных разработок и технологий, в связи с чем, было поручено на законодательном уровне поддержать IT-компании казахстанского производства и обеспечить им приоритетность в конкуренции в государственном сектор (Глава государства провел совещание по реализации..., 2020).

Основной целью интерпретации результатов является первичное диагностирование и трактовка показателей на основе ререференсных значений каждого исследования, где отклонение от референсных значений будет считаться наличием проблем в организме пациента (Рисунок 1).

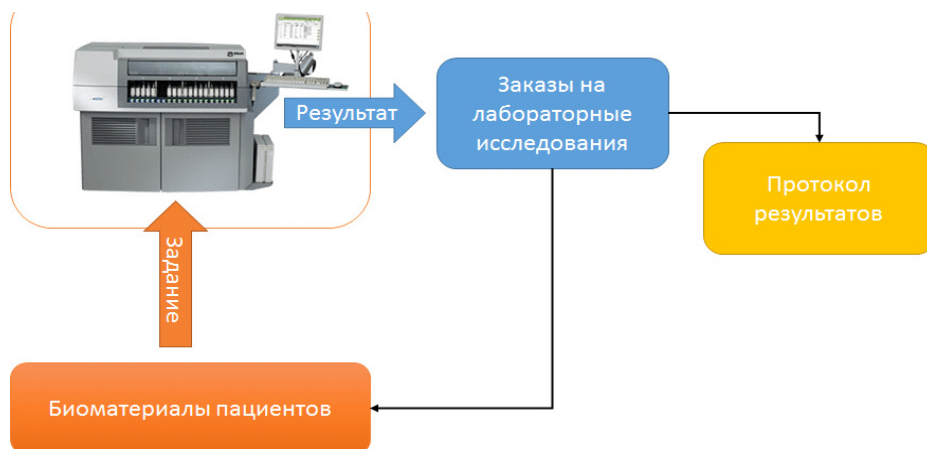


Рисунок 1. Общая схема процесса интерпретации

Преимущества алгоритмов и интерпретаций включают предотвращение ошибочных диагнозов, сокращение количества необходимых лабораторных тестов, сокращение количества процедур, переливаний крови и госпитализаций, сокращение времени, необходимого для постановки диагноза, сокращение ошибок при заказе тестов и предоставление дополнительной информации о том, как результаты лабораторных исследований могут повлиять на другие аспекты ухода за пациентом. Сочетание алгоритмов и интерпретаций для лабораторных исследований имеет множество преимуществ для медицинского обслуживания пациента (Кадиркулов и др., 2020).

Практика здравоохранения включает в себя сбор всех видов данных о пациенте, которые помогут врачу правильно диагностировать состояние здоровья пациента. Этими данными могут быть простые симптомы, наблюдаемые пациентом, первоначальный диагноз врача или подробные результаты анализов, полученные в лаборатории. Таким образом, эти данные используются только для анализа врач, который затем устанавливает заболевание, используя свой личный медицинский опыт (Jaskins и др., 2020). Из-за современного образа жизни болезни стремительно растут. Наш образ жизни и питание привычки оказывают влияние на наше здоровье, вызывая сердечно-сосудистые заболевания и другие проблемы со здоровьем. Техника добычи данных является одной из самых сложных и ведущих исследовательских областей в здравоохранении из-за высокой важности ценных данных (Renjit, Shunmuganathan, 2010).

Результаты клинических и лабораторных анализов часто рассматриваются как нормальные или ненормальные, а не как просто информативными. Для количественных результатов нормальность обычно определяется как диапазон значений, простирающийся на два стандартных отклонения от среднего значения, полученного из популяции "нормальных" людей аналогичного возраста, веса, пола и т.д. Согласно этому общепринятому

определению, нормальный диапазон или референсный интервал должен включать приблизительно 95 % тестируемой популяции. Предположение, лежащее в основе этого подхода, является то, что результат теста (независимая переменная в данном случае) имеет кривую распределения частот, которая имеет форму колокола. Это верно для многих количественных измерений проводимых в медицинской практике. Однако для многих других тестов, для определения "ненормальность", поскольку предположения о нормальном распределении не выполняются (например, аналитические стандартные отклонения увеличиваются с ростом значений в пределах интервала измерения или тестовые значения включают отрицательные значения) (McPherson, Pincus, 2017).

Машинное обучение (ML) оказало огромное влияние на многие сферы жизни современного общества. Например, оно используется для фильтрации спам-сообщений из текстовых документов, таких как электронная почта, анализа различных изображений для выявления различий, а также для извлечения важных данных из больших массивов с помощью интеллектуального анализа данных. ML позволяет выявлять закономерности, строить модели и делать прогнозы на основе обучающих данных (Esteva et al., 2019).

Но как показывает практика, использование структурированных данных намного ускоряет процесс прогнозирования, применения алгоритмов машинного обучения, статистического анализа и др. методик. Структуризация данных производится путем комплексной автоматизации. Комплексная автоматизация медицинской лаборатории означает интеграцию цифровых технологий, таких как искусственный интеллект (ИИ), робототехника и системы электронной медицинской документации (ЭМК), в процесс работы медицинской лаборатории. Целью комплексной автоматизации является рационализация и оптимизация лабораторного процесса, что приводит к повышению эффективности, точности и улучшению результатов лечения пациентов.

Искусственный интеллект способен обрабатывать большие объемы данных, сравнивать и анализировать данные, особенно ценится в медицине и здравоохранении в целом (Pin et al., 2017). Роберт Гриневиц, специалист по анализу данных, считает, что медицинские работники нуждаются в любой помощи, которую они могут получить. Люди могут выявлять закономерности в данных, но это может быть утомительным процессом, для которого лучше подходят машины. особенно при наличии большого количества переменных или сценариев. Объедините этот факт с перегруженностью работой и нехваткой времени, с которыми приходится сталкиваться врачам, становится еще проще еще проще пропустить признаки, которые могут повлиять на диагноз. Искусственный интеллект в здравоохранении может помочь, обнаружив сигналы, которые благонамеренные врачи могут пропустить (Болотова, 2012).

#### **Описание материалов и методов**

Целью представленной работы является процесс формирования и выбор

моделей машинного обучения для интерпретации результатов лабораторных исследований. Система современного здравоохранения ставит перед клинической лабораторной диагностикой задачу быстрого и качественного определения результатов. Решение такой актуальной проблемы лежит в автоматизации процессов на всех этапах диагностических исследований, что позволит сократить время получения результатов, минимизировать участие человека и получить достоверные результаты.

На первом этапе производилась обработка 7,6 млн результатов лабораторных исследований. В данном исследовании был проведен логистический регрессионный анализ для прогнозирования зависимой переменной. Целью анализа было понять взаимосвязь между независимыми переменными и зависимой переменной и сделать прогнозы относительно зависимой переменной на основе независимых переменных. В результате была определена связь между TSH (тиреотропный гормон) + возраст с другими тестами (Рисунок 2).

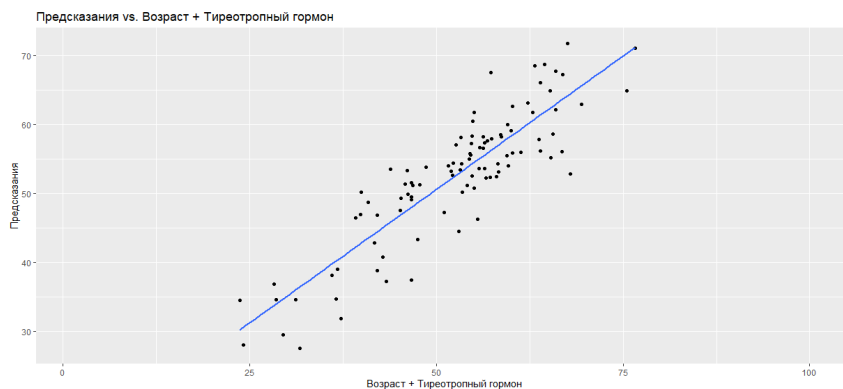


Рисунок 2. Ассоциация между TSH (тиреотропный гормон) + возраст с другими тестами

На втором этапе, после определения теста, были сформированы данные (более 1,8 млн) по тестам имеющие тесную связь с тиреотропный гормоном. Это тесты тиреотропный гормон (ТТГ) в комбинации с трийодтиронином (Т3) и тетраiodтиронином (Т4) необходимых для определения патологии щитовидной железы.

На третьем этапе, было произведено обучение нейронной сети по сформированным данным, по следующей структуре:

- Возраст;
- Пол;
- Значение ТТГ;
- Значение Т3;
- Значение Т4;
- Признак патологии.

Модель нейронной сети был построен по 3 скрытым слоям (Рисунок 3). История обучения модели демонстрирует следующие показатели:

- Потери при обучении: 0,2732;
- Точность обучения: 0,8949;
- Потери при проверке: 0,338;
- Точность валидации: 0,8612

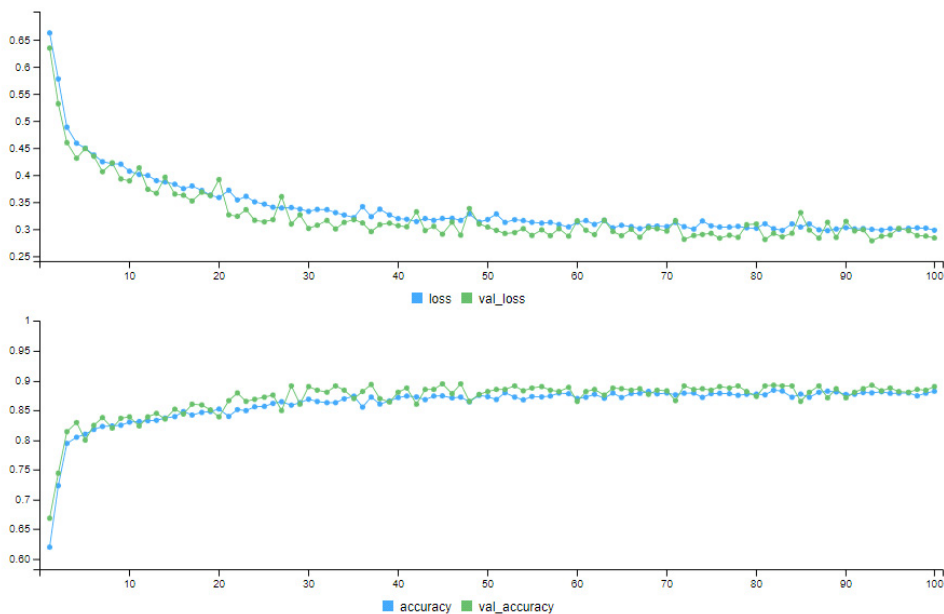


Рисунок 3. График обучения нейронной сети

Модель указала на необходимость применения альтернативных методов машинного обучения или комбинации методов, включая усиление нейронных сетей через градиентный бустинг (GB), случайные леса (RF) и метод опорных векторов (SVM). Ансамблевое машинное обучение представляет собой подход, где различные модели обучаются с целью решения одной задачи и затем комбинируются для достижения оптимальных результатов. Основная идея состоит в том, что комбинированное решение нескольких моделей обычно более точно, чем решение отдельной модели. Для реализации ансамблевого подхода важно провести сравнительный анализ предложенных методов, чтобы выбрать наиболее эффективный. Этот процесс начинается с создания трех моделей на одном и том же наборе данных, затем строится матрица ошибок для оценки и сравнения производительности каждой модели (Рисунок 4).



Рисунок 4. Сравнение производительности моделей

Из рисунка 4 можно сформировать следующую таблицу результатов сравнительного анализа (табл. 1).

Таблица 1. Сравнительный анализ моделей машинного обучения

Модель	Точность	Чувствительность	Специфичность	PPV	NPV	Каппа
SVM	0,8556	0,9709	0,6619	0,8283	0,9312	0,6722
GB	0,9369	0,9673	0,8859	0,9344	0,9416	0,8635
RF	0,9422	0,9588	0,9145	0,9496	0,9296	0,8761

В табл. 1 представлены результаты сравнения трех моделей машинного обучения: метод опорных векторов (SVM), градиентный бустинг (GB) и случайный лес (RF). Рассмотрены следующие метрики: точность (Accuracy), чувствительность (Sensitivity), специфичность (Specificity), положительная прогностическая ценность (PPV), отрицательная прогностическая ценность (NPV) и коэффициент Каппа (Карра). Согласно данным таблицы, модель на основе случайного леса является лидером по точности (0,9422) и коэффициенту Каппа (0,8761), что свидетельствует о ее выдающейся производительности в данном контексте. Градиентный бустинг также показал себя хорошо, достигая точности 0,9369 и коэффициента Каппа 0,8635. В то время как SVM-модель обладает меньшей точностью (0,8556) и коэффициентом Каппа (0,6722), она выделяется высокой чувствительностью (0,9709). Это может быть критично в сценариях, где высокая способность корректно идентифицировать положительные случаи является ключевой.

Сравнительная таблица показала, что для представленных данных более подходящим являются методы машинного обучения случайный лес и градиентный бустинг. Следовательно, используя данные методы можно построить ансамбль [GB+NN] и [RF+NN]. При построении ансамблевой модели производим обучение модели по GB и RF.

### Результаты

В результате ансамбль моделей градиент бустинг (GB) и нейронная сеть



(NN) демонстрируют значительное улучшение производительности по сравнению с отдельными моделями, чему следуют следующие результаты:

- Точность:  $(TP+TN)/(TP+TN+FP+FN) = (1961 + 3300) / (1961 + 3300 + 2 + 3) = 0.99905$ , что означает, что модель правильно классифицировала 99,9 % случаев.

- 95 % CI: (0,9978, 0,9997) представляет собой доверительный интервал для точности; мы можем быть на 95 % уверены, что истинная точность лежит между 99,78 % и 99,97 %.

- P-Value [Acc > NIR]:  $2e-16$ , вероятность соблюдения точности модели, если бы модель не имела информации. Очень маленькое р-значение указывает на то, что модель значительно лучше, чем случайное угадывание.

- Коэффициент отсутствия информации (NIR): 0,627 — базовая точность, если мы всегда предсказываем класс большинства (в данном случае класс 0).

- Карра: 0,998, мера согласия между предсказанными и фактическими значениями с учетом случайности. Значение, близкое к 0, указывает на плохое согласие, а значение, близкое к 1, указывает на идеальное согласие.

### **Заключение**

Исследование показало, что методы машинного обучения для набора данных по выявлению отклонению при диагностике щитовидной железы могут работать как отдельная модель, так и с применением методики ансамбля. В результате согласно рисунку 4 можно сказать, что высокие показатели были у отдельных моделей как случайный лес и градиентный бустинг, и методика ансамбль стек между градиентным бустингом и нейронной сетью.

### **ЛИТЕРАТУРА**

Глава государства провел совещание по реализации Государственной программы «Цифровой Казахстан» События [Электронный ресурс]. URL: [https://www.akorda.kz/ru/events/akorda\\_news/meetings\\_and\\_sittings/glava-gosudarstva-provel-soveshchanie-po-realizacii-gosudarstvennoi-programmy-cifrovoi-kazahstan](https://www.akorda.kz/ru/events/akorda_news/meetings_and_sittings/glava-gosudarstva-provel-soveshchanie-po-realizacii-gosudarstvennoi-programmy-cifrovoi-kazahstan).

К.К. Кадиркулов, А.А. Исмаилова, Г.Ж. Солтан, А.А. Муханова, М.Маханов, Автоматизация идентификации отклонения результатов лабораторных исследований, 2020. Вестник КазНУТУ. ISSN 1680–9211, №4(140), стр. 127–133.

Jackins V., Vimal S., Kaliappan M. & Lee, M.Y. (2020). AI-based smart prediction of clinical disease using random forest classifier and Naive Bayes. The Journal of Supercomputing, 77(5), 5198–5219. doi:10.1007/s11227-020-03481-x

Renjit J.A., Shunmuganathan K.L. (2010). Distributed and cooperative multi-agent based intrusion detection system. Indian J Sci Technol 3(10):1070–1074

McPherson R.A., Pincus M.R., editors. Henry's clinical diagnosis and management by laboratory methods, 23rd ed. St. Louis: Elsevier, 2017.

Esteva A. et al. A guide to deep learning in healthcare. Nat. Med. 25, 24–29 (2019).

I.V. Ilin, A.I. Levina, O.Yu. Iliashenko, Proceedings of the 29th International Business Information Management Association Conference - Education Excellence and Innovation Management through Vision 2020: From Regional Development Sustainability to Global Economic Growth 2017, 1822–1831 (2017)

Болотова Л.С. Системы искусственного интеллекта: модели и технологии, основанные на знаниях. М. 2012. 664 с

#### REFERENCES

The head of state held a meeting on the implementation of the State Program "Digital Kazakhstan" Events [Electronic resource]. URL: [https://www.akorda.kz/ru/events/akorda\\_news/meetings\\_and\\_sittings/glava-gosudarstva-provel-soveshchanie-po-realizacii-gosudarstvennoi-programmy-cifrovoi-kazahstan](https://www.akorda.kz/ru/events/akorda_news/meetings_and_sittings/glava-gosudarstva-provel-soveshchanie-po-realizacii-gosudarstvennoi-programmy-cifrovoi-kazahstan).

K.K. Kadirkulov, A.A. Ismailova, G.J. Soltan, A.A. Mukhanova, M. Makhanov, Automation of identification of deviation of laboratory research results, 2020. Bulletin of KazNITU. ISSN 1680-9211. No. 4(140). Pp. 127–133.

Jackins V., Vimal S., Kaliappan M. & Lee M. Y. (2020). AI-based smart prediction of clinical disease using random forest classifier and Naive Bayes. *The Journal of Supercomputing*, 77(5), 5198–5219. Doi:10.1007/s11227-020-03481-x

Renjit J.A., Shunmuganathan K.L. (2010). Distributed and cooperative multi-agent based intrusion detection system. *Indian J Sci Technol* 3(10):1070–1074

McPherson R.A., Pincus M.R., editors. *Henry's clinical diagnosis and management by laboratory methods*, 23rd ed. St. Louis: Elsevier, 2017.

Esteva A. et al. A guide to deep learning in healthcare. *Nat. Med.* 25, 24–29 (2019).

I.V. Ilin, A.I. Levina, O.Yu. Iliashenko, Proceedings of the 29th International Business Information Management Association Conference - Education Excellence and Innovation Management through Vision 2020: From Regional Development Sustainability to Global Economic Growth 2017, 1822–1831 (2017)

Bolotova L.S. *Artificial intelligence systems: models and knowledge-based technologies*. M. 2012. 664 c

NEWS OF THE NATIONAL ACADEMY OF SCIENCES OF THE REPUBLIC OF  
KAZAKHSTAN  
PHYSICO-MATHEMATICAL SERIES  
ISSN 1991-346X  
Volume 3. Number 347 (2023). 99–111  
<https://doi.org/10.32014/2023.2518-1726.207>

UDC 004.021

© **A. Mukashova, A. Mukanova\*, T. Ospanova, A. Bakiyeva,  
V. Makhatova, 2023**

Eurasian National University named after L.N. Gumilyov, Astana, Kazakhstan.  
E-mail: [ainurmukashovaonlasin@gmail.ru](mailto:ainurmukashovaonlasin@gmail.ru)

### **IMPORTANT ASPECTS OF DEVELOPING EDUCATIONAL PROGRAMS BASED ON THE COMPETENCY-BASED APPROACH**

**Mukashova Ainur** — doctoral student of L.N. Gumilyov Eurasian National University. 010000. Astana, Kazakhstan

E-mail: [ainurmukashovaonlasin@gmail.ru](mailto:ainurmukashovaonlasin@gmail.ru). ORCID ID: <https://orcid.org/0009-0007-2794-4748>;

**Mukhanova Ayagoz** — PhD, Department of Information technology, Faculty of information technologies, L.N. Gumilyov Eurasian National University, Astana, Kazakhstan,

E-mail: [ayagoz198302@mail.ru](mailto:ayagoz198302@mail.ru), <https://orcid.org/0000-0003-3987-0938>;

**Ospanova Tleugaisha** — Candidate of Technical Sciences, P.h.D., Senior Lecturer, L.N. Gumilyov Eurasian National University. 010000, Astana, Kazakhstan

E-mail: [Tleu2009@mail.ru](mailto:Tleu2009@mail.ru). ORCID ID: <https://orcid.org/0000-0002-1729-1321>;

**Bakiyeva Aigerim** — Candidate of Technical Sciences, P.h.D., Senior Lecturer, L.N. Gumilyov Eurasian National University. 010000, Astana, Kazakhstan

E-mail: [m\\_aigerim0707@mail.ru](mailto:m_aigerim0707@mail.ru). ORCID ID: <https://orcid.org/0000-0001-7186-6408>;

**Makhatova Valentina** — Candidate of Technical Sciences, Kh.Dosmukhamedov Atyrau University. 060000, Atyrau, Kazakhstan

E-mail: [mahve@mail.ru](mailto:mahve@mail.ru). ORCID ID: <https://orcid.org/0000-0002-4082-9193>.

**Abstract.** With the emergence of new spheres of professional activity, new professions and professional standards appear. It is necessary to improve the education system, which should provide the digital economy with competent personnel. This article analyzes modern trends in the development of educational programs; the approach of forming educational programs based on professional competencies is considered. The task of training qualified specialists on the basis of competence-based approach to training in IT areas is considered. Competence-based approach to training is currently the main aspect in the sphere of higher education. Formation of competencies for the design of basic educational programs of higher education is declared on the basis of professional standards in the framework of the national chamber of entrepreneurs of the Republic of Kazakhstan "Atameken", but the labor functions described in professional standards often require structuring under the needs for the development of an intelligent system of designing the

optimal competence model of a graduate of higher education institution. The article presents an algorithm for structuring professional programs in .xls format for the development of intelligent design systems, such as the formulation of professional, unique and general educational competencies, expert examination of educational programs, data processing and analysis. We hope that the obtained structured data of the professional standard will serve as a basis for designing a competency model, formulation of competencies providing for the development of future integrated, comprehensive and accessible educational programs with the help of which the result of education will contribute to the training of quality, highly qualified workers with a set of necessary competencies that meet the requirements of employers and allow to effectively carry out professional activities. The first part shows the general structure of the professional standard. In the second part semantic linkage is developed on the example of professional standard on database administration. In the third part data structuring using the library Xceed.Words.NET in C# language.

**Keywords:** Educational program, professional standards, competency-based approach in education

© А. Мукашова, А. Муханова\*, Т. Оспанова, А. Бакиева,  
В. Махатова, 2023

Л.Н.Гумилев атындағы Еуразия Ұлттық университеті, Астана, Қазақстан.  
E-mail: ainurmukashovaonlasin@gmail.ru

## ҚҰЗЫРЕТТІК ТӘСІЛДЕР НЕГІЗІНДЕГІ БІЛІМ БЕРУ БАҒДАРЛАМАЛАРЫН ӘЗІРЛЕУДІҢ МАҢЫЗДЫ АСПЕКТІЛЕРІ

**Мукашова Айну́р** — Л.Н.Гумилев атындағы Еуразия ұлттық университетінің доктаранты.  
010000, Астана, Қазақстан

E-mail: ainurmukashovaonlasin@gmail.ru. ORCID ID: <https://orcid.org/0009-0007-2794-4748>;

**Муханова Аягөз** — PhD, Л.Н. Гумилев атындағы ЕҰУ, Ақпараттық технологиялар факультеті, Ақпараттық жүйелер кафедрасы, Астана, Қазақстан.

E-mail: ayagoz198302@mail.ru, <https://orcid.org/0000-0003-3987-0938>;

**Оспанова Тілеуғайша** — т.ғ.к., Л.Н. Гумилев атындағы Еуразия Ұлттық университетінің м.а. доценті. 010000, Астана, Қазақстан

E-mail: Tleu2009@mail.ru. ORCID ID: <https://orcid.org/0000-0002-1729-1321>;

**Бакиева Айгерім** — т.ғ.к., Ph.D., Л.Н. Гумилев атындағы Еуразия Ұлттық университетінің аға оқытушысы. 010000, Астана Қазақстан.

E-mail: m\_aigerim0707@mail.ru. ORCID ID: <https://orcid.org/0000-0001-7186-6408>;

**Махатова Валентина** — т.ғ.к., Х.Досмұхамедов атындағы Атырау университетінің профессоры. 060000, Атырау, Қазақстан

E-mail: mahve@mail.ru. ORCID ID: <https://orcid.org/0000-0002-4082-9193>.

**Аннотация.** Жаңа кәсіптік қызмет салаларының пайда болуымен қатар жаңа кәсіптер мен кәсіби стандарттар пайда болады. Цифрлық экономиканы құзыретті кадрлармен қамтамасыз етуге тиіс білім беру жүйесін жетілдіру қажет. Мақалада білім беру бағдарламаларын әзірлеудің қазіргі тенденцияларына талдау; кәсіби құзыреттілікке негізделген білім

беру бағдарламаларын қалыптастыру тәсілі қарастырылады. АТ бағыттары бойынша оқытуға құзыреттілік тәсіл негізінде білікті мамандарды даярлау міндеті қарастырылады. Оқытуға құзыреттілік көзқарас қазіргі уақытта жоғары білім берудің негізгі аспектісі болып табылады. Жоғары білім берудің негізгі білім беру бағдарламаларын жобалау үшін құзыреттерді қалыптастыру Қазақстан Республикасының "Атамекен" Ұлттық Кәсіпкерлер палатасы қабылдаған шеңбердегі кәсіби стандарттар негізінде декларацияланады, алайда кәсіптік стандарттарда сипатталған еңбек функциялары көбінесе ЖОО түлегінің оңтайлы құзыреттілік моделін жобалаудың зияткерлік жүйесін әзірлеу үшін құрылымдауды талап етеді. Мақалада кәсіби бағдарламаларды .xls форматына құрылымдау алгоритмі берілген. Ол алгоритм бірегей және жалпы білім беру құзыреттіліктерін тұжырымдау, білім беру бағдарламаларына сараптама жүргізу, деректерді өңдеу және талдау сияқты интеллектуалды жобалау жүйелерін әзірлеуге бағытталған. Сонымен қатар алынған құрылымдық деректер құзыреттілік моделін жобалауға, болашақ интеграцияланған, кешенді және қолжетімді білім беру бағдарламаларын әзірлеуді қамтамасыз етеді деген сенімдеміз. Бірінші бөлімде кәсіби бағдарламаның жалпы құрылымы көрсетілген. Екінші бөлімде мәліметтер базасын басқару кәсіби стандартының мысалында семантикалық байланыс жасалды. Үшінші бөлімде C# тілінің Xceed.Words.NET. кітапханасын пайдалану арқылы деректерді құрылымдау қарастырылған.

**Түйін сөздер.** Білім беру бағдарламасы, кәсіби стандарттар, білім берудегі құзыреттілік тәсіл

© А. Мукашова, А. Муханова\*, Т. Оспанова, А. Бакиева,  
В. Махатова, 2023

Л.Н.Гумилев атындағы Еуразия ұлттық университеті, Астана, Қазақстан.

E-mail: ainurmukashovaonlasin@gmail.ru

### **ВАЖНЫЕ АСПЕКТЫ РАЗРАБОТКИ ОБРАЗОВАТЕЛЬНЫХ ПРОГРАММ, ОСНОВАННЫХ НА КОМПЕТЕНТНОСТНОМ ПОДХОДЕ**

**Мукашова Айнур** — докторант, Евразийский национальный университет имени Л.Н. Гумилева. 010000, Астана, Казахстан

E-mail: ainurmukashovaonlasin@gmail.ru. ORCID ID: <https://orcid.org/0009-0007-2794-4748>;

**Муханова Аягоз** — PhD, кафедра Информационные системы, факультет Информационных технологий. Евразийский национальный университет имени Л.Н. Гумилева, Астана, Казахстан.

E-mail: ayagoz198302@mail.ru, <https://orcid.org/0000-0003-3987-0938>;

**Оспанова Тлеугайша** — к.т.н., и.о.доцент, старший преподаватель, Евразийский национальный университет имени Л.Н. Гумилева. 010000, Астана, Казахстан

E-mail: Tleu2009@mail.ru. ORCID ID: <https://orcid.org/0000-0002-1729-1321>;

**Бакиева Айгерим** — к.т.н., Ph.D., старший преподаватель, Евразийский национальный университет имени Л.Н. Гумилева. 010000, Астана, Казахстан.

E-mail: m\_aigerim0707@mail.ru. ORCID ID: <https://orcid.org/0000-0001-7186-6408>;

**Махатова Валентина** — к.т.н., профессор, Атырауский университет имени Х.Досмухамедова. 060000, Атырау, Казакстан  
E-mail: mahve@mail.ru. ORCID ID: <https://orcid.org/0000-0002-4082-9193>.

**Аннотация.** С появлением новых сфер профессиональной деятельности появляются новые профессии и профессиональные стандарты. Необходимо совершенствовать системы образования, которые должны обеспечивать цифровую экономику компетентными кадрами. В данной статье проведен анализ современных тенденций разработки образовательных программ; рассматривается подход формирования образовательных программ, основанных на профессиональных компетенциях. Рассматривается задача подготовки квалифицированных специалистов на основе компетентного подхода к обучению по направлениям ИТ. Компетентный подход к обучению в настоящее время является основным аспектом в сфере высшего образования. Формирование компетенций для проектирования основных образовательных программ высшего образования декларируется на основе профессиональных стандартов в рамках принятых национальной палате предпринимателей Республики Казакстан «Атамекен», однако трудовые функции, описанные в профессиональных стандартах, зачастую требуют структуризацию под нужды для разработки интеллектуальной системы проектирования оптимальной компетентностной модели выпускника ВУЗа. В статье представлен алгоритм структуризации профессиональных программ в формате .xls для разработки интеллектуальных систем проектирования, таких как формулировки профессиональных, уникальных и общее образовательных компетенций, проведении экспертизы образовательных программ, обработки и анализа данных. Мы надеемся, что полученные структурированные данные профессионального стандарта послужат основой проектированию модели компетенций, формулировку компетенций обеспечивающий для разработки будущих интегрированных, комплексных и доступных образовательных программ с помощью которого результат образования будет способствовать в подготовке качественных, высококвалифицированных работников с набором необходимых компетенций, отвечающих требованиям работодателей и позволяющих эффективно осуществлять профессиональную деятельность. В первой части показана общая структура профессионального стандарта. Во второй части разработана семантическая связь на примере профессионального стандарта по администрированию баз данных. В третьей части структуризация данных с использованием библиотеки Xceed.Words.NET на языке C#.

**Ключевые слова.** Образовательная программа, профессиональные стандарты, компетентный подход в образовании

### **Введение**

В современной системе образования существует острая необходимость формирования образовательных программ с компетентным подходом, отвечающих требованиям рынка труда и ожиданиям абитуриентов.

Компетентностный подход стал широко используемым в образовании, поскольку он обеспечивает глубокое и системное понимание учебных предметов, а также развитие учебных компетенций, которые необходимы для успешной карьеры.

Внедрение компетентностного подхода в систему высшего профессионального образования направлено на улучшение взаимодействия с рынком труда, повышение конкурентоспособности специалистов, обновление содержания, методологии и соответствующей среды обучения. Основная цель профессионального образования — подготовка квалифицированного специалиста соответствующего уровня и профиля, конкурентоспособного на рынке труда, компетентного, свободно владеющего своей профессией и ориентирующегося в смежных областях деятельности, готового к постоянному профессиональному росту, социальной и профессиональной мобильности (Мединцева и др., 2012). «Компетентность — это необходимая способность и качество студента в рамках академической программы, достижение которого указывает на способность или квалификацию студента в области компетенции» (Johnson и др., 2000). Компетентностный подход характеризуется практической направленностью на результат образования и предполагает формирование компетенций, которые отражаются в способности успешно справляться с определенным кругом задач (Зеер и др., 2011). Другими словами — это активная и целенаправленная деятельность по решению поставленных преподавателем задач, в ходе которой формируются знания и умения, а также компетенции, необходимые для эффективного реагирования в проблемных ситуациях как в профессиональной, так и общественной жизни, и адаптации в постоянно меняющемся мире (Пантелеева и др., 2016).

### 1. Профессиональные стандарты

Профессиональные стандарты — это неструктурированный документ, который содержит трудовые функции и описание умения и навыков, знаний необходимых для получения профессии. Они описывают основные компетенции и качества, необходимые для эффективного выполнения работы или функций. *Разработка, введение, замена и пересмотр профессиональных стандартов производятся объединениями (ассоциациями, союзами) работодателей на основе отраслевых рамок квалификаций и утверждаются Национальной палатой предпринимателей Республики Казахстан в установленном уполномоченным государственным органом по труду порядке* (<https://atameken.kz/ru/services/16>). Профессиональные стандарты содержат в себя следующие данные Таблица 1.

Таблица 1. Структура профессионального стандарта

КАРТОЧКА ПРОФЕССИИ: «НАЗВАНИЕ ПРОФЕССИИ»	
Код профессии:	
Наименование профессии:	
Уровень квалификации по ОРК:	

Трудовые функции		
Трудовая функция 1	Умения и навыки:	
	Знания:	
Требования к личностным компетенциям		
Связь с другими профессиями в рамках ОРК	Код	Название профессии
Технические характеристики профессионального стандарта		
Кем разработано		
Номер версии и год выпуска		
Дата ориентировочного пересмотра		
Актуализировано:		
Экспертиза представлена:		
Номер версии и год выпуска:		
Дата ориентировочного пересмотра:		

База данных профессиональный стандарт представляет собой набор неструктурированных .docx-файлов, который включает в себе более 30 направлений подготовки кадров и 582 профессиональных стандартов. Каждый профессиональный стандарт состоит из 8 и более трудовых функций и у каждой трудовой функций есть описание умения и навыков, знаний, необходимые для эффективного освоения направления подготовки.

## 2. Исходный набор данных

В данном исследовании используются профессиональные стандарты (ПС), разработанные совместно с представителями профессиональных сообществ, работодателей, государственных органов и образовательных учреждений, утвержденные *Национальной палатой предпринимателей Республики Казахстан в установленном уполномоченным государственным органом по труду порядке*. Из содержащих данных ПС (таблица 1) отбираем следующие данные: название профессии, трудовые функции, умения и навыки, знания. Рисунок 2.

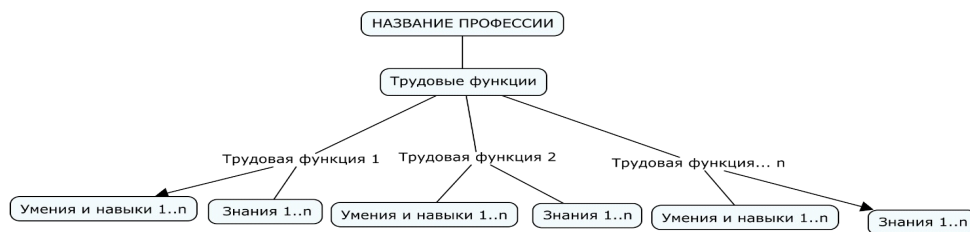


Рисунок 2. Структура ПС

На рисунке 3 показан пример семантической структуры ПС по администрированию баз данных, который содержит в себя 7 трудовых функций и каждая трудовая функция содержит описание необходимых знаний, умения и навыки.



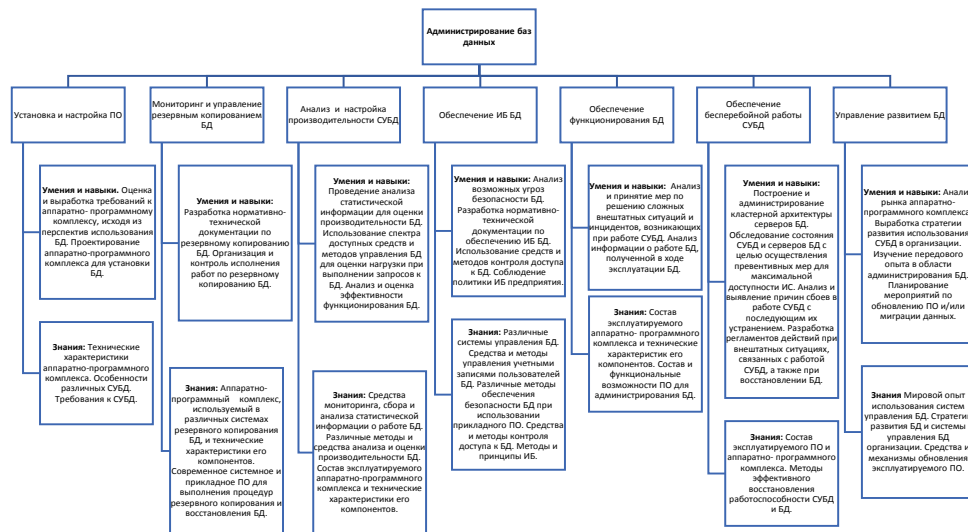


Рисунок 3. Семантическая структура ПК по администрированию баз данных

## Обзор литературы

Автор статьи (Pinto и др., 2010) дает следующее утверждение о понятии компетенция – «Компетенция — это личные способности, которые демонстрируются через измеримые знания, навыки, способности и личные качества, могут способствовать повышению эффективности работы сотрудника, его производительности и в конечном счете, успеха человека и организаций». Определение необходимых компетенций в качестве целей обучения и последующего определения концептуальных требований, является эффективным, при использовании структурированного подхода при разработке учебного плана. Также автор провел анализ результатов эффективности данной стратегии и рассматривает проект учебной программы, основанный, прежде всего на компетенциях и предлагает следующие этапы для разработки учебной программы:

- I. Определить возможности работы и связанные с ней обязанности;
- II. Выявление личностных качеств (soft-skills), необходимых для работы;
- III. Определить компетенции, необходимые для выполнения работы;
- IV. Определить знания, ноу-хау и текущие навыки, необходимые для достижения этих компетенций;
- V. Определить соответствующие предметные концепции (теорию) для приобретения и получения необходимой базы знаний;
- VI. Выявить задания (практика), позволяющие расширить "ноу-хау" по предмету;
- VII. Определить набор навыков (средств), необходимых для выполнения заданий;
- VIII. Определить текстовые и справочные материалы, предметные цели которых соответствуют знаниям, требуемым в пунктах (V), (VI) и (VII);

IX. Разработать тесты, позволяющие проверить знания кандидата на различных уровнях обучения по данной компетенции.

Автор статьи (Андропова и др., 2013) считает что, решением задачи разработки актуальных образовательных программ, возможно с использованием типовых учебных программ и куррикулумов (curricula). В своих исследованиях автор, проводит анализ и систематизирует методы диверсификации учебных программ, применяемых для формирования профилей, треков, специализаций профессиональной подготовки в международных образовательных стандартах (куррикулумах) дисциплины Компьютинг (Computing).

Автор статьи (Сухомлин и др., 2012) отметил, что для развития отечественной высшей школы, актуальным вопросом является эффективное применение зарубежного опыта подготовки ИТ-специалистов. Статья автора посвящена анализу современного состояния международных стандартов образовательных программ подготовки бакалавров и магистров в области информационных технологий: Computer Science 2001 (CS2001 или CCCS2001), Information Systems 2002 (IS2002), Computer Engineering 2004 (CE2004), Software Engineering 2004 (SE2004), Information Technology (IT 2006), а также документ Computing Curricula 2005 (CC2005)

В статье (Пантелеева и другие, 2016) рассматриваются особенности модернизации образования с использованием компетентностного подхода, который выражается в деятельностном практико-ориентированном обучении. Также, представлены компетенций применительно к старшему школьному возрасту позволяющих выявить основные задачи для успешной реализации образовательных программ на основе компетентностного подхода: формирование профессиональных и мотивационных компетенций, путем включения профильного обучения (технический профиль), установление содержательного компонента программ, а также определение эффективности реализуемых подходов за счет применения методов тестирования. В целях реализации поставленных задач, авторы обратились к опыту лицей-интерната для одаренных детей (Россия), университета Лихай (США) и научно-образовательных организаций Азии, в том числе Вьетнама. На основании описанного содержания были сделаны выводы о необходимости и возможности разработки международных обучающих программ, которые в полной мере позволяют реализовать требуемый объем компетенций.

В статье (Азарнова и др., 2018) описан алгоритмическое и программное обеспечение процедур формирования интегральной оценки и практико-ориентированной корректировки механизма свертки для иерархических оценочных компетентностных моделей выпускников образовательных направлений. Формирование оценочных иерархических компетентностных моделей осуществляется на основе компетентностного подхода, реализованного при проектировании структуры образовательных направлений. Механизм свертки по иерархическим моделям, направленный на вычисление

интегральной оценки компетентности выпускника, осуществляется через таблицы свертки, которые вычисляются на основе распределения зачетных единиц между компонентами учебного плана, покрывающими определенные компетенции. Адекватность интегральной оценки проверяется на основе реализации механизмов обратной связи с работодателями, представляющими наиболее важные сегменты рынка труда по направлению магистратуры «Бизнес-информатика», профиль «Информационная бизнес-аналитика». Стандарт направления ориентирован на формирование общекультурных и профессиональных компетенций (3 ОК, 3 ОПК и 19 ПК).

Статья (Gintciak и др., 2019) посвящена применению методов и инструментов обработки больших данных для создания профилей компетенций специалистов. В ней рассматривается подход, разработанный для формулирования ключевых компетенций, необходимых для ИТ-вакансий. Автор использует кластеризацию результатов семантического анализа и использует их для построения карты компетенций.

В следующей исследовательской работе (Majernik и др., 2021) разработана онлайн-платформа EDUportfolio, предназначенная для оптимизации учебных программ, ориентированных на достижение медицинских результатов. Программный инструмент, помогает управлять, составлять карты и анализировать учебные программы в области медицины и здравоохранения. Авторы для визуализации и анализа данных об учебных программах использовали методы обработки естественного языка. Имеются функционалы обнаружения дублирования, пробелы в учебных программах и сравнения учебных программ в разных учебных заведениях и странах.

Научная работа (Kononowicz, 2020) посвящена обзору существующей практики в области внедрения средств картирования учебных планов в медицинском образовании в рамках проекта Building Curriculum Infrastructure in Medical Education (BCIME). Собранный опыт сформировали четыре тематические категории: визуализация, текстовые описания и анализ, подход основанный на результатах, адаптивность при составлении учебных планов.

В статье (Makarova, 2020) предложен алгоритм для проектирования образовательных результатов в виде компетентностной модели выпускника с учетом требований работодателя по направлениям высшего инженерного образования.

В данном обзоре были обобщены последние технические достижения, связанные с процессом разработки образовательных программ. Мы надеемся, что полученные структурированные данные профессионального стандарта и анализ научных исследований в данной области послужит основой для проектирования модели компетенций, формулировку компетенций и получения результатов образования обеспечивающего качественную подготовку высококвалифицированных работников с набором необходимых компетенций, отвечающих требованиям работодателей и рынка труда, позволяющих эффективно осуществлять профессиональную деятельность.

### *3. Подготовка (структуризация) данных*

Формирование компетенций для проектирования основных образовательных программ высшего образования декларируется на основе профессиональных стандартов в рамках принятых национальной палате предпринимателей Республики Казакстан «Атамекен», однако трудовые функции, описанные в профессиональных стандартах, зачастую требуют структуризацию под нужды для разработки интеллектуальной системы проектирования оптимальной компетентностной модели выпускника. Мы предлагаем алгоритм структуризации профессиональных программ от формата .docx в формат.xls.

Для подготовки набора данных в дальнейшей работе необходимо экспортировать данные с Microsoft Word на Microsoft Excel. Для этой цели была использована библиотека Xceed.Words.NET на языке C#. Она предоставляет богатый набор классов и методов, которые позволяют читать данные из Word. Затем с использованием компонента Microsoft.Office.Interop.Excel из Microsoft.NET Framework, который предоставляет программный интерфейс для взаимодействия с приложениями Microsoft Excel. Это дает нам структурировать данные для дальнейшей обработки. Программный код:

```
using System;
using System.Collections.Generic;
using System.IO;
using Xceed.Words.NET;
using ExcelTest.BO;
using Microsoft.Office.Interop.Word;
namespace LabRab5
{class Program
{static void Main(string[] args)
{ string pathDocument=AppDomain.CurrentDomain.BaseDirectory+"1.docx";
  DocX docX = DocX.Load(pathDocument);
  DocX document = docX;
  //Читаем
  string text = document.Text;
  string[] arr = text.Split(new char[] { '\n', '.' });
  //Записываем
  using (ExcelTest.BO.ExcelHelper helper = new ExcelTest.BO.ExcelHelper())
  if (helper.Open(filePath: Path.Combine(Environment.CurrentDirectory,
"1.xlsx"))
      for (int i = 0; i <arr.Length; i++)
      {Console.WriteLine(arr[i]);
        helper.Set(column: "A", row: i + 1, data: i);
        helper.Set(column: "B", row: i + 1, data:arr[i]);
        helper.Save();
```

В результате получены следующие структурированные данные

ПС	Трудовая функция	Умения и навыки	Знания
Администрирование баз данных	Установка и настройка ПО	Оценка и выработка требований к аппаратно- программному комплексу, исходя из перспектив использования БД.	Технические характеристики аппаратно-программного комплекса.
Администрирование баз данных	Установка и настройка ПО	Проектирование аппаратно-программного комплекса для установки БД.	Особенности различных СУБД.
Администрирование баз данных	Установка и настройка ПО	Выбор наиболее эффективной СУБД для установки и настройки ПО.	Требования к СУБД.
Администрирование баз данных	Установка и настройка ПО	Проектирование структуры БД с учетом перспектив использования БД.	Требования к системному и прикладному ПО.
Администрирование баз данных	Установка и настройка ПО	Осуществление эффективной настройки аппаратно-программного комплекса.	Механизмы управления ресурсами аппаратно-программного комплекса.
Администрирование баз данных	Установка и настройка ПО	Использование технической документации по установке и настройке ПО.	Архитектура ИС использующих БД.
Администрирование баз данных	Установка и настройка ПО	-	Проектирование БД.
Администрирование баз данных	Установка и настройка ПО	-	Методы и принципы ИБ.
Администрирование баз данных	Обеспечение функционирования БД	Анализ и принятие мер по решению сложных внештатных ситуаций и инцидентов, возникающих при работе СУБД.	Состав эксплуатируемого аппаратно- программного комплекса и технические характеристики его компонентов.
Администрирование баз данных	Обеспечение функционирования БД	Анализ информации о работе БД, полученной в ходе эксплуатации БД.	Состав и функциональные возможности ПО для администрирования БД.
Администрирование баз данных	Обеспечение функционирования БД	Координация работ по администрированию БД.	Методы мониторинга функционирования БД.
Администрирование баз данных	Обеспечение функционирования БД	Разработка нормативно-технической документации по функционированию БД.	Методы анализа информации.
Администрирование баз данных	Обеспечение функционирования БД	Анализ необходимости модернизации аппаратно-программного комплекса на основе результатов эксплуатации БД.	Основы управления рисками.
Администрирование баз данных	Обеспечение функционирования БД	Прогноз и оценка рисков сбоев в работе БД.	-

## **Заключение**

Рынок труда меняется динамично в связи с высокой скоростью развития ИТ-технологий. Появляются новые сферы профессиональной деятельности, а вместе с ними новые профессии и профессиональные стандарты. Необходимо совершенствовать систему образования, которая должна обеспечивать цифровую экономику компетентными кадрами. В данной обзорной статье рассматривались последние технические достижения, связанные с процессом составления образовательных программ. Мы надеемся, что полученные структурированные данные профессионального стандарта послужат основой для разработки будущих интегрированных, комплексных и доступных образовательных программ.

## **ЛИТЕРАТУРА**

<https://atameken.kz/ru/services/16>

Азарнова Т.В. и др. Разработка практико-ориентированного алгоритма формирования интегральной оценки компетентности выпускников образовательных направлений //Вестник ВГУ. Серия: Системный анализ и информационные технологии. – 2018. – №. 3. – С. 119–131.

Андропова Е.В., Сухомлин В.А. Диверсификация программ профессиональной подготовки в международных образовательных стандартах в области информационных технологий // Вестник Московского университета. Серия 20. Педагогическое образование. – 2013. – №. 1. – С. 73–86.

Гинтсиак А. и др. Разработка профилей компетенций ИТ-специалистов на основе семантического анализа вакансий // Журнал «Физика»: Conference Series. - IOP Publishing, 2019. - Т. 1405. - №. 1. - С. 012003.

Джонсон Д.В. Оценка эффективности образовательных программ на основе компетенций // Журнал «Образование в области информационных систем». - 2000. - Т. 11. - №. 3. - С. 127–132.

Зеер Э.Ф. Практика формирования компетенций: методологический аспект Формирование компетенций в практике преподавания общих и специальных дисциплин в учреждениях среднего профессионального образования / Э.Ф. Зеер, Д.П. Заводчиков // Сборник статей по материалам Всероссийской научно-практической конференции. – Екатеринбург; Березовский, 2011. – С. 5–10.

Кононович А.А. и др. Техническая инфраструктура для составления карт учебных планов в медицинском образовании: обзор //Биоалгоритмы и мед. системы. - 2020. - Т. 16. - №. 2. - С. 20200026.

Маджерник Дж. и др. Разработка и внедрение онлайн-платформы для составления карт учебных планов в медицинском образовании //Биоалгоритмы и мед. системы. - 2021. - Т. 18. - №. 1. - С. 1–11.

Макарова С., Мартынов В., Зайцева А. Проблемы и методы формирования образовательных результатов при реализации инженерного образования в условиях цифровой экономики //2020 V Международная конференция по информационным технологиям в инженерном образовании (Инфорино). - IEEE, 2020. - С. 1–5.

Мединцева И.П. Компетентностный подход в образовании //Педагогическое мастерство: материалы II Междунар. науч. конф. (г. Москва, декабрь 2012 г.).

Пантелеева М.В., Сухристина А.С. Компетентностный подход в образовании: российский и зарубежный опыт //Вестник Южно-Уральского государственного университета. Серия: Образование. Педагогические науки. – 2016. – Т. 8. – №. 4. – С. 100–104.

Пинто Ю. Стратегия, реализация и результаты гибкой учебной программы, основанной на компетенциях //Acta Inroads. - 2010. - Т. 1. - №. 2. - С. 54–61.

Сухомлин В.А. Международные образовательные стандарты в области информационных технологий //Прикладная информатика. – 2012. – №. 1 (37). – С. 33–54.

## REFERENCES

Andropova E.V., Sukhomlin V.A. Diversification of professional training programs in international educational standards in the field of information technology // Vestnik of Moscow University. Series 20. Pedagogical Education. - 2013. - №. 1. - C. 73–86.

Azarnova T.V. et al. Development of practice-oriented algorithm for the formation of integral competence assessment of graduates of educational areas // Vestnik VSU. Series: System Analysis and Information Technology. - 2018. - №. 3. - C. 119–131.

Gintciak A. et al. Developing competency profiles of IT specialists based on semantic analysis of vacancies // Journal of Physics: Conference Series. - IOP Publishing, 2019. - T. 1405. - №. 1. - C. 012003. (in Eng.).

<https://atameken.kz/ru/services/16>

Johnson D.W. Competency-based educational program effectiveness assessment // Journal of Information Systems Education. - 2000. - T. 11. - №. 3. - C. 127–132. (in Eng.).

Kononowicz A.A. et al. Technical infrastructure for curriculum mapping in medical education: a narrative review // Bio-Algorithms and Med-Systems. - 2020. - T. 16. - №. 2. - C. 20200026. (in Eng.).

Majernik J. et al. Development and implementation of an online platform for curriculum mapping in medical education // Bio-Algorithms and Med-Systems. - 2021. - T. 18. - №. 1. - C. 1–11. (in Eng.).

Makarova S., Martynov V., Zaitseva A. Problems and Methods for Forming Educational Results in the Implementation of Engineering Education in the Digital Economy // 2020 V International Conference on Information Technologies in Engineering Education (Inforino). - IEEE, 2020. - C. 1–5. (in Eng.).

Medintseva I.P. Competence approach in education // Pedagogical mastery : proceedings of the II International scientific conference (Moscow, December 2012).

Panteleeva M.V., Sukhrstina A.C. Competence approach in education: Russian and foreign experience // Vestnik of South Ural State University. Series: Education. Pedagogical Sciences. - 2016. - T. 8. - №. 4. - C. 100–104.

Pinto Y.A strategy, implementation and results of a flexible competency based curriculum // Acm Inroads. - 2010. - T. 1. - №. 2. - C. 54–61. (in Eng.).

Sukhomlin V.A. International educational standards in the field of information technologies // Applied Informatics. - 2012. - №. 1 (37). - C. 33–54.

Zeer E.F. Zeer D.P. Zavodchikov // Collection of articles on the materials of the All-Russian scientific-practical conference. - Yekaterinburg; Berezovsky, 2011. - C. 5–10.

NEWS OF THE NATIONAL ACADEMY OF SCIENCES OF THE REPUBLIC OF  
KAZAKHSTAN  
PHYSICO-MATHEMATICAL SERIES  
ISSN 1991-346X  
Volume 3. Number 347 (2023). 112–130  
<https://doi.org/10.32014/2023.2518-1726.208>

MPHTИ 81.96.00

© **Sh. Mussiraliyeva, M. Bolatbek, M. Sagynay\*, Zh. Yeltay,  
K. Bagitova, 2023**

Al-Farabi Kazakh National University, specialty «Information Security Systems»,  
Kazakhstan, Almaty.  
E-mail: [sagynaymoldir11@gmail.com](mailto:sagynaymoldir11@gmail.com)

### **THE CONCEPT OF EXTREMIST DATA AND A SYSTEMATIC REVIEW OF ANTI-EXTREMISM PROJECTS**

**Mussiraliyeva Shynar** — Candidate of Physical and Mathematical Sciences, Head of the Department Information Systems of Al-Farabi Kazakh National University, 050040, Almaty, 71 Al-Farabi Avenue  
E-mail: [mussiraliyevash@gmail.com](mailto:mussiraliyevash@gmail.com); <http://orcid.org/0000-0001-5794-3649>;

**Bolatbek Milana** — PhD., senior lecturer of the Department of Information Systems of Al-Farabi Kazakh National University, 050040, Almaty, 71 Al-Farabi Avenue  
E-mail: [bolatbek.milana@gmail.com](mailto:bolatbek.milana@gmail.com); <http://orcid.org/0000-0002-2153-180X>;

**Sagynay Moldir** — lecturer of the Department of Information Systems of Al-Farabi Kazakh National University, 050040, Almaty, 71 Al-Farabi Avenue  
E-mail: [sagynaymoldir11@gmail.com](mailto:sagynaymoldir11@gmail.com); <http://orcid.org/0009-0004-1377-5742>;

**Yeltay Zhastay** — Master of Information Security Systems, 1st year doctoral student of the specialty "Information Security Systems" of Al-Farabi Kazakh National University, 050040, Almaty, 71 Al-Farabi Avenue

E-mail: [jastayeltay@gmail.com](mailto:jastayeltay@gmail.com); <https://orcid.org/0000-0002-9275-7582>;

**Kalamkas Bagitova** — lecturer of the Department of Information Systems of Al-Farabi Kazakh National University, 050040, Almaty, 71 Al-Farabi Avenue  
E-mail: [kbbagitova@gmail.com](mailto:kbbagitova@gmail.com); <https://orcid.org/0000-0003-1587-1995>.

**Abstract.** The problem of countering the spread of aggressive information on the global Internet is relevant both for society and for government agencies. This paper discusses a method to combat this problem by filtering unwanted resources on the Internet. In the modern world, text data makes up the bulk of content on the Internet. However, terrorist and extremist groups actively use web technologies to disseminate information and propaganda. This poses a threat to national security and requires action at both the global and national levels. The problem of the spread of extremist materials has a global character and is of particular relevance for the Republic of Kazakhstan. Using the Internet, ideologists of extremist groups actively influence the consciousness of young people and the population as a whole. In modern conditions, extremists are able to use the capabilities of the Internet to



organize and carry out criminal actions, including the use of network resources for planning and financing extremist activities. This problem is assessed as a threat to national security. The fight against extremism has become one of the priorities of law enforcement agencies, and it requires constant monitoring, analysis and forecasting of the actions of extremist groups on the Internet. Thus, this work is an attempt to characterize and analyze the complex problem of combating extremist activity on the Internet, especially in the context of the Republic of Kazakhstan. Further research and measures are being taken to identify, prevent and suppress extremist manifestations in the network. The study emphasizes that the fight against extremism and aggressive information on the Internet requires a comprehensive approach and the use of effective measures. The phenomenon of extremism is developing dynamically, and its fight requires continuous monitoring, threat identification and analysis of behavior on the web. In the context of the Republic of Kazakhstan, the fight against extremism has become one of the priorities of law enforcement agencies, and this challenge requires further research and the implementation of timely measures to identify, prevent and suppress extremist manifestations on the Internet.

**Keywords:** extremism, global network, global extremist resources, propaganda, global terrorism, violence, RAND, Trace

© Ш.Ж. Мусиралиева, М.А. Болатбек, М. Сағынай\*, Ж.Ы. Елтай,  
К.Б. Багитова, 2023

Әл-Фараби атындағы Қазақ ұлттық университеті, «Ақпараттық қауіпсіздік жүйесі» мамандығы, Қазақстан, Алматы.

E-mail: sagynaymoldir11@gmail.com

## ЭКСТРЕМИСТИК МӘЛІМЕТТЕР ТҮСІНІГІ ЖӘНЕ ЭКСТРЕМИЗМГЕ ҚАРСЫ КҮРЕС ЖОБАЛАРЫНА ЖҮЙЕЛІК ШОЛУ

**Мусиралиева Шынар** — физика-математика ғылымдарының кандидаты, «Ақпараттық жүйелер» кафедрасының меңгерушісі, Әл-Фараби атындағы Қазақ ұлттық университеті, 050040, Алматы, Әл-Фараби даңғылы, 71

E-mail: mussiraliyevash@gmail.com; <http://orcid.org/0000-0001-5794-3649>;

**Болатбек Милана** — Әл-Фараби атындағы Қазақ ұлттық университеті «Ақпараттық жүйелер» кафедрасының PhD., аға оқытушысы, 050040, Алматы, Әл-Фараби даңғылы, 71

E-mail: bolatbek.milana@gmail.com; <http://orcid.org/0000-0002-2153-180X>;

**Сағынай Мөлдір** — Әл-Фараби атындағы Қазақ ұлттық университеті «Ақпараттық жүйелер» кафедрасының оқытушысы, 050040, Алматы, Әл-Фараби даңғылы, 71

E-mail: sagynaymoldir11@gmail.com; <http://orcid.org/0009-0004-1377-5742>;

**Елтай Жастай** — Әл-Фараби атындағы Қазақ ұлттық университеті Ақпараттық қауіпсіздік жүйелері мамандығының магистрі, Ақпараттық қауіпсіздік жүйелері мамандығының 1-ші курс докторанты, 050040, Алматы, Әл-Фараби даңғылы, 71

E-mail: jastayeltay@gmail.com; <https://orcid.org/0000-0002-9275-7582> ;

**Багитова Каламкас Багитовна** — Әл-Фараби атындағы Қазақ ұлттық университеті «Ақпараттық жүйелер» кафедрасының оқытушысы, 050040, Алматы, Әл-Фараби даңғылы, 71

E-mail: kbbagitova@gmail.com; <https://orcid.org/0000-0003-1587-1995>.

**Аннотация.** Ғаламдық Интернет желісінде агрессивті ақпараттың таралуына қарсы тұру мәселесі қоғам үшін де, мемлекеттік органдар үшін де өзекті болып табылады. Бұл жұмыс интернеттегі қажетсіз ресурстарды сүзу арқылы осы мәселемен күресу әдісін қарастырады. Қазіргі әлемде мәтіндік деректер Интернеттегі мазмұнның негізгі бөлігін құрайды. Алайда террористік және экстремистік топтар ақпарат тарату және насихаттау үшін веб-технологияларды белсенді қолданады. Бұл ұлттық қауіпсіздікке қауіп төндіреді және жаһандық және ұлттық деңгейде әрекет етуді талап етеді. Экстремистік материалдарды тарату проблемасы әлемдік сипатқа ие және Қазақстан Республикасы үшін ерекше өзектілікке ие. Интернетті қолдана отырып, экстремистік топтардың идеологтары жастар мен жалпы халықтың санасына белсенді әсер етеді. Қазіргі жағдайда экстремистер экстремистік қызметті жоспарлау және қаржыландыру үшін желі ресурстарын пайдалануды қоса алғанда, қылмыстық әрекеттерді ұйымдастыру және жүргізу үшін Интернет мүмкіндіктерін шебер пайдаланады. Бұл мәселе ұлттық қауіпсіздікке қауіп ретінде бағаланады. Экстремизмге қарсы күрес құқық қорғау органдарының басымдықтарының біріне айналды және ол Интернет желісіндегі экстремистік топтардың іс-әрекеттерін үнемі бақылауды, талдауды және болжауды талап етеді. Осылайша, бұл жұмыс Интернеттегі, әсіресе Қазақстан Республикасының контекстіндегі Экстремистік қызметке қарсы күрестің күрделі мәселесін сипаттауға және талдауға тырысады. Желідегі экстремистік көріністерді анықтау, алдын алу және жолын кесу мақсатында одан әрі зерттеулер мен шаралар қабылдануда. Зерттеу интернеттегі экстремизм мен агрессивті ақпаратпен күресу кешенді тәсілді және тиімді шараларды қолдануды қажет ететіндігін атап көрсетеді. Экстремизм құбылысы қарқынды дамып келеді және оның күресі үздіксіз бақылауды, қауіптерді анықтауды және желідегі мінез-құлықты талдауды қажет етеді. Қазақстан Республикасының контекстінде экстремизмге қарсы күрес құқық қорғау органдарының басым міндеттерінің біріне айналды және бұл интернет желісіндегі экстремистік көріністерді анықтау, алдын алу және жолын кесу үшін одан әрі зерттеуді және уақтылы шараларды іске асыруды талап етеді.

**Түйін сөздер:** экстремизм, жаһандық желі, жаһандық экстремистік ресурстар, үгіт-насихат, жаһандық терроризм, зорлық-зомбылық, RAND, Trase

*Берілген зерттеу Қазақстан Республикасы Ғылым және жоғары білім министрлігінің Ғылым комитеті қаржыландыратын “Қазақ тіліндегі кибер экстремизмнің идеологиялық бағыттарын жасанды интеллект әдістері көмегімен мультиклассификациялау” жобасы аясында орындалды (грант AP19676342, жоба жетекшісі Мусиралиева Ш.Ж.).*

© Ш.Ж.Мусиралиева, М.А.Болатбек, М.Сағынай\*, Ж.Ы. Елтай,  
К.Б. Багитова, 2023

Казахский национальный университет имени аль-Фараби, специальность  
«Системы информационной безопасности», Казахстан, Алматы.

E-mail: sagynaymoldir11@gmail.com

## ПОНЯТИЕ ЭКСТРЕМИСТСКИХ ДАННЫХ И СИСТЕМНЫЙ ОБЗОР ПРОЕКТОВ ПО БОРЬБЕ С ЭКСТРЕМИЗМОМ

**Мусиралиева Шынар** — кандидат – физико-математических наук, заведующая кафедры «Информационные системы» Казахского национального университета им. Аль-Фараби, 050040, Алматы, проспект Аль-Фараби, 71.

E-mail: mussiraliyevash@gmail.com; <http://orcid.org/0000-0001-5794-3649>;

**Болатбек Милана** — PhD., старший преподаватель кафедры «Информационные системы» Казахского национального университета имени Аль-Фараби, 050040, Алматы, проспект Аль-Фараби, 71.

E-mail: bolatbek.milana@gmail.com; <http://orcid.org/0000-0002-2153-180X>;

**Сағынай Мөлдiр** — преподаватель кафедры «Информационные системы» Казахского национального университета имени Аль-Фараби, 050040, Алматы, проспект Аль-Фараби, 71.

E-mail: sagynaymoldir11@gmail.com; <http://orcid.org/0009-0004-1377-5742>;

**Елтай Жастай** — Магистр систем информационной безопасности, Докторант 1 курса специальности «Системы информационной безопасности» Казахского национального университета имени Аль-Фараби, 050040, Алматы, проспект Аль-Фараби, 71.

E-mail: jastayeltay@gmail.com ; <https://orcid.org/0000-0002-9275-7582>;

**Багитова Каламкас Багитовна** — преподаватель кафедры «Информационные системы» Казахского национального университета имени Аль-Фараби, 050040, Алматы, проспект Аль-Фараби, 71.

E-mail: kbbagitova@gmail.com; <https://orcid.org/0000-0003-1587-1995>.

**Аннотация.** Проблема противодействия распространению агрессивной информации в глобальной сети Интернет является актуальной как для общества, так и для государственных органов. В данной работе рассматривается метод борьбы с этой проблемой путем фильтрации нежелательных ресурсов в Интернете. В современном мире текстовые данные составляют основную часть контента в Интернете. Однако террористические и экстремистские группы активно используют веб-технологии для распространения информации и пропаганды. Это создает угрозу национальной безопасности и требует действий как на глобальном, так и на национальном уровне. Проблема распространения экстремистских материалов имеет мировой характер и особую актуальность для Республики Казахстан. Используя Интернет, идеологи экстремистских групп активно воздействуют на сознание молодежи и населения в целом. В современных условиях экстремисты с умением используют возможности Интернета для организации и проведения преступных действий, включая использование ресурсов сети для планирования и финансирования экстремистской деятельности. Эта проблема оценивается как угроза национальной безопасности. Борьба с экстремизмом стала одним из приоритетов правоохранительных органов, и она требует постоянного

мониторинга, анализа и прогнозирования действий экстремистских групп в сети Интернет. Таким образом, данная работа представляет собой попытку охарактеризовать и проанализировать сложную проблему борьбы с экстремистской деятельностью в Интернете, особенно в контексте Республики Казахстан. Дальнейшие исследования и меры предпринимаются с целью выявления, предотвращения и пресечения экстремистских проявлений в сети. Исследование подчеркивает, что борьба с экстремизмом и агрессивной информацией в интернете требует комплексного подхода и применения эффективных мер. Феномен экстремизма динамично развивается, и его борьба требует непрерывного мониторинга, выявления угроз и анализа поведения в сети. В контексте Республики Казахстан, борьба с экстремизмом стала одной из приоритетных задач правоохранительных органов, и этот вызов требует дальнейших исследований и реализации своевременных мер для выявления, предотвращения и пресечения экстремистских проявлений в сети интернет.

**Ключевые слова:** экстремизм, глобальная сеть, глобальные экстремистские ресурсы, пропаганда, глобальный терроризм, насилие, RAND, Trace

### **Кіріспе**

Жаһандық желіде агрессивті ақпараттың таралуына қарсы тұру қоғам мен мемлекеттік органдардың өзекті мәселесі болып табылады, аталған тапсырма интернеттің қажетсіз ресурстарын сүзу арқылы шешіледі.

XXI ғасыр технологиялары интернеттегі ақпаратты пайдалануды кеңейткенімен, мәтіндік деректер интернеттегі мазмұнның ең көп таралған түрі болып табылады. Алайда, террористік және экстремистік топтар ақпарат тарату, насихаттау, қаражат жинау және экстремистік миссияларын қоса алғанда, әртүрлі функцияларды орындау үшін веб-технологияларды ұтымды пайдалануда. Мұндай жағдайда интернет ұлттық қауіпсіздікке қауіп-қатер төндіреді.

Интернет экстремистік материалдарды орналастыру үшін белсенді қолданылып келеді. Проблема жалпы әлемдік сипатқа ие және әлемдік саяси процестің басты қатысушыларының бірі ретінде Қазақстан Республикасы үшін өте өзекті. Интернеттің ғаламдық желісін және компьютерлік байланыс мүмкіндіктерін қолдана отырып, экстремистік қозғалыстар мен топтардың идеологтары азаматтардың, ең алдымен жастардың санасына белсенді әсер етеді.

Экстремистер бүгінде қылмыстық іс-әрекетте Интернет желісінің шын мәнінде шексіз мүмкіндіктерін белсенді пайдаланады, оның ішінде: қылмыстарды дайындау және жасау кезінде Интернет желісінің ресурстарын пайдалану; әлеуметтік қауіпті әрекеттерді басқару және ұйымдастыру мақсатында ақпаратпен жасырын алмасу; экстремистік қызметті қаржыландыру үшін иесіздендірілген қаржылық желілік құралдарды қолдану, мысалы Darkcoin төлем жүйелері; арнайы құрылған сайттарда және басқа интернет-ресурстарда белсенді насихаттау үшін жоспарланған ақпараттық

операцияларды жүзеге асыру және т.б. нәтижесінде соңғы жылдары экстремизм проблемасы шиеленісе түсуде, ол қазіргі уақытта жалпы мемлекеттік маңызы бар проблема және Қазақстанның ұлттық қауіпсіздігіне қатер ретінде қаралуда. Осылайша, интернет желісін пайдалана отырып, экстремистік көріністерге қарсы іс-қимыл саласындағы ахуал күрделі болып қалуда, бұл, атап айтқанда, экстремизмнің кез келген көріністерін анықтауға, олардың алдын алуға және жолын кесуге бағытталған ғылыми зерттеулерді жүзеге асыру және тиімді әрі уақтылы шаралар кешенін іске асыру қажеттілігін негіздейді.

Экстремизм құбылысы өте серпінді дамып, күн сайын жаңа белгілер мен сипаттамаларға ие болуда. Қазіргі уақытта үгіт-насихат пен экстремистік идеологияның әсерінен террористердің, сондай-ақ желілік және әлсіз байланысқан құрылымы бар ұйымдасқан террористік қауымдастықтардың, террористік шабуылдарының саны артып келеді. Ақпарат алмасудың және осындай құрылымдарды жылжытудың негізгі құралы – интернет, атап айтқанда веб-ресурстар, әлеуметтік желілер және электрондық пошта. Осыған байланысты интернет желісінде террористік және экстремистік ақпаратты генерациялайтын және тарататын жекелеген пайдаланушылардан, топтардан және желілік қоғамдастықтардан туындайтын қатерлерді анықтау, қарым-қатынас тақырыптарын, байланыстарды айқындау, сондай-ақ мінез-құлық мониторингі және болжау міндеті туындайды.

Қазіргі уақытта экстремистік әрекетке қарсы күрес Қазақстан Республикасының аумағында да, одан тыс жерлерде де болып жатқан оқиғалармен анықталған құқық қорғау органдарының басым міндеттерінің бірі болып табылады.

Экстремизм — бұл:

- жеке және (немесе) заңды тұлғаның, белгіленген тәртіппен экстремистік деп танылған ұйымдар атынан жеке және (немесе) заңды тұлғалар бірлестігінің әрекеттер жасауы;

- жеке және (немесе) заңды тұлғаның, жеке және (немесе) заңды тұлғалар бірлестігінің мынадай экстремистік мақсаттарды көздейтін әрекеттер жасауы: Қазақстан Республикасының конституциялық құрылысын күштеп өзгерту, егемендігін, оның аумағының тұтастығын, қол сұғылмауын және бөлінбеуін бұзу, мемлекеттің ұлттық қауіпсіздігі мен қорғаныс қабілетіне нұқсан келтіру, билікті күшпен басып алу немесе билікті күшпен ұстап тұру, заңсыз әскерилендірілген құралым құру, оған басшылық ету және қатысу, қарулы бүлік ұйымдастыру және оған қатысу, әлеуметтік, тектік-топтық алауыздықты қоздыру (саяси экстремизм);

- нәсілдік, ұлттық және рулық алауыздықты, оның ішінде зорлық-зомбылықпен немесе зорлық-зомбылыққа шақырумен байланысты алауыздықты қоздыру (ұлттық экстремизм);

- діни өшпенділікті немесе алауыздықты, оның ішінде зорлық-зомбылықпен немесе зорлық-зомбылыққа шақырумен байланысты өшпенділікті немесе алауыздықты қоздыру, сондай-ақ азаматтардың қауіпсіздігіне, өміріне,

денсаулығына, имандылығына немесе құқықтары мен бостандықтарына қатер төндіретін кез келген діни практиканы қолдану (діни экстремизм) (Қазақстан Республикасының №31 Заңы; Болатбек, 2020).

Экстремистік көріністердің қолдану саласының сипаты бойынша келесі жіктелуін бөліп көрсетуге болады: саяси сипатта; экономикалық сипатта; діни сипатта және психологиялық сипатта. Кез-келген экстремистік қозғалыс өзара байланысты және әр нақты жағдайда күшті немесе әлсіз көрінетін әртүрлі элементтерді қамтиды. Экстремизмнің негізгі түрлері кесте 1.1-де келтірілген:

Кесте 1.1 – Экстремизмнің негізгі түрлері

Экстремизм түрі	Экстремизмнің берілген түрінің негізгі түсінігі
1	2
Сепаратизм	Мемлекеттің бір бөлігін бөліп алып, оны жаңа тәуелсіз мемлекетке немесе автономды бөлікке айналдыруға ұмтылу
Ксенофобия	Басқа мәдениет, ұлт, мемлекет өкілдеріне деген төзімсіздік
Ұлтшылдық	Белгілі бір ұлттың артық екендігін айтуға негізделетін идеология, саяси көзқарастар жүйесі
Шовинизм	Басқа ұлт өкілдерін кемсіту, эксплуатациялау және дискриминациялау мақсатында қандай да бір басқа ұлттық артықшылығын насихаттайтын идеология
Расизм/ нәсілшілдік	Әр түрлі нәсілдердің физикалық және психикалық кемелденбегенін насихаттайтын идеология. Халықты «жоғарыдағылар» және «төмендегілер», «кемелденгендер» мен «кемелденбегендер» деген сияқты топтарға жіктеу, нәсілдік тиістілікке байланысты нәсілдік дискриминация, ұлт геноциді үшін пайдаланылады.
Фашизм	Әскери расизм, «басқа» ұлттық және әлеуметтік топтарға ксенофобиядан басталып, геноцидке, мистикалық дұшпандыққа, тоталитарлық мемлекетке табынуға ауысатын әлеуметтік-саяси қозғалыстардың жалпы атауы.
Терроризм	Тек зорлық-зомбылық құралдарын қолдану арқылы орындалатын саясат.

Экстремизмді бағыты бойынша келесідей топтарға жіктеп көрсетуге болады:

- экономикалық (бір ғана меншік түрін орнату, бәсекелестікті болдырмау және т.б.);
  - рухани (басқа мәдениет өкілдерінің жетістіктерін теріске шығару);
  - экологиялық (табиғатты қорғау саясатына қарсы шығу);
  - діни (басқа конфиссия өкілдеріне деген жеккөрушілік);
  - ұлттық (басқа ұлттардың қызығушылықтары мен құқықтарын теріске шығару);
  - саяси (үкіметтік құрылымдар, мемлекеттік қызметтерге қарсы шығу).
- Іс-әрекеттердің масштабына байланысты:
- мемлекетшілік (өз ұлтына репрессия жасау);
  - мемлекетаралық (өз нормалары мен принциптерін әлемдік масштабға орнатуға тырысу).

Өкілетті құрылымдарға байланысты:

- мемлекеттік (репрессияның өкілетті құрылымдары арқылы орындалады);
- мемлекетке оппозициялық (антирежимдік топтар; теракттар).

*Әдебиеттерге шолу*

Жаһандық терроризм дерекқоры (GTD)

Жаһандық терроризм дерекқоры (GTD) — 1970 жылдан 2019 жылға дейін бүкіл әлем бойынша ішкі және халықаралық террористік шабуылдар туралы ақпаратты қамтитын ашық алғашқы дерекқор қоры және қазіргі уақытта 200 000-нан астам жағдайды қамтиды. Әр оқиға үшін оқиғаның күні мен орны, пайдаланылған қару мен мақсаттың сипаты, құрбан болғандар саны және сәйкестендіру кезінде анықталған топ немесе оған жауапты жеке тұлға туралы ақпаратты қамтиды. Терроризмді зерттеу және терроризмге қарсы іс-қимыл жөніндегі ұлттық консорциум (START) GTD-ді осы онлайн-интерфейс арқылы қол жетімді етеді, террористік зорлық-зомбылық туралы түсінікті кеңейтуге тырысады.

GTD сипаттамалары:

200 000-нан астам террористік шабуылдар туралы ақпарат бар;

Қазіргі уақытта бұл әлемдегі террористік шабуылдар туралы ең толық құпия емес мәліметтер қоры;

1970 жылдан бастап 95000-нан астам жарылыс, 20000 кісі өлтіру және 15000 адам ұрлау және барымтаға алу туралы ақпаратты қамтиды;

Әрбір жағдай үшін кем дегенде 45 айнымалы туралы ақпаратты қамтиды, ал кейінгі оқиғалар 120-дан астам айнымалы туралы ақпаратты қамтиды;

Тек 1998 жылдан 2019 жылға дейін 4 000 000-нан астам жаңалықтар мақалалары мен 25000 жаңалықтар көздері оқыс оқиғалар туралы мәліметтер жинау үшін талданды (1-сурет).

The screenshot shows the website for the Global Terrorism Database (GTD). At the top, there is a navigation bar with the START logo and menu items: Research, Education, Training, Publications, Data & Tools, News & Media, and About. Below the navigation bar is a tagline: "A consortium of researchers dedicated to improving the understanding of the human causes and consequences of terrorism".

The main content area is titled "Global Terrorism Database (GTD)". It contains the following information:

- Description:** The Global Terrorism Database (GTD) is an open-source database including information on terrorist events around the world since 1970 (currently updated through 2018). Unlike many other event databases, the GTD includes systematic data on international as well as domestic terrorist incidents that have occurred during this time period and now includes over 190,000 cases. For each GTD incident, information is available on the date and location of the incident, the weapons used and nature of the target, the number of casualties, and – when identifiable – the identity of the perpetrator.
- Link:** <https://www.start.umd.edu/gtd/>
- START Investigators:** Erin Miller, Gary LaFrance, Laura Dugan
- Dataset:** 1970
- Abbreviation:** GTD
- KEYWORDS:**
  - Projects:** Global Terrorism Database (GTD)
  - Research Area:** Terrorism and Society, Crime and Terrorism
  - Regions:** Global
  - Topic:** Global Security, Terrorism Databases and Resources, Terrorism Trends

On the right side of the page, there is a "Recent Publications" section with a "BACKGROUND REPORT" titled "Global Terrorism Overview: Terrorism in 2019". Below it, there are links to "BOOK: The Psychology of Extremism: A Monocausal Perspective" and "BOOK CHAPTER: Masters of Both: Balancing the Extremes of Innovation Through Tight-Loose Ambidexterity". At the bottom of the page, there is a small text in Russian: "Активизируйте уведомления, перейдите в раздел 'Параметры'".

1-сурет – Жаһандық терроризм дерекқоры (GTD)

Терроризм туралы ғаламдық мәліметтер базасы 2001 жылы доктор Гари Лафри Мэриленд университетіне Пинкертонның Ғаламдық барлау қызметінен (PGIS) қолжазба мұрағатын беруді үйлестірген кезде басталды. 1970 жылдан 1997 жылға дейін PGIS зерттеушілерді негізінен отставкадағы АҚШ әскери күштерін терроризм қаупін бағалау үшін телеграф қызметтерінен, үкіметтік есептерден және ірі халықаралық газеттерден террористік шабуылдар туралы ақпаратты анықтауға және жазуға үйретті. 2005 жылдың желтоқсанына қарай зерттеу тобы түзетулер енгізіп, дерекқорға қосымша ақпарат қосуды аяқтады. 2006 жылдың сәуірінде жанадан құрылған терроризмді зерттеу және терроризмге қарсы ұлттық консорциум (START), терроризмді зерттеу және барлау орталығымен (CETIS) бірлесе отырып, 1997 жылдан кейін GTD кеңейту процесін бастады. Бұған жиналған мәліметтер түрлерін кеңейту және rgis-тегі терроризм анықтамасын зерттеу тобы 1970 жылға дейін барлық оқиғаларға ретроактивті түрде қолданған жеке енгізу критерийлерінің жиынтығы ретінде талдау кірді. CETIS командасы 1998 жылдың қаңтарынан 2008 жылдың наурызына дейін болған шабуылдар туралы ақпаратты жазды. Олар сонымен бірге 1993 жылы жетіспейтін деректерді қайта жинауға тырысты. Өкінішке орай, бұл әрекет сәтсіз болды, ішінара сол кезеңнің бастапқы құжаттары жеткіліксіз сақталғанына байланысты. 2008 жылдың сәуірінде Нью-Хейвен университетінің зорлық-зомбылық топтарын зерттеу институтының (ISVG) сарапшылары GTD-ге қосылу үшін мәліметтер жинауды бастады. ISVG-дің 2012 жылдың наурызына дейін жалғасқан мәліметтер жинау әрекеті 2008 жылдың сәуірінен 2011 жылдың қазанына дейін болған террористік шабуылдар туралы мәліметтерді қамтыды. GTD зерттеушілері басында бұл деректерді GTD-ге біріктірді және GTD мүмкіндігінше толық және дәл болуын қамтамасыз ету үшін 1970 жылға дейін бұрын анықталған жағдайлар туралы қосымша істер мен қосымша ақпаратты анықтау үшін көптеген көздерді үнемі қарап отыруды жалғастырды. Аталған процесс GTD құрамына кіретін оқиғаларды анықтау және құжаттау үшін әлемнің әр түрлі бұқаралық ақпарат құралдарынан басталады – күніне екі миллионнан астам мақала қарастырылған. Табиғи тілді өңдеу, атаулы нысандарды алу және машиналық оқыту модельдері террористік шабуылдар туралы ақпаратты қамтитын жаңалықтар мақалаларын анықтауға және ұйымдастыруға көмектеседі. GTD тобы аналитиктерге бірегей шабуылдарды анықтауға, әр оқиғаның егжей-тегжейін жазуға және жаңа ақпарат түскен кезде бұрын тіркелген оқиғалар туралы жазбаларды жаңартуға мүмкіндік беретін жеке деректерді басқару жүйесін жасады. Технологияны жетілдіру және интернетті кеңейту бастапқы материалдардың қол жетімділігін де, жұмыс процестерінің тиімділігін де арттырды [3].

RAND Database of Worldwide Terrorism Incidents

RAND Database of Worldwide Terrorism Incidents бүкіл әлемдегі терроризмге қатысты оқиғалар туралы RAND дерекқоры (RDWTI) — бұл 1968 жылдан 2009 жылға дейінгі мәліметтер жиынтығы, 40 жыл ішінде RAND корпорациясы терроризм мен терроризмге қарсы зерттеулердің



алдыңғы қатарында болды. Осы жұмысты қолдау үшін RAND 1968 жылдан бастап халықаралық және ішкі терроризм туралы жан-жақты ақпарат беретін террористік оқиғалар туралы мәліметтер базасын жасап шығарды. Көптеген жылдар бойы көптеген мемлекеттік және жеке демеушілер RDWTI мен оның алдындағы адамдарға, Rand терроризм хронологиясына және Rand-MIPT терроризм оқиғаларының мәліметтер базасына қолдау көрсетті. Аталған мәліметтер қорында 40 000-нан астам терроризм жағдайлары кодталған және егжей-тегжейлі сипатталған. RAND қызметкерлері аймақтық тәжірибесі, тиісті тілдік дағдылары және елдегі жергілікті жерлерде жұмыс тәжірибесі бар қызметкерлерді тарта отырып, ықтимал террористік шабуылдар туралы кең зерттеу жүргізді. RDWTI – бұл пайдаланушыларға сапалы және жан-жақты мәліметтер беруге арналған толық қол жетімді және интерактивті мәліметтер базасы. Деректер базасы зерттеу және талдау үшін ақысыз және қол жетімді. RAND дерекқоры 1968 жылдан 2009 жылға дейінгі уақытты қамтиды. Терроризм орындаушылардың жеке басымен немесе себеп сипатымен емес, әрекеттің сипатымен анықталады; негізгі элементтерге мыналар кіреді:

Зорлық-зомбылық қаупі;

Қорқыныш пен алаңдаушылық тудыруға арналған;

Белгілі бір әрекеттерге мәжбүрлеуге арналған;

Мотив саяси мақсатты қамтуы керек;

Әдетте азаматтық мақсаттарға қарсы бағытталған;

Террористік инциденттер туралы хабарламалардың анықтамалары;

Инцидент идентификаторы: есеп жасалғаннан кейін әрбір есепке реттік нөмір беріледі.

Бұл сан RDWTI мәліметтер базасындағы оқиғалардың жалпы санына сәйкес келеді. Оқиға күні: террорлық шабуыл болған күн.

Дереккөз күні: жаңа дереккөздің жарияланған күні.

Ақпарат көзі: ақпарат көзінің атауы. Есептер, әдетте, екі немесе одан да көп дереккөздерге негізделген. Барлық бастапқы құжаттама әр есеп үшін қағаз түрінде сақталады.

Ішкі / халықаралық оқиға: бұл айнымалы үшін әдепкі мән – "ішкі оқиға". "Халықаралық" оқиға деп санау үшін шабуыл элементі шетелдік субъектімен (яғни орындаушы, нысана және т.б.) байланысты болуы керек.

АҚШ-тағы мүлікке/мүлікке шабуыл: әдепкі мән "жоқ"; егер АҚШ азаматы шабуылдың құрбаны болса немесе АҚШ-қа тиесілі мүлік шабуылға ұшыраса немесе бүлінсе, онда "иә"енгізіледі.

Өз-өзіне қол жұмсау миссиясы: әдепкі мән "жоқ"; егер шабуылдаушылар шабуыл аясында "суицид" тактикасын қолданса, онда "иә"енгізіледі.

Шабуыл туралы айтылған: әдепкі мән "жоқ"; егер орындаушылар тобы шабуыл туралы мәлімдесе және бұл мәлімдеме сенімді деп саналса, онда "иә"енгізіледі.

Үйлестірілген шабуыл: әдепкі мән – "жоқ"; егер шабуыл жүйелі түрде жоспарланған жеке шабуылдар сериясының бөлігі болса, онда "иә"енгізіледі.

Ескертпе: бір жерде бір мезгілде болған жарылыстар бір инцидент болып саналады.

Үзілген шабуыл: әдепкі бойынша "жоқ" мәні орнатылған; Егер шабуыл шабуылдаушылар жоспарланған шабуылды жүзеге асырмас бұрын тоқтатылса, онда "иә" енгізіледі.

Ел: шабуыл жасалатын ел.

Қала: шабуыл жасалатын қала.

1-Орындаушы: шабуылға жауапты топ. Жауапкершілікті сенімді мәлімдеме арқылы орнатуға болады немесе талдаушы өзінің аймақтық біліміне сүйене отырып, жауапты топқа кіре алады.

2-Орындаушы: егер екінші топ шабуылға жауапты болса, олар осында тізімделеді; әдепкі мән – «жоқ». Жауапкершілікті сенімді мәлімдеме арқылы орнатуға болады немесе талдаушы өзінің аймақтық біліміне сүйене отырып, жауапты топқа кіре алады.

Бірнеше қылмыскер (>2): әдепкі мән «жоқ» (Байдулла және т.б., 2021).

### **Зерттеу әдістері**

Экстремизмге қарсы іс-қимыл – мемлекеттік органдардың адам мен азаматтың құқықтары мен бостандықтарын, конституциялық құрылыс негіздерін, Қазақстан Республикасының тұтастығы мен ұлттық қауіпсіздігін экстремизмнен қорғауды қамтамасыз етуге, экстремизмнің алдын алуға, оны анықтауға, жолын кесуге және оның салдарларын жоюға, сондай-ақ экстремизмді жүзеге асыруға ықпал ететін себептер мен жағдайларды анықтауға және жоюға бағытталған қызметі.

Экстремизмге қарсы іс-қимыл мынадай негізгі бағыттар бойынша жүзеге асырылады:

экстремизмнің алдын алуға, оның ішінде оны жүзеге асыруға ықпал ететін себептер мен жағдайларды анықтауға және кейіннен жоюға бағытталған профилактикалық шаралар қабылдау;

экстремизмді анықтау және оның жолын кесу;

экстремизмге қарсы іс-қимыл саласындағы халықаралық ынтымақтастық (Қазақстан Республикасының №31 Заңы.).

Мемлекеттік органдар өз құзыреті шегінде экстремизмнің алдын алуға бағытталған мынадай профилактикалық шараларды іске асырады:

1) Діни қызмет саласындағы мемлекеттік реттеуді жүзеге асыратын мемлекеттік орган Қазақстан Республикасының аумағында құрылған діни бірлестіктер мен миссионерлердің қызметіне зерделеу және талдау жүргізеді, өз құзыретіне жататын мәселелер бойынша ақпараттық-насихаттау іс-шараларын жүзеге асырады, Қазақстан Республикасының Діни қызмет және діни бірлестіктер туралы заңнамасын бұзуға қатысты мәселелерді қарайды, діни қызметке тыйым салу туралы ұсыныстар енгізеді;

2) бұқаралық ақпарат құралдары істері жөніндегі уәкілетті орган бұқаралық ақпарат құралдары өнімдерінде экстремизмді насихаттауға және ақтауға жол бермеу, олардың Қазақстан Республикасының заңнамасын сақтауы

тұрғысынан мониторинг жүргізеді, мемлекеттік тапсырысты орындайтын бұқаралық ақпарат құралдарында ұлтаралық және конфессияаралық келісімді нығайту мәселелерінің жария етілуін қамтамасыз етеді;

3) білім беру саласындағы орталық атқарушы орган білім беру ұйымдарында білімалушылардың экстремизм идеяларын қабылдамауын, халықаралық құқық пен ізгіліктің жалпыға танылған қағидаттарын құрметтеуін қалыптастыруға бағытталған тәрбиелік бағдарламалардың бекітілуін және іске асырылуын қамтамасыз етеді, оқу орындарының студенттер алмасу мәселелері бойынша халықаралық шарттарының сақталуын бақылауды жүзеге асырады;

4) Қазақстан Республикасының Ұлттық қауіпсіздік органдары жедел-ізвестіру, қарсы барлау іс-шараларын жүргізеді және Қазақстан Республикасының заңнамасына сәйкес мемлекеттік органдардың дәлелді қорытындылары бойынша өздерінің іс-әрекеттерімен қоғам мен мемлекеттің қауіпсіздігіне қатер төндіретін немесе нұқсан келтіретін шетелдіктер мен азаматтығы жоқ адамдардың Қазақстан Республикасына келуіне жол бермеу жөніндегі шараларды жүзеге асырады;

5) Қазақстан Республикасының Ішкі істер органдары жедел-ізвестіру қызметін, қоғамдық тәртіпті қорғау және қоғамдық қауіпсіздікті қамтамасыз ету жөніндегі атқарушылық және өкімдік функцияларды жүзеге асырады, сондай-ақ өз әрекеттерімен қоғам мен мемлекеттің қауіпсіздігіне қатер төндіретін немесе нұқсан келтіретін шетелдіктер мен азаматтығы жоқ адамдарды Қазақстан Республикасының заңнамасына сәйкес Қазақстан Республикасынан шығарып жіберуді жүзеге асырады;

6) облыстардың (республикалық маңызы бар қалалардың, астананың), аудандардың (облыстық маңызы бар қалалардың) жергілікті атқарушы органдары қоғамдық бірлестіктермен өзара іс-қимылды, тиісті аумақтарда құрылған діни бірлестіктер мен миссионерлердің қызметін зерделеуді жүзеге асырады, олар туралы деректер банкін құрады, өздерінің құзыретіне жататын мәселелер бойынша өңірлік деңгейде ақпараттық-насихаттау іс-шараларын жүзеге асырады, Қазақстан Республикасының заңнамасында белгіленген тәртіппен жергілікті атқарушы өңірдегі діни ахуалды зерделеу және талдау тапсырмасын орындайды.

7) Сыртқы барлау субъектілері өз іс-әрекеттерімен қоғам мен мемлекеттің қауіпсіздігіне қатер төндіретін немесе нұқсан келтіретін шет мемлекеттердің ұйымдарына, шетелдіктер мен азаматтығы жоқ адамдарға қатысты Қазақстан Республикасының мемлекеттік органдарын хабардар етуді жүзеге асырады.

Мемлекеттік органдардың экстремизмді анықтау және оның жолын кесу жөніндегі құзыретіне келетін болсақ:

1. Ұлттық қауіпсіздік, ішкі істер органдары және экономикалық тергеу қызметі Қазақстан Республикасының заңнамасымен осы органдардың қарауына жатқызылған қылмыстық құқық бұзушылықтарды анықтайды, жолын кеседі, ашады және тергейді, сондай-ақ Қазақстан Республикасының заңдарында көзделген өзге де өкілеттіктерді жүзеге асырады.

1-1. Экономикалық тергеу қызметі экстремизмді қаржыландыру көздерінің, арналары мен тәсілдерінің алдын алуды, анықтауды, жолын кесуді жүзеге асырады.

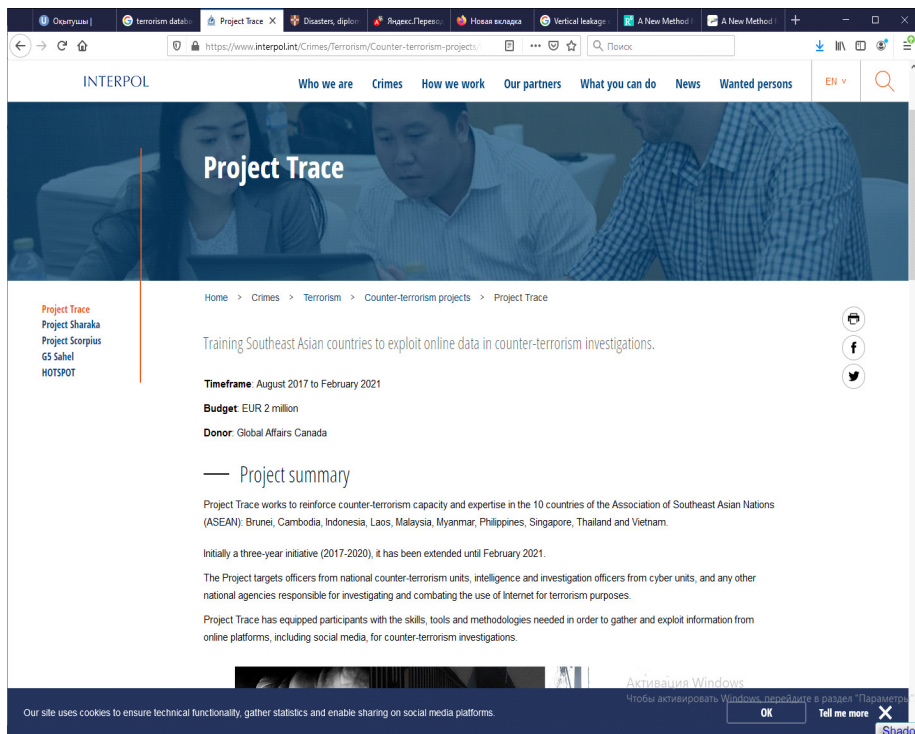
2. Прокурорлар жеке және заңды тұлғалардың, олардың құрылымдық бөлімшелерінің (филиалдары мен өкілдіктерінің) Қазақстан Республикасының Экстремизмге қарсы іс-қимыл саласындағы заңнамасын бұзу фактілері анықталған кезде немесе дайындалып жатқан құқыққа қарсы әрекеттер туралы мәліметтер болған кезде, сондай-ақ бұқаралық ақпарат құралдары арқылы адамның және азаматтың құқықтары мен бостандықтарына, сондай-ақ олардың мүдделеріне зиян келтіруі мүмкін экстремистік материалдар таратылған жағдайда экстремизмнің кез келген көріністерін жою туралы прокурорлық қадағалау актілерін енгізеді, жүзеге асыруға ықпал еткен себептер мен жағдайларды ескере отырып, бұзылған құқықтарды қалпына келтіру туралы сотқа ұйымдар экстремизмді жүзеге асырған жағдайда олардың қызметіне тыйым салу туралы өтініш береді, сондай-ақ Қазақстан Республикасының заңдарында белгіленген тәртіппен және шектерде қылмыстық қудалауды жүзеге асырады.

3. Өзге де мемлекеттік органдар экстремизмді анықтауға және оның жолын кесуге Қазақстан Республикасының заңдарында белгіленген өз құзыреті шегінде қатысады ("Шарака" Жобасы, 2022)

Экстремизммен күресуге арналған жобалар

Тгасе жобасы

Тгасе жобасының мақсаты — Оңтүстік-Шығыс Азия елдерін терроризмге қарсы тергеулерде онлайн-деректерді қолдануға үйрету. Мерзімі: 2017 жылдың тамызынан 2021 жылдың ақпанына дейін. Бюджет: 2 миллион евро. Тгасе жобасы Оңтүстік-Шығыс Азия мемлекеттері қауымдастығының (ASEAN) 10 елінде: Бруней, Индонезия, Камбоджа, Лаос, Малайзия, Мьянма, Филиппин, Сингапур, Таиланд және Вьетнамда терроризмге қарсы күрес саласындағы әлеует пен сараптамалық білімді нығайту бойынша жұмыс істейді. Жоба терроризммен күрес жөніндегі ұлттық бөлімшелердің қызметкерлеріне, кибернетикалық бөлімшелердің барлау және тергеу бөлімшелерінің қызметкерлеріне және интернетті террористік мақсаттарда пайдалануға қарсы тергеу мен күреске жауапты кез келген басқа ұлттық мекемелердің қызметкерлеріне бағытталған. Тгасе жобасы қатысушыларға терроризмге қарсы тергеу жүргізу үшін онлайн платформалардан, соның ішінде әлеуметтік желілерден ақпарат жинауға және пайдалануға қажетті дағдылар, құралдар мен әдістемелерді ұсынды (2-сурет).



2-сурет – Экстремизмге қарсы Трасе жобасы

Әр цикл келесі әрекеттерді қамтиды:

Интернетті террористік мақсатта пайдалануға қарсы тұру үшін апталық базалық дайындық;

Интернетті террористік мақсатта пайдалануға қарсы тұру бойынша бір апталық біліктілікті арттыру курсы;

Бағдарламалық жасақтаманы сыйға тарту және бағдарламалық жасақтаманы мамандандырылған оқыту;

Жаттықтырушыға арналған жаттығу;

Қатысушыларға практикалық тәжірибе беруге арналған үстелдік жаттығу.

Аталған жобаның мақсаттары келесідей:

Қылмыстық тергеу аясында ашық бастапқы (OSINT) және әлеуметтік медиа (SOCMINT) деректер ағындарын қалай пайдалануға болатындығын түсіну;

Интерполдың полиция мүмкіндіктерін, соның ішінде әртүрлі мәліметтер базасы мен интерполдың хабарламаларын қалай пайдалану керектігін түсіну;

Сандық дәлелдемелерді жинау мен қорғаудың дұрыс әдістерін түсіну;

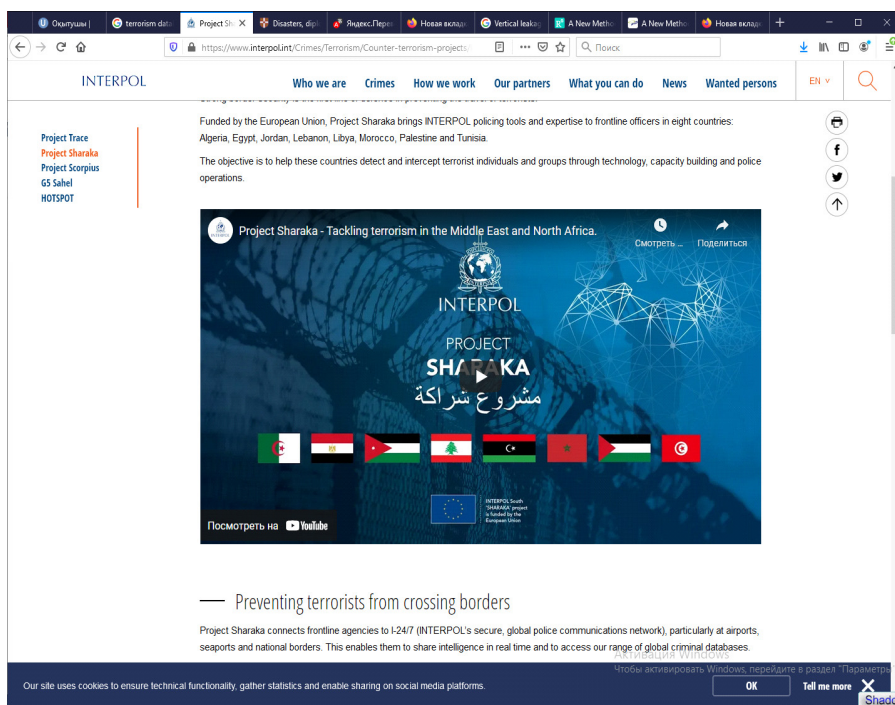
Талдауға көмектесетін құрал ретінде графикалық әдістерді қолдану;

Үшінші тараптардан (Интерпол, басқа да құқық қорғау органдары, интернет-провайдерлер, телекоммуникациялық компаниялар және т.б.) ақпаратты қалай сұрау керектігін түсіну (Чен, 2011).

### Sharaka жобасы

Sharaka жобасы террористердің шекараны кесуінің алдын алуға бағытталған. Аталған жоба озық агенттіктерді I-24/7 (Интерполдың қорғалған жаһандық полиция коммуникациялық желісі), әсіресе әуежайларда, теңіз порттарында және ұлттық шекараларда қосады. Бұл оларға нақты уақыт режимінде барлауды бөлісуге және қылмыс туралы жаһандық мәліметтер базасына қол жеткізуге мүмкіндік береді.

Террористер, әсіресе қақтығыс аймақтарынан оралған шетелдік содыр-террористер пайдаланатын ұрланған жол жүру құжаттарына байланысты шекара қызметі қызметкерлерінің ұрланған және жоғалған жол жүру құжаттары туралы интерполдың деректер базасына тікелей қол жеткізуі өте маңызды (3-сурет).



3-сурет – Экстремизмге қарсы Sharaka жобасы

Бұл жоба мақсатты елдердің терроризмге қарсы күрес саласында қажетті білімге, құрал-жабдықтар мен дағдыларға ие болуын қамтамасыз етеді. Алдыңғы қатардағы қызметкерлер аймақтық тергеулер мен операциялар барысында Интерполдың қылмыстық істер бойынша бірқатар жаһандық дерекқорларын пайдалану бойынша дайындықтан өтеді ("Скорпион" Жобасы.).

### Scorpius жобасы

Scorpius жобасы терроризм мен онымен байланысты трансұлттық

қылмыстың алдын алуға және жолын кесуге бағытталған оңтүстік және Оңтүстік-Шығыс Азиядағы құқық қорғау органдарының әлеуетін арттыру жөніндегі екі жылдық (2017–2019 жж.) бастама болды. Оны интерпол мен Канада үкіметі бірлесіп қаржыландырды.

Біртұтастәсілді қолдана отырып, жоба құқық қорғау қоғамдастығының тиісті қатысушыларын, соның ішінде негізгі шешім қабылдаушыларды, терроризмге және трансұлттық қылмысқа қарсы күрес жөніндегі тергеушілерді, барлау қызметкерлерін, Интерполдың Ұлттық Орталық бюросының қызметкерлерін, қылмыстық сот төрелігі органдарын, прокурорларды және полицияның оқу орындарын біріктірді.

Жаһандық терроризмге қарсы стратегияны қолдай отырып, жоба келесі бенефициар елдерге бағытталған: Бангладеш, Үндістан, Индонезия, Малайзия, Мальдив, Непал, Пәкістан, Филиппин, Шри-Ланка.

Тергеу семинарлары қылмыстық және террористік іс-әрекеттің алдын алу, анықтау және тергеу үшін қажетті практикалық дағдыларды қамтамасыз етеді:

Күдікті трансұлттық қылмыстық желілер мен олардың филиалдарын анықтау;

Ұлттық және халықаралық деңгейлерде тиісті құқық қорғау органдары арасында үйлестіру;

Интерпол және құлақтандыру дерекқорына террористердің бейіндерімен байланысты биометриялық деректерді жүйелі түрде қосу;

Қару-жарақ пен материалдардың заңсыз айналымын анықтау, қадағалау және жолын кесу;

CBRNE инциденттеріне және үйдегі жарылғыш құрылғыларға қатысты заттар мен жұмыс әдістері туралы ақпараттармен бөлісу.

Оқу курстары бағалау кезеңінде анықталған қажеттіліктердің әртүрлі салаларына бағытталған, соның ішінде: ашық бастапқы және әлеуметтік медиа тергеулер – бұл ашық бастапқы барлау және әлеуметтік медиа платформаларын қолдана отырып жүргізілген тергеулер арқылы терроризмнің алдын-алудың және онымен күресудің өзекті және тұрақты әдістері. Сондай-ақ, қатысушылар трансұлттық қылмысқа қарсы терроризмге қарсы және онымен байланысты тергеулер барысында жеке өмірге қол сұқпаушылық, адам құқықтары мен сөз бостандығы туралы білді.

CBRNE қару-жарақтары мен материалдары – мүше елдер арасында жаппай қырып-жою қаруы және қолдан жасалған жарылғыш құрылғылар туралы, әсіресе әскери және құқық қорғау органдары арасында барлау алмасуға жәрдемдесу үшін қарастырылады. Сессия сондай-ақ химиялық биологиялық радиологиялық және ядролық және жарылғыш материалдар туындатқан террористік қатерлер мен инциденттерге ден қою жөніндегі елдердің әлеуетін нығайтуға бағытталған.

Қылмыс орнындағы тергеу және апат құрбандарын анықтау (DVI) – жарылыс орнын тексеру, DVI әдістері және терроризмге қарсы күрес

контекстіндегі қылмыс орнындағы жалпы тергеу. Қатысушылар мұндай оқиғаларға тиімді жауап беруді және Террористік актілерді жасағандарды қудалауды қолдау үшін өмірлік дәлелдер жинауды үйренді.

Аналитикалық семинарлар офицерлерге қылмыстық және террористік әрекеттерді тиімді талдау үшін қажет практикалық дағдыларды ұсынады:

Барлау процесінің тұжырымдамалары мен компоненттерін түсіну;

Талдамалық бағаларды, баяндамалар мен брифингтер дайындау;

Ашық бастапқы және әлеуметтік медиа зерттеулерінің тұжырымдамаларын, әдістері мен құралдарын түсіну және қолдану;

Терроризмді қаржыландыру жағдайларын талдау үшін практикалық ақпарат алу [8].

Dark web project AI Lab Dark Web жобасы

Dark web project AI Lab Dark Web – бұл халықаралық терроризм құбылыстарын (жихадшыларды) зерттеуге және түсінуге бағытталған ұзақ мерзімді ғылыми-зерттеу бағдарламасы. Зерттеушілер халықаралық террористік топтар құрған веб-сайттарды, форумдарды, чаттарды, блогтарды, әлеуметтік желілер сайттарын, бейнелерді, виртуалды әлемді және т. б. қоса алғанда, "бүкіл" веб-контентті жинауға ұмтылады. Олар көп тілді деректерді талдаудың, мәтінді талдаудың және веб-талдаудың әртүрлі әдістерін жасады, сілтемені талдау, мазмұнды талдау, вебкөрсеткіштерді талдау (техникалық күрделілік), көңіл-күйді талдау, авторлықты талдау және зерттеу барысында бейнені талдау есептерін қарастырды. Осы жоба аясында жасалған тәсілдер мен әдістер барлау және қауіпсіздік информатикасы саласын дамытуға ықпал етеді. Мұндай жетістіктер тиісті мүдделі тараптарға терроризм саласында зерттеулер жүргізуге және халықаралық қауіпсіздік пен бейбітшілікті қамтамасыз етуге ықпал етуге көмектеседі. Зерттеушілер сандық кітапханадағы алдыңғы зерттеулер негізінде әртүрлі мамандандырылған өрмекшілер/сканерлер жасады. Құрастырылған өрмекшілер парольмен қорғалған сайттарға кіріп, рандомизацияланған (гуманоидты) үлгіні жасай алады. Олар веб-сайттағы барлық файлдарды, сілтемелерді, PHP, CGI және ASP файлдарын, суреттерді, аудио және бейнелерді шығаруға үйретілген. Жаңалықты қамтамасыз ету үшін әр 2-3 ай сайын таңдалған веб-сайттарды қарап шығады.

Форумдарда өрмекшілерді құруға арналған құрал 15+ форум хостинг бағдарламалық жасақтамасын және олардың форматтарын таниды. Зерттеушілер қатысушылардың өзара әрекеттесуін қайта құруға мүмкіндік беретін авторлар, тақырыптар, жарияланымдар, тақырыптар, уақыт белгілері және т.б. сияқты толық форумды жинайды. Олар форумдардың мерзімді желісін және зерттеу қажеттіліктеріне негізделген біртіндеп жаңартуды жүзеге асырады. Зерттеушілер форумның мазмұнын араб, ағылшын, испан, француз және қытай тілдерінде компьютерлік лингвистиканың таңдаулы әдістерін қолдана отырып жинады және өңдеді.

Көңіл-күйді талдау (полярылық: он/теріс) және әсер ету (эмоциялар: зорлық-



зомбылық, нәсілшілдік, ашу және т.б.) әрі қарай зерттеуді қажет ететін радикалды және зорлық-зомбылық сайттарын анықтауға мүмкіндік береді. Сондай-ақ, зерттеушілер радикалды идеялардың мазмұнына, жіберушілеріне және олардың өзара әрекеттесуіне негізделе отырып, олардың қаншалықты "жұкпалы" болатындығын зерттейді. Сондай-ақ, олар уақыт пен адамдардың көңіл-күйін/өзгеруін зерттеу үшін жеке визуализация әдістерін жасады. Зерттеуге әсер ететін бірнеше ықтималды көптілді лексика және өлшемді азайту мен болжаудың тандалған әдістері (мысалы, негізгі компоненттерді талдау) кіреді.

Авторлықты талдау бойынша зерттеулерге негізделе отырып, анонимді жіберушілерді олардың форумдағы хабарламаларына байланысты қолтаңбалар негізінде бірегей сәйкестендіруге мүмкіндік беретін жазу әдістемесін (кибер) әзірледі. Олар дәстүрлі авторлық талдаудың лексикалық және синтаксистік ерекшеліктерін жүйелік (мысалы, қаріп өлшемі, түс, веб-сілтемелер) және семантикалық (мысалы, зорлық-зомбылық) қамтиды.

### **Қорытынды**

Жұмыста мәтіндік деректер Интернеттегі мазмұнның негізгі бөлігін құрайтыны анықталды және экстремистер бұл ресурсты өз мақсаттарына жету үшін белсенді пайдаланады. Бұл интернеттегі қажетсіз ресурстарды сүзу әдістерін әзірлеу мен енгізудің маңыздылығын көрсетеді.

Алайда, Интернеттегі экстремизмге қарсы күрес стратегияларды үнемі бақылауды, талдауды және бейімдеуді талап ететін күрделі және серпінді міндет болып табылады. Тиімді шаралар тек техникалық шешімдерді ғана емес, сонымен қатар құқық қорғау органдары, интернет-провайдерлер және азаматтық қоғам арасындағы ынтымақтастықты да қамтуы керек.

Интернеттегі экстремизммен және агрессивті ақпаратпен күресу технологияның дамуына және экстремистік топтардың мінез-құлқындағы өзгерістерге сәйкес стратегияларды жан-жақты тәсілді және үнемі жаңартуды қажет ететінін атап өткен жөн.

Берілген мақалада экстремизм түсінігіне анықтама беріліп, оның әр түрлі сипаттары бойынша жіктелімі келтірілген. Сонымен қатар, авторлар әр түрлі халықаралық экстремизмге қарсы күреске қатысты ғылыми жобаларға талдау жасалып, аталған жобалардың ерекшеліктері келтіріледі. Берілген зерттеу жұмысын веб-ресурстардағы экстремистік мазмұнды анықтауға арналған жұмыстар үшін бастапқы дереккөз ретінде пайдалануға болады.

### **ӘДЕБИЕТТЕР**

Қазақстан Республикасының №31 Заңы. Экстремизмге қарсы іс-қимыл туралы: 2005 жылдың 18 ақпанында бекітілген [https://adilet.zan.kz/kaz/docs/Z050000031\\_](https://adilet.zan.kz/kaz/docs/Z050000031_)

Болатбек М.А., Экстремизм түсінігі. Экстремистік мәтіндерді анықтауға арналған белгілер жинағына шолу // "Фараби әлемі" студенттер мен жас ғалымдардың халықаралық ғылыми конференциясы.–Алматы: Қазақ университеті, 2020. –Б. 43–44.

Терроризм туралы Ғаламдық мәліметтер базасы. <https://www.start.umd.edu/gtd/>. 10.03.2022.

RAND бүкіл әлем бойынша террористік актілер туралы мәліметтер базасы. <https://www.rand.org/nsrd/projects/terrorism-incidents.html> 02.02.2022

Байдулла А.М., Мусиралиева Ш.Ж., Болатбек М.А. Экстремистік топтарды анықтау және талдау // "Фараби әлемі" студенттер мен жас ғалымдардың халықаралық ғылыми конференциясы. –Алматы: Қазақ университеті, 2021. – Б.74.

Жобаны қадағалау. <https://www.interpol.int/Crimes/Terrorism/Counter-terrorism> -жобалар / Project-Tracker 2. 03.03.2022.

"Шарака" Жобасы. <https://www.interpol.int/Crimes/Terrorism/Counter-terrorizm> -жоба-Шарака. 03.03.2022.

"Скорпион" Жобасы. <https://www.interpol.int/Crimes/Terrorism/Counter-terrorizm> -жобалар/ Project-Scorpius. 03.03.2022.

Чен, х. қараңғы веб: Интернеттің қараңғы жағын зерттеу және өндіру. - Берлин: Спрингер, Гейдельберг, 2011. –С. 1-450.

#### REFERENCES

Baidulla A.M., Musiralieva SH.ZH., Bolatbek M.A. identification and analysis of extremist groups // International Scientific Conference of students and young scientists "Farabi's world". - Almaty: Kazakh University, 2021. - P. 74.

Law Of The Republic Of Kazakhstan No. 31. On countering extremism: approved on February 18, 2005. <https://adilet.zan.kz/kaz/docs/Z050000031>

Bolatbek M.A., The concept of extremism. Review of the collection of signs for identifying extremist texts // International Scientific Conference of students and young scientists "Farabi's world". - Almaty: Kazakh University, 2020. - Pp. 43–44.

Global Terrorism Database. <https://www.start.umd.edu/gtd/>. 10.03.2022.

RAND Database of Worldwide Terrorism Incidents. <https://www.rand.org/nsrd/projects/terrorism-incidents.html>. 02.02.2022

Baidulla A.M., Musiralieva Sh.Zh., Bolatbek M.A. identification and analysis of extremist groups // International Scientific Conference of students and young scientists "Farabi's world". - Almaty: Kazakh University, 2021. - P. 74.

Project Trace. <https://www.interpol.int/Crimes/Terrorism/Counter-terrorism-projects/Project-Trace2>. 03.03.2022.

Project Sharaka. <https://www.interpol.int/Crimes/Terrorism/Counter-terrorism-projects/Project-Sharaka>. 03.03.2022.

Project Scorpius. <https://www.interpol.int/Crimes/Terrorism/Counter-terrorism-projects/Project-Scorpius>. 03.03.2022.

Chen, H. Dark Web: Exploring and Mining the Dark Side of the Web. –BerlinSpringer, Heidelberg, 2011. –Pp.1–450.

NEWS OF THE NATIONAL ACADEMY OF SCIENCES OF THE REPUBLIC OF  
KAZAKHSTAN  
PHYSICO-MATHEMATICAL SERIES  
ISSN 1991-346X  
Volume 3. Number 347 (2023). 131–146  
<https://doi.org/10.32014/2023.2518-1726.209>

UDC 004.89

© **D. Oralbekova**<sup>1,2\*</sup>, **O. Mamyrbayev**<sup>1</sup>, **A. Zhunussova**<sup>3</sup>,  
**B. Zhumazhanov**<sup>1</sup>, 2023

<sup>1</sup>Institute of information and computational technologies, CS MSHE RK,  
Almaty, Kazakhstan;

<sup>2</sup>Almaty University of Power Engineering and Telecommunications named after  
Gumarbek Daukeyev, Almaty, Kazakhstan;

<sup>3</sup>Narxoz University, Almaty, Kazakhstan.

E-mail: dinaoral@mail.ru

## STUDY OF MODERN METHODS OF LANGUAGE MODELING FOR A LANGUAGE WITH A COMPLEX MORPHOLOGICAL STRUCTURE

**Oralbekova Dina** — Doctor PhD. Senior Researcher, Associate Professor. Institute of information and computational technologies. Almaty, Kazakhstan

E-mail: dinaoral@mail.ru. ORCID ID: 0000-0003-4975-6493;

**Мамырбаев Оркен** — Doctor PhD. Associate Professor, Deputy General Director. Institute of information and computational technologies. Almaty, Kazakhstan

E-mail: morkenj@mail.ru. ORCID ID: 0000-0001-8318-3794;

**Zhunussova Aliya** — Senior Lecturer. Narxoz University, Almaty, Kazakhstan

E-mail: alia\_94-22@mail.ru. ORCID ID: 0000-0002-3641-8260;

**Zhumazhanov Bagashar** — Candidate of technical sciences. Senior Researcher. Institute of information and computational technologies. Almaty, Kazakhstan

E-mail: bagasharj@mail.ru. ORCID ID: 0000-0002-5035-9076.

**Abstract.** This scientific paper presents a comparative analysis of contemporary language modeling methods and their application to the Kazakh language, which is characterized by its complex morphological structure. Language modeling involves training machine learning models to predict word probabilities within a given context. The primary focus of this study is the investigation of the BERT model (Bidirectional Encoder Representations from Transformers) and its effectiveness in modeling languages with diverse morphological patterns. The article provides an overview of n-gram models and recurrent neural networks, highlighting their limitations in capturing long-term dependencies and semantic relationships in text. Then the BERT model, its architecture and principles of operation, including attention mechanisms and multi-level Transformer blocks, are considered. The following are the results of the study, including the adaptation of the BERT model

to languages with a complex morphological structure, including Kazakh. It is shown that the BERT model demonstrates high accuracy in modeling contextual dependencies and semantic relationships between words in such languages. The article emphasizes the importance and prospects of applying modern methods of language modeling, especially the BERT model, for languages with a complex morphological structure. She also points out the need for further research in the field of adaptation of the BERT model to specific languages, the development of new architectures and methods, as well as solving the challenges associated with rare and sparsely spoken languages, the results of this study will help improve the understanding and effectiveness of language processing of text in the Kazakh language, and also contribute to the development of the NLP field as a whole.

**Keywords:** language modeling, Kazakh language, n-gram, BERT, GPT, LSTM

**Financing:** *This research has been/was/is funded by the Science Committee of the Ministry of Education and Science of the Republic of Kazakhstan (Grant No. AP19174298).*

**Conflict of interest:** *The authors declare that there is no conflict of interest.*

© Д. Оралбекова<sup>1,2\*</sup>, О. Мамырбаев<sup>1</sup>, А. Жунусова<sup>3</sup>, Б. Жумажанов<sup>1</sup>, 2023

<sup>1</sup>Институт информационных и вычислительных технологий КН МНВО РК,  
Алматы, Казахстан;

<sup>2</sup>Алматинский Университет Энергетики и Связи им. Г. Даукеева,  
Алматы, Казахстан;

<sup>3</sup>Narхоз University, Алматы, Казахстан.  
E-mail: dinaoral@mail.ru

## КҮРДЕЛІ МОРФОЛОГИЯЛЫҚ ҚҰРЫЛЫМЫ БАР ТІЛГЕ АРНАЛҒАН ЗАМАНАУИ ТІЛДІК МОДЕЛЬДЕУ ӘДІСТЕРІН ЗЕРТТЕУ

**Оралбекова Дина** — PhD докторы. Аға ғылыми қызметкер, доцент. Ақпараттық және есептеуіш технологиялар институты. Алматы, Қазақстан  
E-mail: dinaoral@mail.ru. ORCID ID: 0000-0003-4975-6493;

**Мамырбаев Оркен** — PhD докторы. Қауымдастырылған профессор, бас директордың орынбасары. Ақпараттық және есептеуіш технологиялар институты. Алматы, Қазақстан  
E-mail: morkenj@mail.ru. ORCID ID: 0000-0001-8318-3794;

**Жунусова Алия** — Аға оқытушы. Нархоз университеті. Алматы, Қазақстан  
E-mail: alia\_94-22@mail.ru. ORCID ID: 0000-0002-3641-8260;

**Жумажанов Бағашар** — Техникалық ғылымдардың кандидаты. Аға ғылыми қызметкер, доцент. Ақпараттық және есептеуіш технологиялар институты. Алматы, Қазақстан  
E-mail: bagasharj@mail.ru. ORCID ID: 0000-0002-5035-9076.

**Аннотация.** Бұл ғылыми мақала тілдік модельдеудің заманауи әдістерін салыстырмалы талдауды және оларды қазақ тілі сияқты агглютинативті тілге қолдануды ұсынады. Тілдік модель – бұл сөз бойынша ықтималдылықты бөлуді жүзеге асыруға үйретілген машиналық оқыту моделінің бір түрі. Тілдік

модель белгілі бір мәтіннің контекстіне сүйене отырып, сөйлемдегі немесе сөз тіркесіндегі бос орынды толтыру үшін келесі ең қолайлы сөзді болжауға тырысады. BERT (Bidirectional Encoder Representations from Transformers) моделін және оның морфологиялық әртүрлілігімен сипатталатын тілдерді тиімді модельдеу қабілетін зерттеуге баса назар аударылды. Мақалада n-грамм модельдеріне, қайталанатын нейрондық желілерге және олардың мәтіндегі ұзақ мерзімді тәуелділіктер мен семантикалық қатынастарды түсірудегі шектеулеріне шолу жасалды. Содан кейін BERT моделі, оның архитектурасы және жұмыс принциптері, соның ішінде назар аудару механизмдері және көп деңгейлі Transformer блоктары егжей-тегжейлі қарастырылды. Бұдан әрі BERT моделін қазақ тілін қоса алғанда, күрделі морфологиялық құрылымы бар тілдерге бейімдеуді қамтитын зерттеу нәтижелері ұсынылды. BERT моделі осындай тілдердегі сөздер арасындағы контекстік тәуелділіктер мен семантикалық қатынастарды модельдеуде жоғары дәлдікті көрсететіні анықталды. Мақалада күрделі морфологиялық құрылымы бар тілдер үшін заманауи тілдік модельдеу әдістерін, әсіресе BERT модельдерін қолданудың маңыздылығы мен перспективалары көрсетілген. Ол сондай-ақ BERT моделін нақты тілдерге бейімдеу, жаңа архитектуралар мен әдістерді әзірлеу, сондай-ақ сирек кездесетін тілдерге байланысты сын-қатерлерді шешу саласында одан әрі зерттеу қажеттігін көрсетеді. Осы зерттеудің алынған нәтижелері қазақ тіліндегі мәтінді тілдік өңдеудің түсінігі мен тиімділігін арттыруға өз үлесін қосады, сондай-ақ жалпы NLP саласының дамуына ықпал етеді.

**Түйін сөздер:** тілдік модельдеу, қазақ тілі, n-грамм, BERT, GPT, LSTM

**Қаржыландыру:** Бұл зерттеу Қазақстан Республикасы Ғылым және жоғары білім министрлігі, Ғылым комитетімен қаржыландырылған (Грант № AP19174298).

**Мүдделер қақтығысы:** Авторлар осы мақалада мүдделер қақтығысы жоқ деп мәлімдейді.

© Д. Оралбекова<sup>1,2\*</sup>, О. Мамырбаев<sup>1</sup>, А. Жунусова<sup>3</sup>, Б. Жумажанов<sup>1</sup>, 2023

<sup>1</sup>Институт информационных и вычислительных технологий КН МНВО РК, Алматы, Казахстан;

<sup>2</sup>Алматинский Университет Энергетики и Связи им. Г. Даукеева, Алматы, Казахстан;

<sup>3</sup>Narxoz University, Алматы, Казахстан.

E-mail: dinaoral@mail.ru

## ИССЛЕДОВАНИЕ СОВРЕМЕННЫХ МЕТОДОВ ЯЗЫКОВОГО МОДЕЛИРОВАНИЯ ДЛЯ ЯЗЫКА СО СЛОЖНОЙ МОРФОЛОГИЧЕСКОЙ СТРУКТУРОЙ

**Оралбекова Дина** — Доктор PhD. Старший научный сотрудник, доцент. Институт информационных и вычислительных технологий. Алматы, Казахстан  
E-mail: dinaoral@mail.ru. ORCID ID: 0000-0003-4975-6493;

**Мамырбаев Оркен** — Доктор PhD. Ассоциированный профессор, заместитель генерального директора. Институт информационных и вычислительных технологий, Алматы, Казахстан  
E-mail: morkenj@mail.ru. ORCID ID: 0000-0001-8318-3794;

**Жунусова Алия** — Старший преподаватель. Университет Нархоз. Алматы, Казахстан  
E-mail: alia\_94-22@mail.ru. ORCID ID: 0000-0002-3641-8260;

**Жумажанов Багашар** — Кандидат технических наук. Старший научный сотрудник. Институт информационных и вычислительных технологий. Алматы, Казахстан  
E-mail: bagasharj@mail.ru. ORCID ID: 0000-0002-5035-9076.

**Аннотация.** Данная научная статья представляет сравнительный анализ современных методов языкового моделирования и их применение к агглютинативному языку, такому как казахский язык. Языковая модель — это тип модели машинного обучения, призванной проводить распределение вероятностей по словам. Языковая модель пытается предсказать следующее наиболее подходящее слово для заполнения пробела в предложении или фразе, исходя из контекста определенного текста. Основной акцент сделан на изучение модели BERT (Bidirectional Encoder Representations from Transformers) и ее способности эффективно моделировать языки, характеризующиеся морфологической разнообразностью. В статье представлен обзор n-граммных моделей, рекуррентных нейронных сетей и их ограничений в улавливании долгосрочных зависимостей и семантических отношений в тексте. Затем подробно рассмотрена модель BERT, ее архитектура и принципы работы, включая механизмы внимания и многоуровневые Transformer блоки. Далее представлены результаты исследования, включающие адаптацию модели BERT к языкам со сложной морфологической структурой, включая казахский язык. Показано, что модель BERT демонстрирует высокую точность в моделировании контекстуальных зависимостей и семантических отношений между словами в таких языках. Статья подчеркивает значимость и перспективы применения современных методов языкового моделирования, особенно модели BERT, для языков со сложной морфологической структурой. Она также указывает на необходимость дальнейших исследований в области адаптации модели BERT к конкретным языкам, разработки новых архитектур и методов, а также решения вызовов, связанных с редкими и малораспространенными языками. Полученные результаты этого исследования помогут улучшить понимание и эффективность языковой обработки текста на казахском языке, а также способствуют развитию области NLP в целом.

**Ключевые слова:** языковое моделирование, казахский язык, n-граммы, BERT, GPT, LSTM

**Финансирование:** Данное исследование финансировалось Комитетом науки Министерства науки и высшего образования Республики Казахстан (Грант No AP19174298).

**Конфликт интересов:** Авторы заявляют об отсутствии конфликта интересов.

### **Кіріспе**

Тілдік модельдер табиғи тілді өңдеудің (NLP) негізгі компоненті болып табылады, өйткені олар машиналарға адам тілін түсінуге, құруға және талдауға мүмкіндік береді. Олар негізінен кітаптар немесе мақалалар жинағы сияқты мәтіннің үлкен жиынтығын қолдана отырып оқытылады. Содан кейін модельдер сөйлемдегі келесі сөзді болжау немесе грамматикалық тұрғыдан дұрыс және семантикалық тұрғыдан сәйкес келетін жаңа мәтін жасау үшін осы оқу деректерінен алатын үлгілерді пайдаланады.

Тіл моделін жасау табиғи тілді өңдеудегі (NLP) негізгі міндет болып табылады және көптеген қосымшаларда жаңа мәтін құру, машиналық аударма, Автоматты реферат және басқалар сияқты маңызды элементті ұсынады. Түркі тілдерінің тобына кіретін морфологиялық тұрғыдан күрделі тілдер жеткілікті, олар тиімді модельдеуде басқа талаптарды қажет етеді.

Агглютинативті тіл, қазақ тілі сияқты, грамматикалық тұлғаны, санды, уақытты, жағдайды және басқа морфологиялық сипаттамаларды ескере отырып, сөздер өзгертін бай морфологиясы бар тілдің мысалы болып табылады. Бұл қазақ мәтінін өңдеу мен талдауда қиындықтар туғызады, сондай-ақ нәтижелердің жоғары сапасы мен дәлдігіне қол жеткізу үшін арнайы модельдеу әдістерін талап етеді.

Соңғы жылдары тілдік модельдеу саласында күрт ілгерілеу байқалады, бұл қазақ тілін қоса алғанда, күрделі тілдік құрылымдарды тиімді өңдеу үшін жаңа мүмкіндіктер ашты. Ең сәтті және кеңінен қолданылатын тәсілдердің бірі — мәтіндегі сөздер тізбегін статистикалық талдауға негізделген n-грамм модельдерін қолдану (Браун, 1992). Бұл модельдер алдыңғы сөздерге негізделген келесі сөздің контексті мен ықтималдығын ескеруге мүмкіндік береді (Хайруллина, 2018).

Алайда, n-грамм модельдерінің кемшілігі – олардың сөздер арасындағы ұзақ мерзімді тәуелділіктер мен семантикалық қатынастарды байланыстыру қабілетінің шектеулі болуы (Поречный, 2020). Соңғы жылдары терең оқыту тілдік модельдеуде үлкен жетістік болды, бұл күрделі және қуатты модельдер жасауға мүмкіндік берді. BERT (Bidirectional Encoder Representations from Transformers) моделі ерекше көзге түседі, ол Transformer архитектурасына негізделген және мәтіндегі ұзақ мерзімді тәуелділіктер мен контекстік ерекшеліктерді анықтай алады (Салып, 2022). BERT мәтіндік деректердің үлкен көлемінде өзін-өзі оқыту механизмін қолданады, бұл модельге әртүрлі тілдік құрылымдар мен олардың контекстік ерекшеліктерімен байытуға мүмкіндік береді. Бұл BERT-ті күрделі морфологиялық құрылымы бар тілдер үшін әсіресе пайдалы етеді. BERT тілдік модельдеу нәтижелерін айтарлықтай жақсартқанымен, оның шектеулері бар. Ең алдымен, оны оқыту үлкен есептеу ресурстары мен уақытты қажет етеді. Сонымен қатар, BERT сөйлемдегі барлық сөздер бір-бірінен тәуелсіз деп болжайды, бұл сөздер бір-бірімен тығыз байланысатын бай морфологиясы бар тілдер үшін шектеу болуы мүмкін.

Тағы бір перспективалық бағыт – GPT сияқты генеративті модельдерді тілдік модельдеуге қолдану (Частикова, 2022). GPT сонымен қатар Transformer архитектурасына негізделген және оның контексті мен семантикалық қатынастары мен мағыналарын ескере отырып, мәтін құруға қабілетті. Бұл модель шектеулі мәтіндік деректері бар тілдерде мәтінді талдау мен құрудың тиімді құралы бола алады.

Жұмыстың мақсаты — тілдік модельдеу әдістеріне шолу және салыстырмалы талдау, сонымен қатар оларды қазақ тілі сияқты күрделі морфологиялық құрылымы бар тілге қолдану. Сонымен қатар, әр әдістің артықшылықтары мен шектеулері келтірілген.

Мақала келесідей ұйымдастырылған: 2-бөлімде тілдік модельдеу саласындағы қолданыстағы жұмыстарға шолу жасалды және оларды күрделі морфологиялық құрылымы бар тілдерге қолдану келтірілген. 3-бөлім тілдік модельдеу әдістерін, соның ішінде n-грамм модельдерін, BERT, GPT және басқаларын, олардың негізгі жұмыс принциптері мен артықшылықтарын егжей-тегжейлі сипаттауға арналған. 4-бөлімде тәжірибелер мен зерттеулердің нәтижелері жасалған, ал қорытындыда зерттеу нәтижелері мен осы бағыттағы болашақ жұмыстардың қорытындысы келтірілген.

### **Әдістер мен материалдар**

#### *n-грамм моделі*

n-грамм моделі — бұл n-1 сөзден кейінгі келесі сөзді олардың жұптасу ықтималдығына негізделген болжайтын статистикалық модельдер. Мәселен, қазақ тіліндегі «Мен Алматыға бара жатырмын» біріктірудің ықтималдығы жоғары, ал «Мен бара Алматыға жатырмын» біріктірудің ықтималдығы төмен. Қарапайым тілмен айтқанда, n-грамм – n сөздердің тізбегі. Мысалы, биграммалар – екі сөзден тұратын тізбектер (Мен Алматыға, Алматыға бара, баа жатырмын), триграммалар – үш сөзден тұратын тізбектер (мен Алматыға бара, Алматыға баа жатырмын) және т.б.

Мұндай ықтималдық үлестірімдері машиналық аудармада, орфографияны автоматты түрде тексеруде, сөйлеуді тануда және автоматты түрде енгізуде кеңінен қолданылады. Барлық жағдайларда біз келесі сөздің немесе сөз тізбегінің ықтималдығын есептейміз. Мұндай есептеулер тілдік модельдер деп аталады.

Сөйлемдегі сөздер санына байланысты  $P(w)$  есептеудің жалпы формуласы келесі түрде болады (1):

$$P(w_1, \dots, w_n) = P(w_n | w_{n-1}, \dots, w_1) \cdot P(w_{n-1}, \dots, w_1) \quad (1)$$

мұнда  $w$  – жеке сөздер,  $n$  – сөздер саны.

Осылайша, шартты ықтималдықтарды көбейту арқылы бүкіл тізбектің бірлескен ықтималдығын бағалай аламыз. Алайда, алдыңғы сөздердің ұзақ тізбегі жағдайында сөздің нақты ықтималдығын есептей алмаймыз, өйткені мүмкін болатын тізбектер өте көп және деректерімізде бұл тізбектер



болмауы мүмкін. Сондықтан, барлық алдыңғы сөздерді ескере отырып, сөздің ықтималдығын есептеудің орнына, біз оны жеңілдету арқылы ықтималдылықты жуықтай аламыз. Бұл Марков тізбектерінің негізі, оның көмегімен біз тым кең контекстті ескермей, реттілік элементінің ықтималдығын болжай аламыз (2):

$$P(w_n | w_{n-1}, \dots, w_1) \approx P(w_n | w_{n-1}, \dots, w_{n-k}) \quad (2)$$

мұнда  $k$  – Марков тізбегінің реті.

Бірінші ретті Марков тізбегін қолдана отырып, кез келген сөз тізбегінің ықтималдығын оңай есептеуге болады. Осылайша, біз биграммаларды, триграммаларды, квадрограммаларды және т.б. есептей аламыз, ал тізбек неғұрлым ұзағырақ болса, біздің модель соғұрлым егжей-тегжейлі болады, яғни ұзын сөйлемдер қысқа сөйлемдерге қарағанда көбірек грамматиканы қамтиды.

n-грамм моделі салыстырмалы түрде қарапайым және тиімді, бірақ олар дәйектіліктегі сөздердің ұзақ мерзімді контекстін ескермейді.

*Лесне жұмыстар*

Жұмыс авторлары (Си, 2021) n-грамм модельдерінің кемшіліктерін едәуір азайтқан n-distant-max тілдік моделін ұсынды. Контекстегі орфография қателерін түзету жүйесін бағалау кезінде салыстырмалы талдау жүргізілгеннен кейін, әзірленген модель n-грамм тілінің классикалық моделімен салыстырғанда осы жүйенің тиімділігін едәуір арттырғаны көрсетілді.

C. Shelba және басқалары (Челба, 2017) n-грамм модельдеу үшін RNN модельдерінің жадының қасиетін қолдана отырып тиімді тереңдігін зерттеді. Шағын UPenn Treebank корпустың тәжірибелер LSTM ұяшығы RNN тікелей байланысымен салыстырғанда n-gram күй деректерін кодтау үшін жақсы модель екенін көрсетті. Сөйлемнің тәуелсіздігі туралы болжамды сақтай отырып, LSTM n-граммасы  $n=9$  үшін LSTM LM өнімділігіне сәйкес келді және  $n=13$  кезінде одан сәл асып түсті.

Зерттеушілер (Ким, 2022) агглютинативті тілдер тобына кіретін корей тілін модельдеуді, атап айтқанда, тілді бейнелеу әдістері мен алдын ала оқыту әдістерін зерттеді. Авторлар кеңінен қолданылатын Transformer архитектурасы мен екі жақты тіл көрінісіне негізделген корей тілін фрагментті түрде қайта құруды ұсынады. Олар алдын-ала оқыту кезінде осындай ақпаратты қолдана отырып, тілді түсінуге сөйлеу бөлігі (сөйлеу бөлігі, PoS) сияқты морфологиялық ерекшеліктерді енгізді. Жұмыстың алынған нәтижелері ұсынылған әдістер корей тілін түсінудің зерттелетін міндеттерінің модельдік өнімділігін жақсартатынын көрсетті.

Myrzakhmetov (Мырзахметов, 2018) қазақ тіліне арналған кеңейтілген тілдік модельдеу экспериментінде веб-газеттердегі, сондай-ақ қазақ тіліндегі басқа сайттардағы әртүрлі мақалалардан тұратын қазақ тіліне арналған корпус ұсынды. Дәстүрлі n-грамм модельдерімен бірге олар вербальды тіл моделі

(LM) үшін нейрондық желі модельдерін жасады. Үлкен параметрленген ұзақ қысқа мерзімді жад (LSTM) моделі ең жақсы өнімділікті көрсетті. Сонымен қатар, зерттеушілер LM негізіндегі морфемаларды пайдаланды. Эксперименттер субсөзге негізделген LM қазақ тілі үшін сөз негізіндегі LM-мен салыстырғанда n-грамм және нейрондық желі модельдерінде жақсы жұмыс істейтінін көрсетті.

### *BERT*

BERT – бұл Google-дің нейрондық желісі, ол көптеген міндеттер бойынша state-of-the-art нәтижелерін үлкен айырмашылықпен көрсетті. BERT көмегімен табиғи тілді өңдеуге арналған жасанды интеллект (AI) бағдарламаларын жасауға болады: еркін түрде қойылған сұрақтарға жауап беру, чатботтар, автоматты аудармашылар құру, мәтінді талдау және т.б. (Девлин, 2018).

Тілдік модельдеуге арналған бірнеше назар аудару механизмдері бар BERT моделіне екі бағытты оқытуы бар Transformer архитектурасы енгізілді (Васвани, 2017). Тілдік модель солдан оңға және оңнан солға қарай екі бағытта оқыту кезінде сөйлемнің контексті мен семантикасын толығымен қабылдайды.

BERT контекстен жасырылған сөзді болжауға үйретілген (1-сурет) және екі сөйлемнің дәйекті немесе сәйкестігін жіктеуге анықтауға жасалған. Кіріс деректері алдымен векторларға енгізілген, содан кейін нейрондық желіде өңделетін токендер тізбегі болып табылады. Шығару векторлар тізбегі болып табылады, онда әрбір вектор бірдей индексі бар кіріс лексемасына сәйкес келеді. BERT-ке сөздер тізбегін бермес бұрын, әрбір тізбектегі сөздердің 10%-ы [маска] белгісімен ауыстырылады. Содан кейін модель басқа, жасырын емес сөздермен берілген контекст негізінде бүркеніш сөздердің бастапқы мағынасын болжауға тырысады.

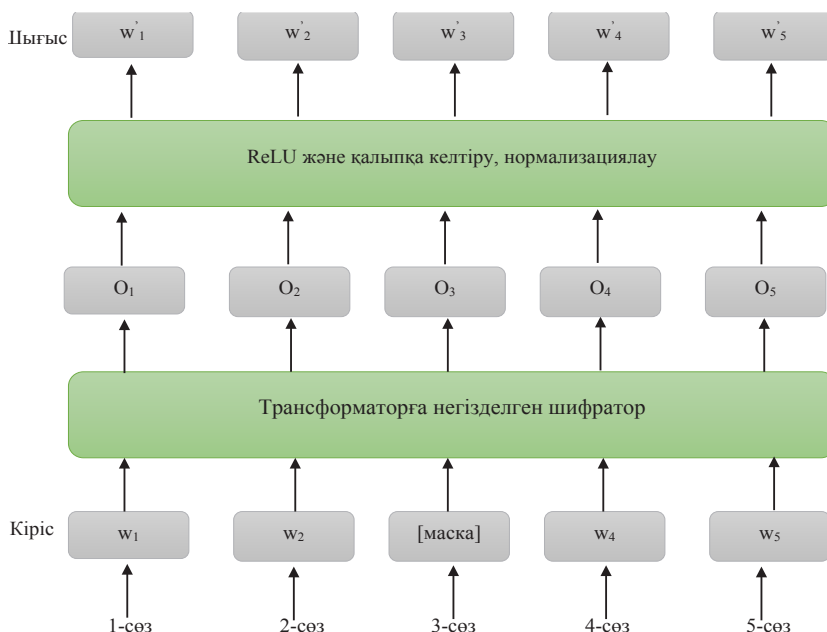
BERT GLUE тестіндегі он бір NLP тапсырмасы бойынша алдыңғы заманауи модельдерден едәуір асып түседі. Бұл керемет нәтиже BERT тіл туралы құрылымдық ақпаратты жақсы байланыстыра алады деп болжайды. Маңызды даму қадамы Голдберг болып табылады, ол BERT субъекті-егістік келісімін қадағалау қабілетін бағалау кезінде синтаксистік құбылыстарды жақсы сәйкестендіретіндігін көрсетеді.

### *Ілеспе жұмыстар*

Авторлар (Чандра, 2021) 2020 жылы АҚШ президенттік сайлауы қарсаңында Twitter-дегі көңіл-күйді талдау үшін LSTM және BERT тілдік модельдерін қолданды. Олардың зерттеулері әлеуметтік медианы талқылау сайлау кезінде көпшіліктің мінез-құлқы мен көзқарасын түсінуге көмектесетінін анықтады. Президенттік сайлауға байланысты шамамен 1,2 миллион твиттер талданды. Модельдеу мен талдаудан кейін көңіл-күйді талдау сайлау нәтижелерін модельдеудің жалпы негізі бола алатынын анықтады. BERT моделі Байденнің сайлау науқандары кезінде Twitter-ге сүйене отырып жеңіске жету мүмкіндігі жоғары екенін көрсетті. Олар BERT моделі Трампты, Байденді және даулы мемлекеттерді анықтауда дәл болды деп есептеді. Демек, көбірек деректер

мен географиялық ақпаратпен көңіл-күйді талдау сайлау нәтижелерін болжау үшін пайдалы болу мүмкіндігін дәлелдеді.

Ganesh және басқалар (Ганеш, 2019) BERT фразалық көрінісі ақпаратты төменгі қабаттардағы фразалар деңгейінде түсіретінін, сондай-ақ BERT аралық қабаттары лингвистикалық ақпараттың бай иерархиясын төменгі жағында беттік функциялармен, ортасында синтаксистік функциялармен және жоғарғы жағында семантикалық функциялармен кодтайтынын көрсетті. Мысалы, алыс қашықтықтағы тәуелділіктер туралы ақпарат қажет болған кезде BERT тереңірек деңгейлерді қажет етті. Сонымен қатар, зерттеушілер BERT лингвистикалық ақпаратты классикалық ағаш құрылымдарына еліктейтін композициялық түрде жинайтынын анықтады.



Сурет 1 – BERT жалпы архитектурасы

Авторлар (Ли, 2020) корей KR-BERT моделін кішірек сөздік пен деректер жиынтығын пайдалана отырып оқытты. Корей тілі латын графикасын пайдаланбайтын морфологиялық ресурсы аз тілдердің бірі болғандықтан, авторлар көптілді BERT моделі жіберіп алған тілге тән тілдік құбылыстарды түсірудің маңыздылығын атап өтті. Олар бірнеше токенизаторларды, соның ішінде әзірленген WordPiece қос бағытты токенизаторыын сынады және модельдері үшін жақсы сөздік қорын жасау үшін ішкі таңбадан таңба деңгейіне дейін токендеу үшін ең аз диапазонды реттеді. Осы түзетулермен әзірленген KR-BERT моделі шамамен 1/10 өлшемді корпусты пайдаланатын бұрыннан белгілі басқа модельдерге қарағанда салыстырмалы және тіпті жақсырақ орындады.

Контексте болжам жасау үшін тілдік модельдер қолданатын ақпарат туралы мақсатты сұрақтар қоюға мүмкіндік беретін табиғи тілдің эксперименттеріне негізделген диагностика жиынтығын ұсынылды (Эттингер, 2020). Мысал ретінде олар диагнозды BERT моделіне қолданды және ол әдетте адамдарға қарағанда сезімталдығы төмен болса да, жалпы санатқа немесе рөлдердің өзгеруіне байланысты жақсы және жаман аяқталуларды ажырата алатынын және зат есімдердің гипернимдерін сенімді түрде шығаратынын анықтады, бірақ рөлдерге негізделген оқиғалар туралы қорытындылар мен болжамдармен күреседі және сынайды — және, атап айтқанда, теріске шығарудың контекстік әсеріне айқын сезімталдықты көрсетеді. Осылайша, бұл жұмыста BERT моделінің негізгі кемшіліктері анықталды.

Зерттеу жұмыстың авторлары (Ли, 2020) алдын ала дайындалған BERT моделін дәл баптауға және оны патенттік жіктеуге қолдануға бағытталған. Екі миллионнан астам патенттері бар үлкен деректер жиынтығына қатысты ұсынылған тәсіл CNN сөз ендірілген тәсілінің арқасында заманауи тәсілден асып түседі. Осылайша келесі жақсартулар алынды: 1) алдын-ала дайындалған BERT моделіне және патенттерді жіктеуге арналған дәл баптауға негізделген жаңа заманауи нәтиже, 2) патенттік талаптардың өзі жіктеу тапсырмасы үшін қазіргі заманғы нәтижелерге қол жеткізу үшін жеткілікті екенін көрсету.

### *GPT*

OpenAI GPT модельдерінің арқасында табиғи тілді өңдеуде айтарлықтай жетістіктерге жетті. GPT-1-ден GPT-4-ке дейін бұл модельдер проза мен поэзияны құрудан чатботтарға, тіпті кодтауға дейін AI құрған мазмұнның алдыңғы қатарында болды.

Генеративті алдын-ала дайындалған түрлендіргіштер (GPT) — бұл табиғи тілді өңдеу тапсырмалары үшін қолданылатын машиналық оқыту моделінің түрі. Бұл модельдер контекстке сәйкес және семантикалық тұрғыдан біртұтас тіл құру үшін кітаптар мен веб-беттер сияқты үлкен көлемде алдын ала оқытылады.

Басқаша айтқанда, GPT – бұл адам тәрізді мәтінді нақты бағдарламалаусыз жасай алатын компьютерлік бағдарламалар. Нәтижесінде оларды сұрақтарға жауап беру, тілдік аударма және мәтінді жалпылауды қоса алғанда, табиғи тілді өңдеудің бірқатар мәселелерін шешу үшін дәл реттеуге болады. Сонымен қатар, бұл модель табиғи тілді өңдеудегі үлкен жетістік, бұл машиналарға тілді бұрын-соңды болмаған еркін және дәлдікпен түсінуге мүмкіндік береді.

*GPT-1.* GPT-1 OpenAI Transformer архитектурасын қолдана отырып, тілдік модельдің алғашқы итерациясы ретінде шығарылды. Оның 117 миллион параметрі бар, бұл алдыңғы заманауи тіл модельдерін айтарлықтай жақсартты. Осы модельдің артықшылығы — оның анықтамасы немесе контексті болған кезде тегіс және біртұтас тіл жасау қабілеті болды. Модель екі деректер жиынтығын біріктіруге үйретілді: Common Crawl, миллиардтаған сөздері бар веб-беттердің үлкен деректер жинағы және bookcorpus деректер жинағы, әртүрлі жанрдағы 11 000-нан астам кітаптар жинағы. Осы әртүрлі деректер

жиынтығын пайдалану GPT-1-ге тілдік модельдеудің күшті қабілеттерін дамытуға мүмкіндік берді.

GPT-1 табиғи тілді өңдеудегі (NLP) маңызды жетістік болғанымен, оның белгілі бір шектеулері бар: модель қайталанатын мәтінді жасауға бейім болды, бұл оған оқыту деректерінен тыс ақпарат берілген кезде айқын көрінді. Ол сондай-ақ мәтіндегі ұзақ мерзімді тәуелділіктерді бақылай алмайды. Сонымен қатар, оның үйлесімділігі мен еркіндігі тек қысқа мәтіндік тізбектермен шектелді, ал ұзағырақ үзінділерде байланыс болмады.

Осы шектеулерге қарамастан, GPT-1 Transformer архитектурасына негізделген үлкен және қуатты модельдердің негізін қалады.

*GPT-2.* Оның құрамында 1,5 миллиард параметр болды. Модель Common Crawl және Web Text біріктіретін әлдеқайда үлкен және әртүрлі деректер жиынтығында оқытылды. GPT-2-нің күшті жақтарының бірі – оның мәтіннің дәйекті және шынайы тізбегін құру қабілеті. Сонымен қатар, ол адам сияқты жауаптар жасай алады, бұл оны мазмұнды құру және аудару сияқты табиғи тілді өңдеудің әртүрлі тапсырмалары үшін құнды құралға айналдырады. Алайда, GPT-2-нің кемшіліктері бар. Ол күрделі ойлау мен контекстті түсінуді қажет ететін міндеттермен басқара алмайды. GPT2 мәтіннің қысқа абзацтары мен үзінділерінде жақсы жұмыс істегенімен, ол ұзағырақ үзінділерде контекст пен келісімді сақтай алмайды.

*GPT-3.* GPT-3 BookCorpus, Common Crawl және Wikipedia сияқты әртүрлі деректер көздерінде оқытылады. Деректер жиынтығында триллионға жуық сөз бар, бұл GPT-3-ке NLP тапсырмаларының кең спектріне күрделі жауаптар жасауға мүмкіндік береді. Алдыңғы модельдермен салыстырғанда GPT-3-тің негізгі жақсартуларының бірі –үйлесімді мәтін құру, компьютерлік код жазу, тіпті өнер туындыларын жасау мүмкіндігі. Алдыңғы модельдерден айырмашылығы, GPT-3 берілген мәтіннің мәнмәтінін түсінеді және тиісті жауаптар генерациялайды. Табиғи дыбыстық мәтін құру мүмкіндігі чатботтар, мазмұн жасау және тілдік аударма сияқты қосымшалар үшін өте маңызды. Осындай мысалдардың бірі – ChatGPT, жасанды интеллект диалогтық боты.

GPT-3-де дегенмен өз кемшіліктері жеткілікті. Модель біржақты, дәл емес немесе орынсыз жауаптарды қайтара алады. Бұл мәселе GPT-3 жалған ақпаратты қамтуы мүмкін мәтіннің көп мөлшерінде оқытылатындықтан туындайды. Сондай-ақ, модель контекст пен фондық білімді түсінуде әлі де қиындықтарға тап болғанын көрсетеді және кейде мүлдем маңызды емес мәтінді жасайтын кездері болады.

*GPT-4.* GMT-4 Тек ChatGPT Plus пайдаланушыларына арналған, бірақ пайдалану шегі шектеулі. GPT-4-тің көрнекті ерекшелігі – оның мультимодальды мүмкіндіктері. Яғни модель енді кескінді кіріс ретінде қабылдай алады және оны мәтіндік формат ретінде түсінеді. Модель сонымен қатар күрделі мәтіндерді жақсы түсінеді, бірнеше кәсіби және дәстүрлі сынақтарда адам деңгейіндегі өнімділікті көрсетеді.

*MT-NLG.* MT-NLG (Megatron-Turing Natural Language Generation) — бұл

Transformer архитектурасына негізделген қуатты және жетілдірілген тілдік модель. Ол табиғи тілде көптеген тапсырмаларды орындай алады, соның ішінде логикалық тұжырымдар мен оқуды түсіну. Бұл Microsoft және Nvidia әзірлеген тілдік модельдердің соңғы нұсқасы және сөйлемдерді автоматты түрде аяқтау, ақылға қонымды пайымдауды түсіну және оқуды түсіну сияқты көптеген нәрселерді жасай алады. Модель көптеген деректерде оқытылды, атап айтқанда ағылшын тілді веб-сайттардан барлығы 339 миллиард токендерден (сөзден) тұратын 15 деректер жиынтығы. MT-NLG – бұл жаңадан жасалған модель, сондықтан ол үшін нақты пайдалану жағдайлары әлі де толығымен зерттелмеген. Дегенмен, модельді жасаушылар бұл табиғи тілді өңдеу технологиялары мен өнімдерінің болашағын анықтай алады деп болжады.

*LaMDA.* LaMDA – Google әзірлеген диалогтық қосымшаларға арналған тілдік модель. Ол ауызша диалогты еркін түрде құруға арналған, бұл оны әдетте тапсырмаларға негізделген дәстүрлі модельдерге қарағанда табиғи және егжей-тегжейлі етеді. LaMDA 137 миллиард параметрлері бар диалогтық деректерде оқытылды. Бұл оған ашық әңгіменің нюанстарын түсінуге мүмкіндік береді. Google бұл модельді іздеу, Google Assistant және Workspace сияқты өнімдерінде қолдануды жоспарлап отыр. MT-SNG сияқты, бұл жаңадан жасалған модель, сондықтан осы модельді қолдану бойынша зерттеулер өте аз.

Тілдік модельдеудің тағы бір маңызды әдісі – ұзақ, қысқа мерзімді жады бар қайталанатын нейрондық желілерді пайдалану (LSTM - Long Short-Term Memory). LSTM-мәтіндегі ұзақ мерзімді тәуелділіктерді тиімді модельдеуге қабілетті нейрондық желі архитектурасы. LSTM ақпаратты ұзақ уақыт бойы сақтауға және жаңартуға мүмкіндік беретін арнайы жад механизмін пайдаланады []. Бұл әсіресе морфологиялық өзгерістер мен тәуелділіктер айтарлықтай қашықтыққа ие болуы мүмкін мәтіндерді өңдеу кезінде пайдалы. LSTM-нің басты артықшылықтарының бірі – оның ұзақ мерзімді тәуелділіктерді модельдеу және алдыңғы контексттердегі ақпаратты есте сақтау қабілеті. Бұл модельге мәтіндегі келесі сөзді оның контексті мен семантикасын ескере отырып дәлірек болжауға мүмкіндік береді. LSTM модельдер әртүрлі тілдік құрылымдар мен морфологиялық ерекшеліктерді алуға мүмкіндік беретін мәтіндік деректердің үлкен көлемінде оқытылуы мүмкін.

### **Нәтижелер және оларды талқылау**

*Корпус және тәжірибелер.* Сөз тізбегінің ықтималдығын болжауға арналған нейрондық тіл модельдері әдетте мәтіндік деректердің үлкен корпусында оқытылады және тілдің негізгі құрылымын үйренуге қабілетті. Ол үшін қазақ тіліндегі электронды көркем кітаптарды жинау туралы шешім қабылданды. tilalemi.kz сайтынан қазақ тілінде *txt* форматында 30 электронды кітап жүктелді. Модельді оқыту үшін сәйкесінше 24 кітап, ал тексеру және тестілеу үшін 3 кітаптан ақпарат қолданылды (кесте 1).

1 кесте. Модельді оқыту үшін мәтіндік корпусты бөлу

Дереккөз	tilalemi.kz		
	Кітаптар саны	Сөйлемдер саны	Сөз саны
Оқытуға арналған жиынтығы	24	131369	1434149
Валидацияға арналған жиынтық	3	22637	238601
Тестілеуге арналған жиынтық	3	8168	89853

### Деректерді алдын ала өңдеу

Модельдерді оқытпас бұрын деректерді алдын ала өңдеу сөйлемдердің басталуы мен аяқталуын токендеу, тазарту, өңдеу және тегістеу әдістерін қолдану сияқты бірнеше қадамдарды қамтиды.

Токенизация — мәтінді жеке токендерге немесе сөздерге бөлу процесі. n-грамм моделі жағдайында әр сөз жеке токенге айналады. Токенизация *nlk* табиғи тілді өңдеу кітапханасы арқылы жүзеге асырылды.

Токенизациядан кейін деректерді қажетсіз таңбалардан және арнайы таңбалардан тазарту жүргізілді. Сондай-ақ, модель үшін семантикалық жүкте-месі жоқ барлық тыныс белгілері, сандар мен таңбалар алынып тасталды.

BERT моделі үшін келесі параметрлер анықталды: 5 эпоха саны, пакеттің өлшемі 8, оңтайландырғыш ретінде AdamW алгоритмі белгіленді. Модельдерді қайта өңдеуден және оқытудан кейін тест жиынтығында нәтижелер алынды және 2-кестеде келтірілген (сурет. 2).

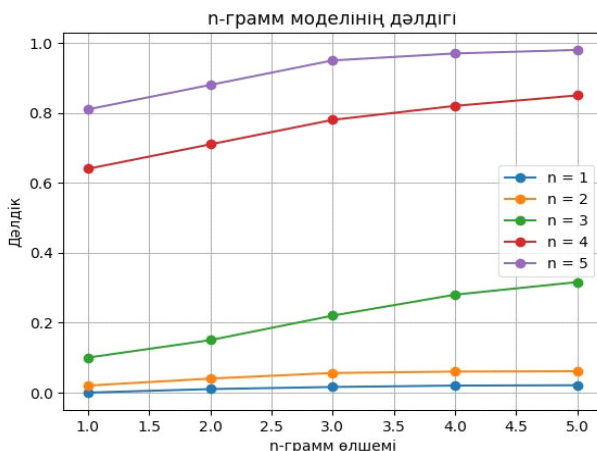
2 кесте. n-грамм және BERT модельдерінің нәтижелері

Модельдер	Болжау дәлдігі, %
n-грамм моделі	
n = 1	1,6
n = 2	5,6
n = 3	31
n = 4	84,6
n = 5	97,7
BERT моделі	98,9

### Алынған нәтижелерді талқылау

Эксперименттік тапсырмаларды орындау барысында n-грамм және BERT модельдері құрылды және оларды қазақ тіліне арналған келесі сөзді болжау жұмыстары іске асырылды. Екі модель де токенизацияны қолданады, ал BERT токенизатор жаттығу және сынақ датасеттерін токенизациялау үшін қолданылады, содан кейін оларды жаттығу кезінде қолданылатын PyTorch форматына түрлендіреді.

Тек n = 4 және n = 5 кезінде модель максимумға жетеді, бірақ n-грамм санының ұлғаюы сонымен қатар үлкен есептеу ресурстарын қажет етті (сур.2), ал модельді оқыту мен тестілеу BERT моделімен салыстырғанда көп уақытты қажет етпеді.



Сурет 2 – n-грамм моделін болжау дәлдігі

BERT моделі бойынша нәтиже алу үшін 2 күндей уақыт кетті, бірақ нәтижесі жақсы болып шықты, n-грамм моделін 1,2 %-ға дәлірек болжады.

n-грамм тіл модельдерінен айырмашылығы, BERT контекстке сезімтал көріністерді үйретеді. Мәтінмәнді ескеретін модельдер әдетте токеннің сол немесе оң контекстін ғана есепке алуға мүмкіндік береді. BERT, керісінше, екі жақты контекстті ескереді, бұл модельге көп мағыналы сөздердің мағынасын жақсы түсінуге көмектеседі.

### Қорытынды

Бұл ғылыми мақалада тілдік модельдеудің заманауи әдістері және олардың күрделі морфологиялық құрылымы бар тілдерге қолданылуы зерттелді. Негізгі фокус n-грамм, BERT модельдерін және олардың агглютинативті сипаттағы тілдерді тиімді модельдеу қабілетін зерттеу болды.

Зерттеу нәтижелері BERT моделі қазақ тіліне арналған сөздер арасындағы контекстік тәуелділіктер мен семантикалық қатынастарды түсінуде жоғары дәлдікті көрсететінін анықтады. Оның контекстті екі бағытта модельдеу және назар аудару механизмдерін қолдану қабілеті мәтіндегі ұзақ мерзімді тәуелділіктер мен күрделі қатынастарды түсінуге мүмкіндік береді. Зерттеудің бұл нәтижелері күрделі морфологиялық құрылымы бар тілдер үшін заманауи тілдік модельдеу әдістерін, әсіресе BERT модельдерін қолданудың маңыздылығы мен перспективаларын растайды. Мұндай модельдер мәтінді автоматты түрде өңдеу сапасын едәуір жақсарты алады, табиғи тілді өңдеудің әртүрлі міндеттерінде дәлдік пен сенімділікті арттырады.

Алдағы жұмыстарда NLP-дің басқа да перспективалық әдістері негізінде түрлі эксперименттік жұмыстар жүзеге асырылатын болады.

### ӘДЕБИЕТТЕР

Васвани и др. (2017). Внимание — это все, что вам нужно. В материалах 31-й конференции по нейронным системам обработки информации (NeurIPS 2017), Лонг-Бич, Калифорния, США, 2017 (in Eng.)



Ганеш Джавахар, Бенуа Саго, Джаме Седда. (2019). Что BERT узнает о структуре языка? ACL 2019 — 57-е ежегодное собрание Ассоциации компьютерной лингвистики, июль 2019 г., Флоренция, Италия. (in Eng.)

Девлин Джейкоб, Минг-Вэй Чанг, Кентон Ли и Кристина Тутанова. (2019). BERT: Предварительная подготовка глубоких двунаправленных преобразователей для понимания языка. В материалах конференции Североамериканского отделения Ассоциации компьютерной лингвистики 2019 года: технологии человеческого языка, Т. 1 (Длинные и короткие статьи), стр. 4171–4186, Миннеаполис, Миннесота. (in Eng.)

Ли Санга, Чан Хансоль, Байк Юнми, Пак Сюзи, Шин Хёпиль. (2020). Малая языковая модель BERT, специфичная для корейского языка. Журнал KIISE. 47. 682–692. 10.5626/JOК.2020.47.7.682 (in Eng.)

Мырзахметов Б. и Кожирбаев З. (2018). Расширенные эксперименты по языковому моделированию казахского языка. Материалы семинара CEUR, 2303 (in Eng.)

Браун Питер Ф., Винсент Дж. Делла Пьетра, Питер В. де Соуза, Дженифер К. Лай, Роберт Л. Мерсер. (1992). n-граммные модели естественного языка на основе классов, Компьютерная лингвистика, Том. 18, стр. 467–479. (in Eng.)

Поречный А.С. (2020). Создание инструмента семантико-синтаксического анализа текстов на русском языке. Современные информационные технологии и ИТ-образование, DOI: 10.17308/sait.2020.1/2630.

Салып Б., Смирнов А. (2022). Анализ модели BERT как инструмента определения меры смысловой близости предложений естественного языка. StudNet, №5, URL: <https://cyberleninka.ru/article/n/analiz-modeli-bert-kak-instrumenta-opredeleniya-mery-smyslovoy-blizosti-predlozheniy-estestvennogo-yazyka>.

Си Луссейн Аурага, Абделла Юсфиб, Саида Лааруссик, Хишам Гuedдахд, Мохаммед Неджа. (2021). Новая оценка языковой модели n-грамм. В материалах 5-й Международной конференции по искусственному интеллекту в компьютерной лингвистике, стр. 211–215. DOI: 10.1016/j.procs.2021.05.111 (in Eng.)

Ким Х., Ким С., Кан И., Квак Н., Фунг П. (2022). Моделирование корейского языка с помощью синтаксического руководства. В материалах тринадцатой конференции по языковым ресурсам и оценке, стр. 2841–2849, Марсель, Франция. Европейская ассоциация языковых ресурсов (in Eng.)

Хайруллина Р.Х., Рахимова Э.Ф., Сагитова А.Ф. (2018). Лингвокогнитивные основы языкового моделирования. Мир науки, культуры, образования, №3 (70), URL: <https://cyberleninka.ru/article/n/lingvokognitivnye-osnovy-yazykovogo-modelirovaniya>.

Ли Цзе-Шэн, Цзе Сян. (2020). Патентная классификация путем тонкой настройки языковой модели BERT. Мировая патентная информация, Т. 61, 101965. DOI: 10.1016/j.wpi.2020.101965 (in Eng.)

Чандра Р., Сайни Р. (2021). Байден против Трампа: моделирование всеобщих выборов в США с использованием языковой модели BERT, в IEEE Access, Т. 9, стр. 128494–128505, 2021, doi: 10.1109/ACCESS.2021.3111035 (in Eng.)

Частикова В.А., Гуляй В.Г., Жерлицын С.А. (2022) Подход к решению проблемы контроля качества в сфере услуг на основе построения системы интеллектуального анализа данных." Вестник Адыгейского государственного университета. Серия 4: Естественно-математические и технические науки, №4 (311). URL: <https://cyberleninka.ru/article/n/podhod-k-resheniyu-problemy-kontrolya-kachestva-v-sfere-uslug-na-osnove-postroeniya-sistemy-intellektualnogo-analiza-dannyh>.

Челба К., Норузи М., Бенджио С. (2017). Моделирование языка N-грамм с использованием рекуррентной оценки нейронной сети. ArXiv, abs/1703.10724 (in Eng.)

Этингер Э. (2020). Чем не является BERT: уроки нового набора психолингвистической диагностики языковых моделей. Труды Ассоциации компьютерной лингвистики, 2020, 8, стр. 34–48. DOI: 10.1162/tacl\_a\_00298 (in Eng.)

## REFERENCES

Allyson Ettinger. (2020). What BERT is not: Lessons from a new Suite of Psycholinguistic Diagnostics for Language Models. *Transactions of the Association for Computational Linguistics*, 2020, 8. Pp. 34–48. DOI: 10.1162/tacl\_a\_00298 (in Eng.)

Chastikova V.A., Gulyai V.G., Zherlitsyn S.A. (2022). Podkhod k resheniyu problemy kontrolya kachestva v sfere uslug na osnove postroeniya sistemy intellektualnogo analiza dannykh." *Vestnik Adygeiskogo gosudarstvennogo universiteta. Seriya 4: Estestvenno-matematicheskie i tekhnicheskie nauki*, №4 (311). URL: <https://cyberleninka.ru/article/n/podhod-k-resheniyu-problemy-kontrolya-kachestva-v-sfere-uslug-na-osnove-postroeniya-sistemy-intellektualnogo-analiza-dannyh> (in Rus.)

Chelba C., Norouzi M., Bengio S. (2017). N-gram Language Modeling using Recurrent Neural Network Estimation. *ArXiv*, abs/1703.10724 (in Eng.)

Chandra R., Saini R. (2021). Biden vs Trump: Modeling US General Elections Using BERT Language Model, in *IEEE Access*. Vol. 9. Pp. 128494–128505, 2021, doi: 10.1109/ACCESS.2021.3111035 (in Eng.)

Ganesh Jawahar, Benoît Sagot, Djamé Seddah. (2019). What does BERT learn about the structure of language? In *ACL 2019 - 57th Annual Meeting of the Association for Computational Linguistics*, Jul 2019, Florence, Italy (in Eng.)

Hyeondey Kim, Seonhoon Kim, Inho Kang, Nojun Kwak, and Pascale Fung. (2022). Korean Language Modeling via Syntactic Guide. In *Proceedings of the Thirteenth Language Resources and Evaluation Conference*. Pp. 2841–2849, Marseille, France. European Language Resources Association (in Eng.)

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, Vol. 1 (Long and Short Papers). Pp. 4171–4186, Minneapolis, Minnesota (in Eng.)

Jieh-Sheng Lee, Jieh Hsiang. (2020). Patent classification by fine-tuning BERT language model. *World Patent Information*, Vol. 61. 101965. DOI: 10.1016/j.wpi.2020.101965 (in Eng.)

Khairullina R.Kh., Rakhimova E.F., Sagitova A.F. (2018). Lingvokognitivnye osnovy yazykovogo modelirovaniya. *Mir nauki, kultury, obrazovaniya*, №3 (70), URL: <https://cyberleninka.ru/article/n/lingvokognitivnye-osnovy-yazykovogo-modelirovaniya> (in Rus.)

Lee Sangah, Jang Hansol, Baik Yunmee, Park Suzi, Shin Hyopil. (2020). A Small-Scale Korean-Specific BERT Language Model. *Journal of KIISE*. 47. 682–692. 10.5626/JOK.2020.47.7.682 (in Eng.)

Myrzakhmetov B. & Kozhirbayev Z. (2018). Extended language modeling experiments for Kazakh. *CEUR Workshop Proceedings*, 2303 (in Eng.)

Peter F. Brown, Vincent J. Della Pietra, Peter V. deSouza, Jenifer C. Lai, Robert L. Mercer. (1992). Class-based n-gram models of natural language, *Computational Linguistics*. Vol. 18. Iss. 4. Pp. 467–479 (in Eng.)

Porechnyi A.S. (2020). Sozdanie instrumenta semantiko-sintaksicheskogo analiza tekstov na russkom yazyke. *Sovremennye informatsionnye tekhnologii i IT-obrazovanie*. DOI: 10.17308/sait.2020.1/2630 (in Rus.)

Salyp B., Smirnov A. (2022). Analiz modeli BERT kak instrumenta opredeleniya mery smyslovoi blizosti predlozhenii estestvennogo yazyka. *StudNet*, №5, URL: <https://cyberleninka.ru/article/n/analiz-modeli-bert-kak-instrumenta-opredeleniya-mery-smyslovoy-blizosti-predlozheniy-estestvennogo-yazyka> (in Rus.)

Si lhousain Aouragha, Abdellah Yousfib, Saida Laaroussic, Hicham Gueddahd, Mohammed Nejja. (2021). A New Estimate of the n-gram Language Model. In *Proceedings of the 5th International Conference on AI in Computational Linguistics*. Pp. 211–215. DOI: 10.1016/j.procs.2021.05.111 (in Eng.)

Vaswani et al. (2017). Attention Is All You Need. In *Proceedings of the 31st Conference on Neural Information Processing Systems (NeurIPS 2017)*, Long Beach, CA, USA, 2017 (in Eng.)

NEWS OF THE NATIONAL ACADEMY OF SCIENCES OF THE REPUBLIC OF  
KAZAKHSTAN  
PHYSICO-MATHEMATICAL SERIES  
ISSN 1991-346X  
Volume 3. Number 347 (2023). 147–160  
<https://doi.org/10.32014/2023.2518-1726.210>

UDC 004.056  
МРНТИ 81.93.29

© **B. Rzayev\***, **Zh. Beldeubayeva**, **I. Uvaliyeva**, 2023

S. Seifullin Kazakh Agro-Technical Research University, Astana, Kazakhstan.  
E-mail: pathinchaos@gmail.com

### IDENTIFICATION OF MALICIOUS DATA IN THE INFORMATION NETWORK BY USING THE STACKING METHOD

**Rzayev Babyr Temirbekuly** — doctoral student, S. Seifullin Kazakh Agro-Technical Research University. 010000. Astana, Kazakhstan

E-mail: pathinchaos@gmail.com. ORCID ID: <https://orcid.org/0000-0002-9671-650X>;

**Beldeubayeva Zhanar Toleubayevna** — PhD, Senior lecturer, S. Seifullin Kazakh Agro-Technical Research University. 010000. Astana, Kazakhstan

E-mail: zh.beldeubayeva@mail.ru. ORCID ID: <https://orcid.org/0000-0003-4056-6220>;

**Uvaliyeva Indira Makhmutovna** — PhD, associate professor, D. Serikbayev East Kazakhstan Technical University. 070000. Ust-Kamenogorsk, Kazakhstan

E-mail: iuvaliyeva@mail.ru. ORCID ID: <https://orcid.org/0000-0002-2117-5390>.

**Abstract.** Information security is now more relevant than ever, and information is now as valuable to criminals as our physical property. The attacker's motives may include stealing information, obtaining financial benefits, spying or sabotage. Organizations should allocate funds to ensure security and be ready to detect, respond and proactively prevent attacks such as phishing, malicious software, viruses, malicious insiders and ransomware. Since the number of cyber threats is growing rapidly, organizations cannot prepare for all of them. Often, the information security systems used are not enough to identify new types of attacks and vulnerabilities — it is necessary to complement existing security systems with intelligent solutions. This paper proposes an approach to solving the problem of detecting malicious traffic in data transmission networks based on processing the received tuples of information sequences of network packets by the ensemble classification method — stacking machine learning algorithms. The approach does not require special data preparation, the resulting classification errors of individual algorithms are smoothed out by the solution of the metaclassifier. The proposed solution, in order to increase the accuracy and completeness of detecting destructive effects, makes it possible to use its classification algorithms optimized for different types of anomalies, which are trained on their own subsets of data

presented as a tuple of values of information sequences of network packets. The experiment is described using the machine learning classifiers Naïve Bayes, Hoeffding Three, Random Tree, REP Tree and J48. The evaluation was carried out using classifiers separately and using stacking, which was based on the same classifiers. Experimental results were obtained on the NSL-KDD public dataset. The software implementation of the approach, as a full-fledged intellectual solution, will make it possible to more effectively identify destructive effects. The approach can be applied as an addition to the existing monitoring systems of organizations related to network traffic processing. The essential advantages of the approach are its versatility for various technologies and data processing systems, the purpose of which is the accurate classification of data, and scalability, through the use of additional algorithms beyond those used in the approach. The purpose of this study is to increase the accuracy of detecting malicious network traffic by using stacking machine learning algorithms.

**Keywords:** information security, machine learning, stacking of algorithms, network traffic, NSL-KDD

© Б.Т. Рзаев\*, Ж.Т. Бельдеубаева, И.М. Увалиева, 2023

С. Сейфуллин атындағы Қазақ агротехникалық зерттеу университеті,  
Астана, Қазақстан.

E-mail: pathinchaos@gmail.com

## СТЕКИНГ ӘДІСІН ҚОЛДАНУ АРҚЫЛЫ АҚПАРАТТЫҚ ЖЕЛІДЕГІ ЗИЯНДЫ ДЕРЕКТЕРДІ АНЫҚТАУ

**Рзаев Бабыр Темірбекұлы** — докторант. С. Сейфуллин атындағы Қазақ агротехникалық зерттеу университеті. 010000. Астана, Қазақстан

E-mail: pathinchaos@gmail.com. ORCID ID: <https://orcid.org/0000-0002-9671-650X>;

**Бельдеубаева Жанар Толеубаевна** — PhD, С. Сейфуллин атындағы Қазақ агротехникалық зерттеу университеті. 010000. Астана, Қазақстан

E-mail: zh.beldeubayeva@mail.ru. ORCID ID: <https://orcid.org/0000-0003-4056-6220>;

**Увалиева Индира Махмутовна** — PhD, қауымдастырылған профессор, Д. Серікбаев атындағы Шығыс Қазақстан техникалық университеті. 010000. Өскемен, Қазақстан

E-mail: iuvalieva@mail.ru. ORCID ID: <https://orcid.org/0000-0002-2117-5390>.

**Аннотация.** Қазіргі таңда ақпараттық қауіпсіздік бұрынғыдан да өзекті ал ақпарат біздің физикалық мүлкіміз сияқты қылмыскерлер үшін құнды болып келеді. Қылмыскердің мотивтеріне ақпаратты ұрлау, қаржылық пайда табу, тыңшылық немесе диверсия кіруі мүмкін. Ұйымдар қауіпсіздікті қамтамасыз етуге қаражат бөліп, фишинг, зиянды бағдарламалық жасақтама, вирустар, зиянды инсайдерлер және төлем бағдарламалары сияқты шабуылдарды анықтауға, әрекет етуге және алдын-алуға дайын болуы керек. Киберқауіптердің саны тез өсіп келе жатқандықтан, ұйымдар олардың барлығына дайын бола алмайды. Көбінесе ақпараттық қауіпсіздікті қамтамасыз ету жүйелері шабуылдардың жаңа түрлерін және осалдықтарды анықтау үшін жеткіліксіз,

сондықтан қолданыстағы қауіпсіздік жүйелерін зияткерлік шешімдермен толықтыру қажет. Бұл жұмыста ансамбльдік жіктеу әдісі — машиналық оқыту алгоритмдерінің стекингімен желілік пакеттердің ақпараттық тізбектер кортеждерін өңдеуге негізделген деректерді тарату желілеріндегі зиянды трафикті анықтау мәселесін шешу тәсілі ұсынылады. Тәсіл деректерді арнайы дайындауды қажет етпейді, алынған жеке алгоритмдердің жіктеу қателері метаклассификатор шешімімен тегістеледі. Диструктивті әсерлерді анықтаудың дәлдігі мен толықтығы көрсеткіштерін арттыру мақсатында ұсынылған шешім желілік пакеттердің ақпараттық тізбектер мәндерінің кортежі түрінде ұсынылған деректердің өзіндік ішкі жиындарында оқытылған түрлі аномалиялар үшін оңтайландырылған олардың жіктеу алгоритмдерін пайдалануға мүмкіндік береді. Naïve Bayes, Hoeffding Three, Random Tree, REP Tree және J48 машиналық оқыту классификаторларын қолданылған эксперименттердің сипаттамасы берілген. Бағалау классификаторларды жеке қолдану және сол классификаторларды жіктеу негізіндегі стекингті қолдану арқылы жүргізілді. Эксперименттік нәтижелер NSL-KDD жалпыға ортақ деректер жинағын өңдеу арқылы алынған. Диструктивті әсерлерді тиімді анықтауға мүмкіндік беретін толыққанды интеллектуалды шешім ретінде тәсілді бағдарламалық қамтамасыз ету қажет. Тәсіл желілік трафикті өңдеуге қатысты ұйымдардың қолданыстағы бақылау жүйелеріне қосымша ретінде қолданылуы мүмкін. Тәсілдің маңызды артықшылықтары оның әртүрлі технологиялар мен деректерді өңдеу жүйелері үшін әмбебаптығы және тәсілде қолданылғаннан тыс қосымша алгоритмдерді қолдана отырып масштабталуы болып табылады. Зерттеудің мақсаты машиналық оқыту алгоритмдер стекингі арқылы зиянды желілік трафикті анықтау дәлдігін арттыру болып табылады.

**Түйін сөздер:** ақпараттық қауіпсіздік, машиналық оқыту, алгоритмдер стекингі, желілік трафик, NSL-KDD

© **Б.Т. Рзаев\***, **Ж.Т. Бельдеубаева**, **И.М. Увалиева**, 2023

Казахский агротехнический исследовательский университет

имени С. Сейфуллина, Астана, Казахстан.

E-mail: pathinchaos@gmail.com

## **ИДЕНТИФИКАЦИЯ ВРЕДНОСНЫХ ДАННЫХ В ИНФОРМАЦИОННОЙ СЕТИ С ИСПОЛЬЗОВАНИЕМ МЕТОДА СТЕКИНГА**

**Рзаев Бабыр Темирбекулы** — докторант, Казахский агротехнический исследовательский университет имени С.Сейфуллина. 010000. Астана, Казахстан

E-mail: pathinchaos@gmail.com. ORCID ID: <https://orcid.org/0000-0002-9671-650X>;

**Бельдеубаева Жанар Толеубаевна** — PhD, Казахский агротехнический исследовательский университет имени С.Сейфуллина. 010000. Астана, Казахстан

E-mail: zh.beldeubayeva@mail.ru. ORCID ID: <https://orcid.org/0000-0003-4056-6220>;

**Увалиева Индира Махмутовна** — PhD, ассоциированный профессор, Восточно-Казахстанский технический университет им. Д. Серикбаева. 010000. Усть-Каменогорск, Казахстан  
E-mail: iuvalieva@mail.ru. ORCID ID: <https://orcid.org/0000-0002-2117-5390>.

**Аннотация.** Информационная безопасность сейчас как никогда актуальна, а информация теперь так же ценна для преступников, как и наше физическое имущество. Мотивы злоумышленника могут включать кражу информации, получение финансовой выгоды, шпионаж или саботаж. Организации должны выделять средства на обеспечение безопасности и быть готовыми к обнаружению, реагированию и упреждающему предотвращению таких атак, как фишинг, вредоносное программное обеспечение, вирусы, вредоносные инсайдеры и программы — вымогатели. Поскольку количество киберугроз быстро растет, организации не могут подготовиться ко всем из них. Зачастую, имеющихся систем обеспечения информационной безопасности недостаточно для выявления новых видов атак и уязвимостей. Необходимо доукомплектовывать существующие системы безопасности новыми интеллектуальными решениями. В данной работе предлагается подход к решению проблемы выявления вредоносного трафика в сетях передачи данных, основанный на обработке полученных кортежей информационных последовательностей сетевых пакетов ансамблевым методом классификации — стекингом алгоритмов машинного обучения. Подход не требует специальной подготовки данных и полученные ошибки классификации отдельных алгоритмов сглаживаются решением метаклассификатора. Предложенное решение с целью повышения показателей точности и полноты выявления деструктивных воздействий дает возможность использовать оптимизированные для разных типов аномалий свои алгоритмы классификации, которые обучены на собственных подмножествах данных, представленных в виде кортежа значений информационных последовательностей сетевых пакетов. Приведено описание эксперимента с использованием классификаторов машинного обучения Naïve Bayes, Hoeffding Tree, Random Tree, REP Tree и J48. Оценка производилась с использованием классификаторов в отдельности и с применением стекинга, в основе которого были использованы те же классификаторы. Экспериментальные результаты получены на публичном наборе данных NSL-KDD. Программная реализация подхода как полноценного интеллектуального решения позволит более эффективно выявлять деструктивные воздействия. Подход может быть применим как дополнение к существующим системам мониторинга организаций, связанных с обработкой сетевого трафика. Существенными преимуществами подхода является его универсальность для различных технологий и систем обработки данных, целью которых является точная классификация данных и масштабируемость путем применения дополнительных алгоритмов сверх используемых в подходе.

Целью данного исследования является повышение точности выявления

вредоносного сетевого трафика, путем применения стекинга алгоритмов машинного обучения.

**Ключевые слова:** информационная безопасность, машинное обучение, стекинг алгоритмов, сетевой трафик, NSL-KDD

### **Introduction**

The functioning of corporate telecommunications networks (CTS) requires constant monitoring of the occurrence of all types of failures, collisions associated with the processing of network traffic. The development of the concept of the industrial Internet, the Internet of Things makes it necessary to assess the operability, functional security of individual network devices and the network segments formed by them. The analysis of the CTS operability is carried out using various monitors that process internal and external information containing statistical data of network packets and indicators of their processing. As a result, multidimensional time series are formed, which can contain many time-varying parameters reflecting the functioning of the system.

The variety of elements of the Internet of Things, a large number of objects, protocols of interaction in network traffic, data processing technologies, heterogeneity of formats, constantly changing architecture, configuration changes and improvement of attacks cause problems with prompt detection and response to subsequent information security (IS) incidents. This shows that existing approaches and methods may not always be effective in conditions of constant changes in CTS.

The task of identifying malicious data refers to the tasks of detecting anomalies (Cyril, 2022). There are specialized anomaly detection systems (Knapp, 2011), which allow you to detect unusual behavior or events in network traffic. They can help network administrators detect and respond to security threats, network errors, and performance issues. In recent years, there has been a growing interest of the scientific community in investigating the problem of anomaly detection using machine learning and deep learning methods (Yuan Gao et al., 2023).

### **Related works**

Here is an overview of scientific studies on key aspects of detecting anomalies in network traffic.

Types of anomalies. Study (Hayes et al., 2015) offers the division of anomalies in network traffic into three types: point anomalies, contextual anomalies and collective anomalies. Point anomalies refer to individual data points that differ significantly from the rest of the data. Contextual anomalies occur when the behavior of the system deviates from the expected context or pattern. Collective anomalies occur when a group of data points deviates from an expected pattern.

Problems with detecting anomalies. Detection of anomalies in network traffic faces a number of problems, including high dimensionality (Zheng et al., 2022) and the complexity of network data, the dynamic nature of network traffic (Stephen Ranshous et al., 2015) and the presence of noisy and incomplete data. In addition, it is necessary to carefully balance the ratio between false positive and false negative results in order not to miss real anomalies and minimize false positives.

Areas of application of anomaly detection. Anomaly detection systems in network traffic have a wide range of applications, including intrusion detection, network monitoring and performance analysis. For example, anomaly detection can be used to detect malicious activity in network traffic, such as DDoS attacks (Purwanto et al., 2014; Haiping et al., 2022; Purwanto et al., 2015; Chovanec et al., 2023; Lopez et al., 2019) and botnet activity (Zhao et al., 2013; Alaa Obeidat et al., 2022). It can also be used to detect network performance issues (Wawrowski et al., 2023; Igor Fosić et al., 2023), such as packet loss and latency, and to identify optimization opportunities.

Approaches to detecting anomalies in network traffic include statistical methods, machine learning methods, and rule-based methods.

One of the common approaches to detecting anomalies in network traffic are statistical methods (Iglesias Vázquez et al., 2014; Liu et al., 2023). This includes analyzing the statistical properties of network traffic data to detect unusual patterns or behaviors. For example, an anomaly can be detected if the volume of traffic or the frequency of certain types of traffic deviates significantly from the expected levels. An example of a statistical method for detecting anomalies is the use of moving averages (Zhang et al., 2020; Zhou Zeng-Guang et al., 2016) or exponential smoothing (Tang et al., 2022) to identify trends and anomalies in network traffic data.

Machine learning methods (Eduardo Weber Wächter et al., 2022; Nassif Ali et al., 2021; Thudumu et al., 2020). They can also be used to detect anomalies in network traffic. These methods include training the model on a large set of network traffic data and using the model to identify unusual patterns or behaviors in new data. For example, clustering methods can determine the current state of IoT devices (Sukhoparov et al., 2020). Also, an example of a machine learning method for detecting anomalies is the use of neural networks (Benjamin Staar et al., 2019; AlDahoul et al., 2021), who are able to study complex patterns and relationships in data.

Rule-based methods (Elfaki Abdelrahman, 2014; Duffield et al., 2009), They can also be used to detect anomalies in network traffic. These methods include defining a set of rules or thresholds that trigger an alert when certain conditions are met. For example, a rule can be defined to trigger an alert if the number of failed login attempts exceeds a certain threshold within a specified time period. An example of a rule-based anomaly detection method is the use of Snort (Szmit Maciej et al., 2007). An open source intrusion detection system that uses a set of predefined rules to detect various types of network threats.

In conclusion, it should be noted that each conducted study can make a significant contribution to the implementation of anomaly detection systems, which are an important tool for network administrators and security specialists. They help detect unusual behavior and events in network traffic, identify security threats and network performance issues. With the advent of machine learning and deep learning methods, the accuracy and efficiency of anomaly detection systems continue to increase.



**Proposed solution**

This work involves the use of a stacking algorithm to identify abnormal, potentially malicious data in network traffic.

The operation of the algorithm can be represented as follows (Figure 1).

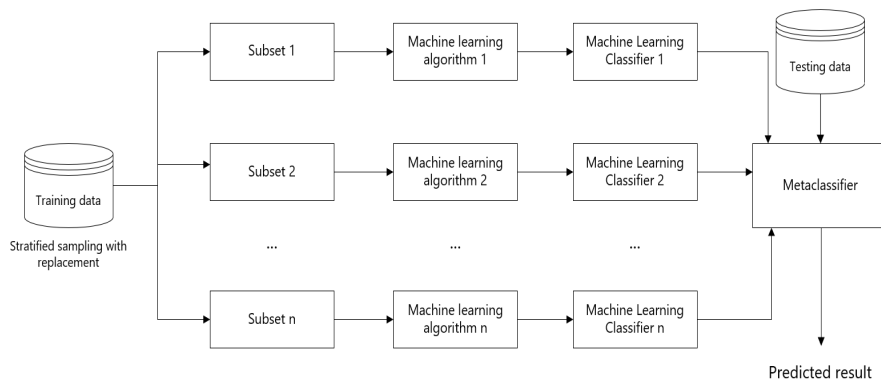


Fig. 1. Stacking operation scheme

There are  $n$ -distinct subsets of the training dataset, which are created using a stratified sample with substitution, where the relative proportion of different classes is preserved in all subsets (Sikora Riyaz et al., 2014). Each subset of the training set is used to determine the performance of classifiers in the training set. A metaclassifier in the form of a relative weight for each classifier is created by assigning a weight proportional to its performance to the classifier.

When evaluating an instance from a test set, each algorithm outputs a class distribution vector for this instance, which gives the probability that this particular instance belongs to this class. We can represent the vector of class distribution over  $c$  classes for the  $j$ -th classifier by the  $1 \times c$  vector as follows:

$$\Delta_j = [\delta_{1j} \delta_{2j} \dots \delta_{cj}] \quad 1 \leq j \leq n \tag{1}$$

where,

$$0 \leq \delta_{ij} \leq 1 \quad \forall 1 \leq i \leq c$$

$$\sum_i \delta_{ij} = 1$$

The class distribution vectors for  $n$  classifiers can then be represented by an  $n \times c$  matrix as follows:

$$\Delta = [\Delta_1 \Delta_2 \dots \Delta_n]^T \tag{2}$$

The metaclassifier creates a weight distribution vector that gives relative weight to various advertisements. The vector of weight distribution over  $n$  classifiers is represented as follows:

$$\theta = [\theta_1 \theta_2 \dots \theta_n] \tag{3}$$

where,

$$\begin{aligned} 0 \leq \delta_j \leq 1 \\ \sum_j \theta_j = 1 \end{aligned}$$

Given a class distribution matrix and a weight distribution vector, the metaclassifier evaluates each instance of the test set using the following class distribution  $I \times c$  vector:

$$\Delta' = \theta \cdot \Delta = [\delta'_1 \delta'_2 \dots \delta'_c] \tag{4}$$

где,

$$\delta'_i = \sum_j \theta_j \delta_{ij}$$

At the same time, as mentioned above, the stacking algorithm assumes that the weight distribution vector of the metaclassifier  $\theta$  it is created by assigning a weight to the classifier proportional to its performance.

*Evaluation of the proposed solution and results*

The experimental evaluation of the proposed solution was carried out using the publicly distributed NSL-KDD dataset (Bhupendra Ingre et al., 2015), which contains 125973 entries for training and 22544 entries for testing.

The description of the features of the NSL-KDD set is presented in Table 1.

Table 1. Description of the features of the NSL-KDD dataset

№	Feature	Description
1	Duration	Duration of connection time
2	Protocol_type	Protocol used for connection
3	Service	Destination network service used
4	Flag	Connection Status - Normal or Error
5	Src_bytes	The number of bytes of data transferred from the source to the destination in a single connection
6	Dst_bytes	The number of bytes of data transferred from the destination to the source in a single connection
7	Land	if the source and destination IP addresses and port numbers are equal, then this variable takes the value 1, another 0
8	Wrong_fragment	The total number of incorrect fragments in this connection
9	Urgent	The number of urgent packets in this connection. Urgent packets are packets with the urgency bit activated.
10	Hot	The number of "burning" indicators in the content, such as: entering the system directory, creating programs and executing programs
11	Num_failed_logins	Number of failed login attempts
12	Logged_in	Login status: 1 on successful login; 0 otherwise
13	Num_compromised	Number of "compromise" conditions
14	Root_shell	1 if the root shell is obtained; 0 otherwise
15	Su_attempted	1 if the "su root" command was tried or used; 0 otherwise

16	Num_root	The number of "root" accesses or the number of operations performed as root in the connection
17	Num_file_creations	The number of operations to create a file when connecting
18	Num_shells	Number of shell hints
19	Num_access_files	Number of operations with access control files
20	Num_outbound_cmds	Numbering of outgoing commands in an ftp session
21	Is_hot_login	1 if the login belongs to the "hot" list, i.e. root or administrator; otherwise 0
22	Is_guest_login	1 if login is "guest"; 0 otherwise
23	Count	The number of connections to the same destination node as the current connection in the last two seconds.
24	Srv_count	The number of connections to the same service (port number) as the current connection in the last two seconds.
25	Serror_rate	Percentage of connections that activated the flag (4) s0, s1, s2 or s3 among the connections combined in count (23)
26	Srv_serror_rate	Percentage of connections that activated the flag (4) s0, s1, s2 or s3 among the connections combined in srv_count (24)
27	Rerror_rate	Percentage of connections that activated the REJ flag (4) among the connections combined in count (23)
28	Srv_rerror_rate	Percentage of connections that activated the REJ flag (4) among connections combined in srv_count (24)
29	Same_srv_rate	Percentage of connections to the same service, among the connections combined in the account (23)
30	Diff_srv_rate	Percentage of connections to various services, among the connections combined in the column (23)
31	Srv_diff_host_rate	Percentage of connections that were to different destination machines among the connections combined in srv_count (24)
32	Dst_host_count	The number of connections having the same destination host IP address
33	Dst_host_srv_count	The number of connections having the same port number
34	Dst_host_same_srv_rate	Percentage of connections to the same service, among connections combined in dst_host_count (32)
35	Dst_host_diff_srv_rate	Percentage of connections to various services, among connections combined in dst_host_count (32)
36	Dst_host_same_src_port_rate	Percentage of connections that were to the same source port among the connections combined in dst_host_srv_count (33)
37	Dst_host_srv_diff_host_rate	Percentage of connections that were to different destination machines among the connections combined in dst_host_srv_count (33)
38	Dst_host_serror_rate	Percentage of connections that activated the flag (4) s0, s1, s2 or s3 among the connections combined in dst_host_count (32)
39	Dst_host_srv_serror_rate	Percentage of connections that activated the flag (4) s0, s1, s2 or s3 among the connections combined in dst_host_srv_count (33)
40	Dst_host_rerror_rate	Percentage of connections that activated the (4) REJ flag among the connections combined in dst_host_count (32)
41	Dst_host_srv_rerror_rate	Percentage of connections that activated the REJ flag (4) among the connections combined in dst_host_srv_count (33)
42	Class	Class of data

Training is performed based on KDDTrain data, which contains 22 types of attacks, and testing is performed on KDDTest data, which contains an additional 17 types of attacks. These attacks are divided into four classes of attacks:

Denial of Service (DoS) - malicious attempt to block system or network resources and services.

Probe – this attack collects information about potential vulnerabilities of the target system, which can later be used to launch attacks on these systems.

Remote to Local (R2L) – Unauthorized ability to send data packets to a remote system over the network and gain access either as a user or as root to perform their unauthorized actions.

User to Root (U2R) – In this case, attackers gain access to the system as a normal user and hack vulnerabilities to gain administrative privileges.

In addition to attacks, for classification purposes, the set also contains normal data that does not contain malicious components. The number of NSL-KDD set entries distributed by classes is shown in the Table 2.

Table 2. Number of NSL-KDD set entries distributed by category

Testing dataset		Training dataset	
Class	Number of entries	Class	Number of entries
Normal	67343	Normal	9711
DOS	45927	DOS	7458
Probe	11656	Probe	2421
R2L	995	R2L	2754
U2R	52	U2R	200
Total	125973	Total	22544

The freely distributed Weka application for data processing and machine learning was used for the experiment. This application is written in Java at the University of Waikato (New Zealand), distributed under the GNU GPL license (Wikipedia, 2021).

The first part of the experiment involved the use of classifiers separately. Naïve Bayes (NB), Hoeffding Tree (HT), Random Tree (RT), REP Tree (REP) and J48 were selected from the classifiers.

The classifiers were evaluated using the ROC error curve, which displays the ratio between correctly classified records and incorrectly classified ones, Figure 2.

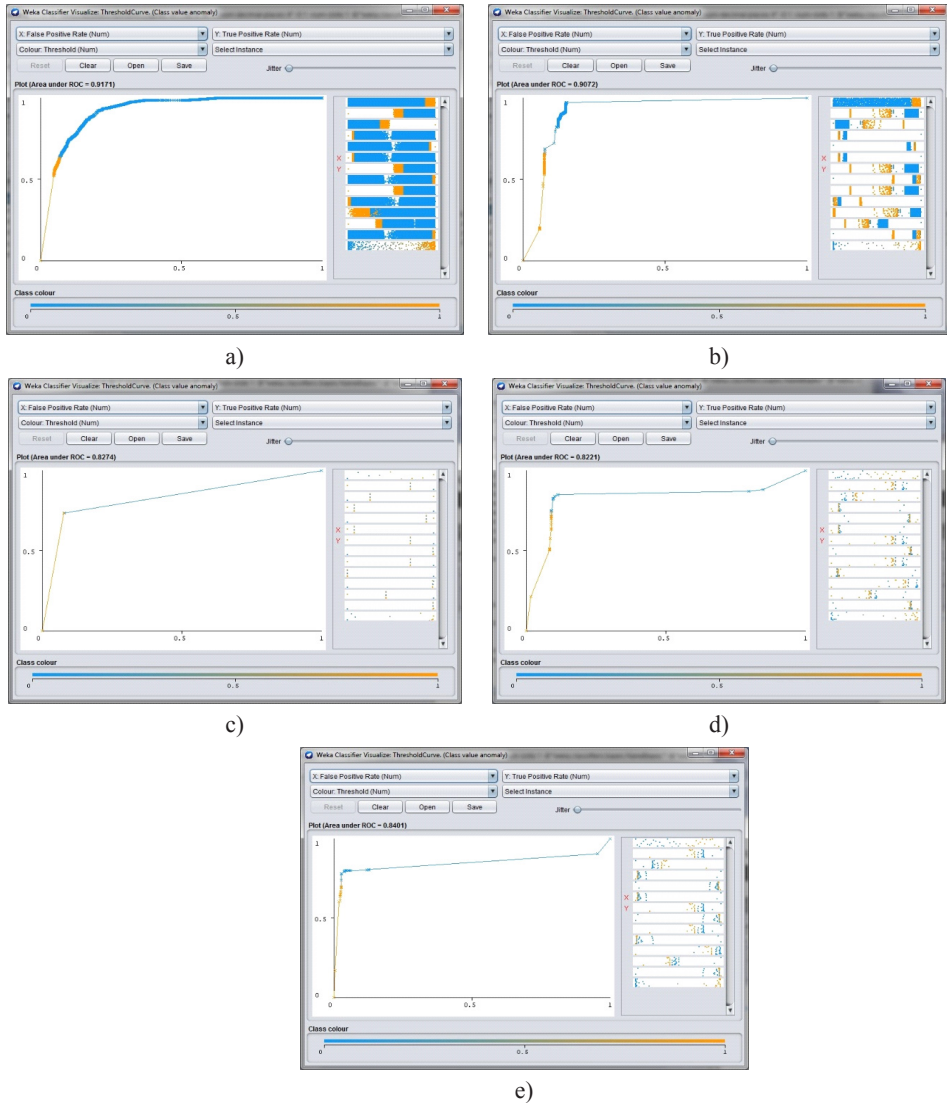


Fig. 2. ROC curves: a) NB, b) HT, c) RT, d) REP, e) J48

The obtained results of data classification using ML algorithms are presented in the Table 3.

Table 3. Classification results

Parameter\Classifier	NB	HT	RT	REP	J48
Accuracy, %	76,1	77,1	81,3	81,5	81,5
Precision, %	80,9	81,2	83,7	83,5	85,8
Recall, %	76,1	77,2	81,4	81,5	81,5
F-measures, %	75,9	77,1	81,4	81,6	81,5
ROC	91,7	90,7	82,7	82,2	84,0

The best results on the ROC curve were shown by the NB classifier (81,5 %).

In the second part of the experiment, the stacking of ML algorithms (ST) was implemented using the same classifiers - Naïve Bayes (NB), Hoeffding Tree (HT), Random Tree (RT), REP Tree (REP) and J48. The metaclassifier for stacking was chosen – Logistic Regression.

The assessment of the use of stacking was carried out according to the ROC curve, Figure 3.

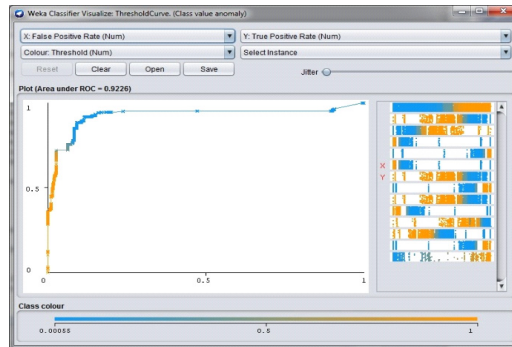


Fig. 3. Indicators of the ROC curve when using stacking

The results of classifiers individually and with the use of stacking are shown in the Table 4.

Table 4. Comparative classification results

Parameter\Classifier	NB	HT	RT	REP	J48	ST
Accuracy, %	76,1	77,1	81,3	81,5	81,5	82,2
Precision, %	80,9	81,2	83,7	83,5	85,8	86,2
Recall, %	76,1	77,2	81,4	81,5	81,5	82,3
F-measures, %	75,9	77,1	81,4	81,6	81,5	82,2
ROC	91,7	90,7	82,7	82,2	84,0	92,2

A comparative histogram using classifiers separately and using stacking is shown in Figure 4.

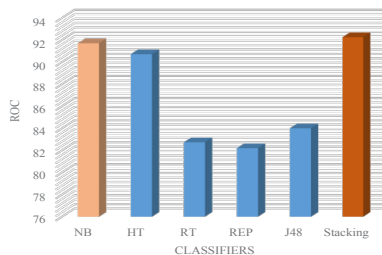


Fig. 4. Comparative histogram with using stacking

The stacking results are superior to the result of the best single classifier. Therefore, it can be argued about the effectiveness of the use of stacking algorithms more than the use of algorithms separately.

## Conclusion

This article proposes an approach to the identification of malicious traffic based on the use of an ensemble method of machine learning – stacking. The approach is based on the use of individual ML classifiers as the basic ones, and on their basis the training of the metaclassifier, in our case, logistic regression is used for subsequent data classification.

The application of the proposed solution based on stacking makes it possible to improve classification accuracy indicators rather than using separate classifiers.

The advantage of using stacking is versatility for various data processing systems and scalability, through the use of additional algorithms beyond those used in the approach.

The disadvantage of the proposed approach is sensitivity to data quality, as one of the main requirements for data classification by MO methods, as well as the need for additional computing resources when the model becomes more complex.

## REFERENCES

- Cyral, 2022. Anomaly detection. URL: <https://cyral.com/glossary/anomaly-detection/>
- Knapp E., 2011. Exception, Anomaly, and Threat Detection. *Industrial Network Security*. Pp. 189–214. DOI:10.1016/b978-1-59749-645-2.0000.
- Yuan Gao, Xianhui Yin, Zhen He, Xueqing Wang, 2023. A deep learning process anomaly detection approach with representative latent features for low discriminative and insufficient abnormal data. *Computers & Industrial Engineering*. Volume 176, 108936. ISSN 0360-8352. DOI: <https://doi.org/10.1016/j.cie.2022.108936>.
- Hayes M.A., Capretz M.A., 2015. Contextual anomaly detection framework for big sensor data. *Journal of Big Data 2*. DOI: <https://doi.org/10.1186/s40537-014-0011-y>.
- Zheng J., Li J., Liu C. et al., 2022. Anomaly detection for high-dimensional space using deep hypersphere fused with probability approach. *Complex Intell. Syst.* 8. 4205–4220. DOI: <https://doi.org/10.1007/s40747-022-00695-9>.
- Stephen Ranshous, Shitian Shen, Danai Koutra, Steve Harenberg, Christos Faloutsos, Nagiza F. Samatova, 2015. Anomaly detection in dynamic networks: a survey *WIREs Comput Stat* 2015. 7. Pp. 223–247. DOI: 10.1002/wics.1347.
- Y. Purwanto, Kuspriyanto Hendrawan, B. Rahardjo, 2014. Traffic anomaly detection in DDos flooding attack. 8th International Conference on Telecommunication Systems Services and Applications (TSSA), Kuta. Bali. Indonesia. Pp. 1–6. DOI: 10.1109/TSSA.2014.7065953.
- Haiping Lin, Chengwen Wu, Mohammad Masdari, 2022. A comprehensive survey of network traffic anomalies and DDoS attacks detection schemes using fuzzy techniques. *Computers and Electrical Engineering*. Volume 104. Part B. 108466. ISSN 0045-7906. DOI: <https://doi.org/10.1016/j.compeleceng.2022.108466>.
- Purwanto Yudha, Kuspriyanto, Hendrawan Temmy, Rahardjo Budi, 2015. Traffic anomaly detection in DDos flooding attack. *Proceedings of 2014 8th International Conference on Telecommunication Systems Services and Applications*. TSSA 2014. DOI: 10.1109/TSSA.2014.7065953.
- Chovanec M., Hasin M., Havrilla M., Chovancová E., 2023. Detection of HTTP DDoS Attacks Using NFStream and TensorFlow. *Applied Sciences*. 13(11):6671. DOI: <https://doi.org/10.3390/app13116671>.
- Lopez Alma D., Mohan Asha P., Nair Sukumaran, 2019. Network Traffic Behavioral Analytics for Detection of DDoS Attacks. *SMU Data Science Review*. Vol. 2. No. 1. Article 14.
- Zhao David, Traore Issa, Sayed, Bassam Lu, Wei Saad, Sherif Ghorbani, Ali Garant Dan, 2013. Botnet detection based on traffic behavior analysis and flow intervals. *Computers & Security*. 39. 2–16. DOI: 10.1016/j.cose.2013.04.007

Alaa Obeidat, Rola Yaqbeh, 2022. Smart Approach for Botnet Detection Based on Network Traffic Analysis. *Journal of Electrical and Computer Engineering*. vol. 2022. Article ID 3073932. DOI: <https://doi.org/10.1155/2022/3073932>

Wawrowski L., Białas A., Kajzer A., Kozłowski A., Kurianowicz R., Sikora M., Szymańska-Kwiecień A., Uchroński M., Białczak M., Olejnik M., Michalak M., 2023. Anomaly Detection Module for Network Traffic Monitoring in Public Institutions. *Sensors*. 23(6):2974. DOI: <https://doi.org/10.3390/s23062974>.

Igor Fosić, Drago Žagar, Krešimir Grgić, Višnja Križanović, 2023. Anomaly detection in NetFlow network traffic using supervised machine learning algorithms. *Journal of Industrial Information Integration*. Volume 33. 100466. ISSN 2452–414X. DOI: <https://doi.org/10.1016/j.jii.2023.100466>.

Iglesias Vázquez, Félix, Zseby Tanja, 2014. Analysis of network traffic features for anomaly detection. *Machine Learning*. 101. DOI: [10.1007/s10994-014-5473-9](https://doi.org/10.1007/s10994-014-5473-9).

Liu H., Wang H., 2023. Real-Time Anomaly Detection of Network Traffic Based on CNN. *Symmetry*. 15(6):1205. DOI: <https://doi.org/10.3390/sym15061205>.

Zhang M., Guo J., Li X., Jin R., 2020. Data-Driven Anomaly Detection Approach for Time-Series Streaming Data. *Sensors (Basel)*. 20(19):5646. DOI: [10.3390/s20195646](https://doi.org/10.3390/s20195646).

Zhou, Zeng-Guang, Tang Ping, 2016. Improving time series anomaly detection based on exponentially weighted moving average (EWMA) of season-trend model residuals. DOI: [10.1109/IGARSS.2016.7729882](https://doi.org/10.1109/IGARSS.2016.7729882).

Tang H., Wang Q., Jiang G., 2022. Time Series Anomaly Detection Model Based on Multi-Features. *Comput Intell Neurosci*. 2022:2371549. DOI: [10.1155/2022/2371549](https://doi.org/10.1155/2022/2371549).

Eduardo Weber Wächter, Server Kasap, Şefki Kolozali, Xiaojun Zhai, Shoaib Ehsan, Klaus D. McDonald-Maier, 2022. Using machine learning for anomaly detection on a system-on-chip under gamma radiation. *Nuclear Engineering and Technology*. Volume 54. Issue 11. Pp. 3985–3995. ISSN 1738-5733. DOI: <https://doi.org/10.1016/j.net.2022.06.028>.

NassifAli, Abu Talib Manar, Nasir Qassim, Dakalbab Fatima, 2021. Machine Learning for Anomaly Detection: A Systematic Review. *IEEE Access*. Pp. 1–1. DOI: [10.1109/ACCESS.2021.3083060](https://doi.org/10.1109/ACCESS.2021.3083060).

Thudumu S., Branch P., Jin J. et al., 2020. A comprehensive survey of anomaly detection techniques for high dimensional big data. *J Big Data* 7. 42. DOI: <https://doi.org/10.1186/s40537-020-00320-x>.

Sukhoparov M. E., Lebedev I. S., 2020. Identification of the state of information security of Internet of Things devices in information and telecommunication systems. *Control systems, communications and security*. № 3. Pp. 252–268. DOI: [10.24411/2410-9916-2020-10310](https://doi.org/10.24411/2410-9916-2020-10310).

Benjamin Staar, Michael Lütjen, Michael Freitag, 2019. Anomaly detection with convolutional neural networks for industrial surface inspection. *Procedia CIRP*. Volume 79. Pp. 484–489. ISSN 2212–8271. <https://doi.org/10.1016/j.procir.2019.02.123>.

AIDahou N., Abdul Karim H., Ba Wazir A.S., 2021. Model fusion of deep neural networks for anomaly detection. *J Big Data* 8. 106 (2021). DOI: <https://doi.org/10.1186/s40537-021-00496-w>.

Elfaki Abdelrahman, 2014. Using a Rule-based Method for Detecting Anomalies in Software Product Line. *Research Journal of Applied Sciences, Engineering and Technology*. 7.

N. Duffield, P. Haffner, B. Krishnamurthy, H. Ringberg, 2009. Rule-Based Anomaly Detection on IP Flows. *IEEE INFOCOM 2009*. Rio de Janeiro. Brazil. Pp. 424–432. DOI: [10.1109/INFCOM.2009.5061947](https://doi.org/10.1109/INFCOM.2009.5061947).

Szmit Maciej, Wężyk Radosław, Skowroński Maciej, Szmit Anna, 2007. Traffic anomaly detection with Snort.

Sikora Riyaz, Al-laymoun O'la Hmoud, 2014. A Modified Stacking Ensemble Machine Learning Algorithm Using Genetic Algorithms. *Journal of International Technology and Information Management*. Vol. 23. Iss. 1. Article 1. DOI: <https://doi.org/10.58729/1941-6679.1061>. Available at: <https://scholarworks.lib.csusb.edu/jitim/vol23/iss1/1>.

Bhupendra Ingre, Anamika Yadav, 2015. Performance Analysis of NSL-KDD dataset using ANN. 2015 International Conference on Signal Processing And Communication Engineering Systems (SPACES). Pp. 92–96. DOI: [10.1109/SPACES.2015.7058223](https://doi.org/10.1109/SPACES.2015.7058223).

Wikipedia, 2022. Weka. URL: <https://ru.wikipedia.org/>



NEWS OF THE NATIONAL ACADEMY OF SCIENCES OF THE REPUBLIC OF KAZAKHSTAN  
PHYSICO-MATHEMATICAL SERIES

ISSN 1991-346X

Volume 3. Number 347 (2023). 161–175

<https://doi.org/10.32014/2023.2518-1726.211>

UDC 681.518.3

© N.S. Baimuldina<sup>1</sup>, G.N. Skabayeva<sup>2\*</sup>, A. Zhaksybayeva<sup>1</sup>, 2023

<sup>1</sup>Al-Farabi Kazakh National University, Almaty, Kazakhstan;

<sup>2</sup>Kazakh national agrarian research university, Almaty, Kazakhstan.

E-mail: Gskbayeva@mail.ru

### PROJECT MANAGEMENT SOFTWARE IN THE FIELD OF BIOTECHNOLOGY

**Baimuldina Nazira** — Candidate of pedagogical Science, associate professor, Al-Farabi Kazakh National University, Almaty, Kazakhstan

E-mail: baimuldinanaziko@mail.ru. ORCID ID: <https://orcid.org/0000-0002-2976-7454>;

**Skabayeva Gulmira** — Candidate of pedagogical Science, associate professor, Kazakh national agrarian research university, Almaty, Kazakhstan

E-mail: Gskbayeva@mail.ru. ORCID ID: <https://orcid.org/0000-0002-2649-4454>;

**Zhaksybayeva Aizere** — Master of the Kazakh National University named after Al Farabi, Almaty, Kazakhstan

E-mail: aizerezhaksybaeva@gmail.com. ORCID ID: <https://orcid.org/0000-0001-7589-4529>.

**Abstract.** The list of currently relevant topics and areas that have moved to the global market include: biotechnology, project management and Information Systems. This article analyzes the interweaving of these three areas. Both in the world and in Kazakhstan, biotechnology methods and project management are used in medicine, agriculture and industry. If biology as a science is inherited by Kazakhstan from the USSR, then the project management industry will be dominated by sovereign Kazakhstan. The need to create a unified information system for Project Management in the field of Biotechnology is analyzed and considered in this article. The work is designed to optimize the aspect of Project Management in the field of Biotechnology in a modern way. In addition, it is intended for all specialists in the field of biotechnology and people who are interested in this field. Biotechnology and project management for the first time a set of technological methodology and general requirements for scientific research aimed at developing high-level work, projects. Biotechnology refers to the use of living systems, cells and organisms for human needs. Biotechnology is divided into many industries, including the following: medical, agricultural, industrial and environmental. It is very important for Kazakhstan to develop and raise biotechnology projects in these areas to the highest level.

**Keywords:** biotechnology, project management, software, IT Solutions, Project Management, Microsoft Office Project

© Н.С. Баймулдина<sup>1</sup>, Г.Н. Скабаева<sup>2\*</sup>, А.Д. Жақсыбаева<sup>1</sup>, 2023

<sup>1</sup>Әл-Фараби атындағы Қазақ ұлттық университеті, Алматы, Қазақстан;

<sup>2</sup>Қазақ ұлттық аграрлық зерттеу университеті, Алматы, Қазақстан.

E-mail: gskbayeva@mail.ru

## **БИОТЕХНОЛОГИЯ САЛАСЫНДАҒЫ ЖОБАЛАРДЫ БАСҚАРУДЫҢ БАҒДАРЛАМАЛЫҚ ҚАМТАМАСЫЗ ЕТУІ**

**Баймулдина Назира Сахимжановна** — педагогика ғылымдарының кандидаты, қауымдастырылған профессор, Әл-Фараби атындағы Қазақ ұлттық университеті, Алматы, Қазақстан

E-mail: baimuldinanaziko@mail.ru. ORCID: <https://orcid.org/0000-0002-2976-7454>;

**Скабаева Гүлмира Несипбаевна** — педагогика ғылымдарының кандидаты, қауымдастырылған профессор, Қазақ ұлттық аграрлық зерттеу университеті, Алматы, Қазақстан

E-mail: gskbayeva@mail.ru. ORCID: <https://orcid.org/0000-0002-2649-4454>;

**Жақсыбаева Айзере Дәулеткелдықызы** — Әл-Фараби атындағы Қазақ ұлттық университетінің магистрі, Әл-Фараби атындағы Қазақ ұлттық университеті, Алматы, Қазақстан  
E-mail: aizerezhaksybaeva@gmail.com. ORCID: <https://orcid.org/0000-0001-7589-4529>.

**Аннотация.** Қазіргі уақытта өзекті тақырыптардың тізіміне және әлемдік нарықта көшті бастап тұрған салаларға: биотехнология, жобаларды басқару және ақпараттық жүйелер кіреді. Бұл мақалада осы үш саланың тоғысуына талдау жасалынды. Барлық әлемде де, Қазақстанда да биотехнология әдістері және жобаларды басқару медицинада, ауыл шаруашылығында және өндірісте қолданылады. Биология ғылым ретінде Қазақстанға КСРО-дан мұра болып қалса, жобаларды басқару саласы егеменді Қазақстанның мемлекетті басқаруынан бас алады. Биотехнология саласындағы жобаларды басқару үшін бірыңғай ақпараттық жүйені құру қажеттілігін осы мақалада талданып, қарастырылады. Жұмыс қазіргі заманға сай биотехнология саласындағы жобаларды басқару аспектісін оңтайландыруға арналған. Одан бөлек биотехнология саласындағы барлық мамандарға және осы салаға қызығушылығы бар адамдарға арналған. Биотехнология және жобаларды басқару алғаш рет жоғары деңгейдегі жұмыстарды, жобаларды әзірлеуге бағытталған технологиялық әдістеме және ғылыми зерттеулеріне қойылатын жалпы талаптар жиынтығы. Биотехнология адам қажеттіліктеріне тірі жүйелерді, жасушаларды және организмдерді пайдалану. Биотехнология көп салаларға бөлінеді, соның ішінде келесідей салалар бар: медициналық, ауыл шаруашылығы, өнеркәсіптік және экологиялық. Қазақстан үшін осы салаларда биотехнология жобарадын дамыту және жоғарғы деңгейге көтеру өте маңызды.

**Түйін сөздер:** биотехнология, жобаларды басқару, бағдарламалық жасақтама, IT шешімдер, Project Management, Microsoft Office Project

© Н.С. Баймулдина<sup>1</sup>, Г.Н. Скабаева<sup>2\*</sup>, А.Д. Жақсыбаева<sup>1</sup>, 2023

<sup>1</sup>Казахский национальный университет имени аль-Фараби,

г. Алматы, Казахстан;

<sup>2</sup>Казахский национальный аграрный исследовательский университет,

г. Алматы, Казахстан.

E-mail: gskbayeva@mail.ru

## ПРОГРАММНОЕ ОБЕСПЕЧЕНИЕ ДЛЯ УПРАВЛЕНИЯ ПРОЕКТАМИ В ОБЛАСТИ БИОТЕХНОЛОГИИ

**Баймулдина Назира Сахимжановна** — кандидат педагогических наук, ассоциированный профессор, Казахский национальный университет имени Аль-Фараби, Алматы, Казахстан

E-mail: baimuldinanaziko@mail.ru. ORCID: <https://orcid.org/0000-0002-2976-7454>;

**Скабаева Гульмира Нсипбаевна** — кандидат педагогических наук, ассоциированный профессор, Казахский национальный аграрный исследовательский университет, Алматы, Казахстан

E-mail: Gskbayeva@mail.ru. ORCID: <https://orcid.org/0000-0002-2649-4454>;

**Жақсыбаева Айзере Даулеткелдықызы** — магистр Казахского национального университета имени Аль-Фараби, Казахский национальный университет имени Аль-Фараби, Алматы, Казахстан

E-mail: aizerezhaksybaeva@gmail.com. ORCID: <https://orcid.org/0000-0001-7589-4529>.

**Аннотация.** В перечень актуальных в настоящее время тем и отраслей, начиная с мировых рынков, входят: биотехнологии, управление проектами и информационные системы. В этой статье анализируется слияние этих трех областей. Методы биотехнологии и управления проектами используются в медицине, сельском хозяйстве и производстве как во всем мире, так и в Казахстане. В то время как Биология, как наука, унаследовала Казахстан от СССР, сфера управления проектами отказывается от управления суверенным Казахстаном. Необходимость создания единой информационной системы для управления проектами в области биотехнологии анализируется и рассматривается в данной статье. Работа посвящена оптимизации аспекта управления проектами в современной биотехнологической отрасли. Кроме того, она предназначена для всех специалистов в области биотехнологии и людей, интересующихся этой областью. Биотехнология и управление проектами – совокупность общих требований к научным исследованиям и технологической методике, впервые направленных на разработку работ, проектов высокого уровня. Биотехнология — это использование живых систем, клеток и организмов для удовлетворения потребностей человека. Биотехнология делится на несколько отраслей, в том числе следующие: медицинские, сельскохозяйственные, промышленные и экологические. Для Казахстана очень важно развивать и повышать биотехнологические проекты в этих областях.

**Ключевые слова:** биотехнология, управление проектами, программное обеспечение, ИТ-решения, Project Management, Microsoft Office Project

## **Кіріспе**

*Биотехнология, жобаларды басқару және бағдарламалық қамтамасыз ету.*

Бүгінгі таңда Қазақстанда биотехнологияның алдында көптеген шешілмеген технологиялық міндеттер мен бағыттар бар. Биотехнология білім саласының бөлігі және қарқынды дамып келе жатқан өнеркәсіп саласы ретінде "адам – табиғат – қоғам" концепциясында өзара қарым-қатынас жүйесінде теңгерімді сақтауды қамтамасыз ете отырып, қазіргі заманның көптеген түйінді проблемаларын шешуге қабілетті. Микробиология, биохимия, генетика, гендік инженерия, иммунология туралы білімге сүйене отырып және заманауи жабдықтар мен құрылғыларды қолдана отырып, биотехнология табиғи және жасанды жүйелерді жасауға және жаңартылатын материалдардан адам өміріне қажетті барлық заттар мен қосылыстарды алуға мүмкіндік береді. Әлемдік конференция материалдары биотехнология саласындағы инновациялық дамытуға бағытталған мәселелердің кең спектрін көрсетеді. Нанобиотехнология, биоинформатика және нейроинформатика саласындағы жұмыстарға, сондай-ақ биоқауіпсіздік пен гендік-инженерлік қызметті мемлекеттік реттеудің әртүрлі аспектілеріне көп көңіл бөлінеді. Биотехнология саласы адамның қажеттіліктерін жасушалық және гендік-инженерлік әдістермен шешуді ойластырады. Бұл оңай емес және осы салада болашағы бар жобаларды ұйымдастыру үшін сізде үлкен білім мен өңіріңізде болып жатқан жаңа жобалардың базасы болуы керек.

Ендеше, биотехнология және жобаларды басқару дегеніміз не? Биотехнология - қазіргі заманғы ең маңызды ғылым. Бұл технология мен өнеркәсіптік өндірісте биологиялық процестерді қолдану туралы ғылым. Оның атауы грек сөздерінен шыққаны белгілі: *bios* – өмір; *teken-өнер*; *logos* - сөз, ілім, ғылым. Биологиялық процестерге әртүрлі сипаттағы (микробтық, өсімдік немесе жануар) биологиялық объекттерді қолданатын заттар кіреді, мысалы, бірқатар медициналық, тағамдық және басқа мақсаттағы өнімдерді өндіру — антибиотиктер, вакциналар, ферменттер, жемшөп пен тамақ ақуыздары, полисахаридтер, гормондар, гликозидтер, аминқышқылдары, алкалоидтар, биогаз, тыңайтқыштар және басқалар (Handbook of Genetics, 2014).

Жобаларды басқару — менеджмент пәннің құрамдас бөлігі және оның құралдар арсеналының бір бөлігін, оның тілін және негізгі анықтамаларын қолданады, сонымен бірге өзінің нақты түсініктерін қосады. Жобаны басқару берілген мақсаттарға тез және тиімді қол жеткізуге көмектеседі. Сонымен қатар, бұл процесте кез келген жобаның жалпы мақсаттары үшін қолдануға болатын кешендердің бүкіл жүйесі қалыптасады және ресурстарды сауатты бөлу схемасы жасалады. Қазіргі уақытта жобаларды басқару — бұл білім, ережелер мен стандарттар жиынтығы бар тұтас ғылым. Ең танымал ресурс — РМВОК. Бұл жобалық менеджмент саласындағы белгілі бір "білім кітабы", онда осы саладағы ең жақсы практикалық кеңестер мен білім сипатталған ("BARS Consulting Group" ЖШС, 2007).

Анықтамаларды ала отырып биотехнология саласындағы жобаларды зерттеудің маңыздылығын айта кету керек. Ғылыми институттар мен университеттер биотехнология саласын зерттеуді жалғастыруда, бірақ әзірге әлемдік деңгейде үлкен нәтижелерді көрсете алмаймыз, бұл зерттеулер нашар коммерцияланады, оларға аз көңіл бөлінуде. Бұған себепкер болған қаражат көзі және қызығушылық танытпайтын инвесторлар. Сол себептен біздің мамандар жақсы әрі тең жағдайда әлемдік жетекші компаниялармен бәсекеге түсе алмайды. Бірақ Қазақстан тарапынан осы уақытқа дейін көптеген идеялар жарияланған. Бұған дәлел ретінде Нұр-Сұлтан қаласында орналасықан Ұлттық Биотехнология орталығы болмақ.

Қазақстанда биотехнология саласын дамытудың негізгі перспективалары өндіріске ғылымды қажетсінетін технологияларды және денсаулық сақтау, ауыл шаруашылығы, қоршаған ортаны қорғау үшін бәсекеге қабілетті өнімдерді енгізумен байланысты. Сондай-ақ тамақ және қайта өңдеу өнеркәсібі, тау-кен ісі және пайдалы қазбаларды қайта өңдеу (бағалы металдарды өндірудегі биогеотехнологиялар) атап айтуымыз керек. Бұл бағыттар Ұлттық биотехнология орталығының міндеттеріне енгізілген және олардың бірнеше түрлері бойынша табысты атқарымдар бар (Томаг, 2014).

#### **Негізгі мақала материалы**

*Қазақстандағы биотехнология және жобаларды басқару.*

Қазіргі уақытта әлемнің жетекші елдері инновациялық экономика мен технологиялық көшбасшылыққа ұмтылуда, ғылымды дамытуға елеулі қаржы жұмсауда, бірақ әлемнің бірде-бір елі бір мезгілде барлық ғылыми бағыттар бойынша ауқымды зерттеулер жүргізе алмайды. Сондықтан қатаң бәсекелестік күресте елдік артықшылықтардың анықтау қажет, сонда олардың дамуы әлемдік нарықта жетекші орынға ие болады. Елдің инновациялық дамуының басымдықтарын анықтау, әлемнің көптеген елдерінде халықтың өмір сүру деңгейі мен сапасын арттыруға ықпал ететін ғылым мен технологиялардың "өсу нүктелерін" анықтау мақсатында форсайт-зерттеулер жүргізілуде. "Ұлттық мемлекеттік ғылыми-техникалық сараптама орталығы" АҚ 2013 жылы Қазақстан дамуының 2030 жылға дейінгі кезеңге арналған басым ғылыми және технологиялық бағыттарын, сондай-ақ басым тақырыптарды айқындауға бағытталған зерттеу нысанында "ғылым және технологиялар саласындағы жүйелі талдау және болжау" жобасын іске асырды, ғылыми зерттеулер жүргізді (Блинов, 1995; Додонов, 2011).

Жоба аясында сараптамалық және талдамалық зерттеулер 8 бағыт бойынша жүргізілді, оның ішінде денсаулық саласы, биотехнология, аграрлық сектордың орнықты дамуы, тамақ өнімдерін өңдеу және қауіпсіздігі бар. Қазақстандық сарапшылардың жұмысына Манчестер университетінің, Кореяның ғылым мен технологияларды бағалау және жоспарлау институтының және Финляндия академиясының беделді ғалымдары кеңес берді. Қазақстанда өткізілген жобаның экономиканың практикалық қажеттіліктеріне және белгілі бір әлеуметтік маңызды нәтиже алуға, отандық құзыреттерді дамытуға

бағдарланған оның әлеуметтік-экономикалық құрамдас бөлігіндегі айрықша ерекшелігі

Жоба нәтижелері негізінде Қазақстанда ғылыми бағыттарды дамыту үшін мүмкіндіктер (өз ғылыми құзыреттерінің болуы немесе даму мүмкіндігі); шетелдік ғылыми ұйымдармен бірлескен ғылыми зерттеулер жүргізу мүмкіндіктері; бұрыннан бар құзыреттер мен технологияларды трансферттеу қажеттілігі айқындалды. Жоба шеңберінде Қазақстанда 2030 жылға дейінгі кезеңге жүргізілетін пәнаралық зерттеулерді таңдау негіздемесі жүзеге асырылды (Раманқұлов, 2018).

2030 жылға дейінгі кезеңге Қазақстан үшін басым бағыттар ретінде пәнаралық зерттеулердің 4 бағыты іріктелді, олардың бірі экспортқа бағдарланған ұлттың денсаулығын сақтауды және арттыруды қамтамасыз ететін экологиялық таза ауыл шаруашылығы өнімін жасауға бағытталған зерттеулер болып табылады, ауыл шаруашылығының заманауи технологияларын әзірлеуге және жетілдіруге бағытталған осы пәнаралық зерттеулер процесінде түпкілікті нәтиже экологиялық таза жоғары сапалы азық-түлік өнімдерін өндіру технологиялары және толық азық-түлік қауіпсіздігі жағдайында халықтың өмір сүру стандарттарының жоғары деңгейін қамтамасыз ету болады. Олар елдің экспорттық әлеуетін кеңейтуді қамтамасыз етуге мүмкіндік береді, халықтың денсаулығын жақсартуды қамтамасыз ететін жаңа пайдалы қасиеттері бар жоғары сапалы азық-түлік өнімдерін жасау. Бұл бағыттарды биотехнология саласы қамтиды. Бұл нәтижеге қол жеткізу үшін келесі тақырыптық бағыттар бойынша ғылыми зерттеулер мен әзірлемелер жүргізу қажет екендігі анықталды:

- Ауылшаруашылығышикізатынөңдеудіңжоғарытиімдітехнологияларында қазіргі заманғы физика-химиялық және электрофизикалық әдістерді әзірлеу;

- Ауыл шаруашылығы шикізатын, оның ішінде қайталама шикізатты өңдеу кезінде нано - және биотехнология әдістерін пайдалана отырып технологиялық процестерді әзірлеу;

- Тағамды жобалаудың заманауи әдістері негізінде халықтың әртүрлі жас топтары үшін (балалар тағамы, мектепке дейінгі, мектеп, геродиетикалық және т.б.) жаппай тұтыну өнімдерін жасау техникасы мен технологиясын әзірлеу.

Осындай тақырыптарды қамтып, жоспарсыз жүру мүмкін емес, осы себептер биотехнология саласына жобаларды басқару өте маңызды екенін атап өту қажет. Бүгінгі таңда ауыл шаруашылығы шикізатын, оның ішінде қайталама шикізатты тереңдетіп өңдеудің инновациялық технологиялары биотехнологияға негізделген. Бірқатар ферменттік препараттарды қолдану қосымша құны жоғары өнімдерді алуға мүмкіндік береді. Ауыл шаруашылығы өндірісінің қалдықтары ақуыз, органикалық қышқылдар, бірқатар құнды өнімдер және т. б. алу үшін негіз ретінде пайдаланылуы мүмкін[9].

Сапалы нәтижелерге қол жеткізу үшін тақырыптық нәтижелерді де іске асыру қажет "Биотехнология" секторының бағыттары:

Фармацевтикалық, азық-түлік және т.б. үшін бағытталған биокатализ технологияларын әзірлеу;

- қайта өңдеу өнеркәсібі және ауыл шаруашылығы;

- этномәдени ескеріле отырып, азық-түлік, өңдеу өнеркәсібі және ауыл шаруашылығы үшін ашытқылар жасау мақсатында микроорганизмдердің жаңа консорциумдарын әзірлеу және аймақтық ерекшеліктер.

2021 жылы Мәскеуде Беларусь, Қазақстан Республикаларының және Ресей Федерациясының Еуразиялық экономикалық комиссия (ЕЭК) туралы шартына қол қойылғанына он жыл толды. Ресми дереккөздерден көрініп тұрғандай, осы уақыт ішінде жоғары еңбек өнімділігімен де, зияткерлік компоненттің жоғары үлесімен де ерекшеленетін салаларда өнеркәсіптік кооперацияның жаңа бағыттарын іздеу жүріп жатыр: биотехнология, жасанды интеллект, роботтандыру және т.б. (Трофимов, 2012).

Осының ішіне жобаларды басқаруда кіреді, қазіргі уақытта Қазақстан Республикасының ұлттық стандарты "ҚР СТ ISO 21500–2014 жобаларын басқару жөніндегі Нұсқаулық" 2014 жылы әзірленіп, 2016 жылғы 1 қаңтардан бастап күшіне енді. Осылайша, жобалық менеджмент еліміздегі жаңа бағыт болып табылады. Қазақстанда жобаларды басқаруды негізінен PMI PMBOK(r) және IPMA бойынша зерттейді және жүргізеді. Жобаларды басқару ақпараттық технологияларсыз мүмкін емес. Сол себепті биотехнология саласында жобаларды басқару үшін ақпараттық технологияның маңызы өте үлкен. Ақпараттық технологиялар саласындағы Жобаларды басқару мәселелері биотехнологияның басқа салаларымен қатар, дайын технологиялық шешімдерді Қазақстан Республикасындағы қызметтің әртүрлі салаларына бейімдеу үрдісі басталған кезден бастап Қазақстанда өзекті бола бастады. Өткен ғасырдың 90-жылдарының басында және ортасында Қазақстанда банктік, мұнай, өндіруші, өндірістік, телекоммуникациялық, құрылыс және өзге де коммерциялық ұйымдар компьютерлік және офистік техниканы, кейбір жағдайларда серверлік жабдықтарды сатып алумен шектелді, оларды енгізумен шетелдік ұйымдар жиі айналысатын. Олар өз кезегінде дамыған елдердің тиісті қызмет салаларында жұмыс істеген дайын шешімдерді ұсынды және алға тартты. Бизнес-процестерді автоматтандыру және ақпараттық жүйелерді енгізу қажеттілігі, өткен ғасырдың 90-шы жылдарының аяғында ғана пайда болды. Оның бастамасы мемлекеттік органдардың, компаниялардың, банктердің және ірі коммерциялық ұйымдардың қызметіне бейімделген, халықаралық стандарттар негізінде жобаларды басқару жөніндегі басшылар мен менеджерлерді даярлау туралы мәселені көтере бастады. Жобаларды басқару стандарттарын ауқымты тарату мәселесі 2000 жылдардың басында актуалды болды. Ұлттық компаниялар мен ірі ұйымдар, сондай-ақ қызметі тікелей немесе жанама түрде жобаларды іске асыруды қамтитын Қазақстанның мемлекеттік органдары корпоративтік курстар ұйымдастырып, өз қызметкерлерін жобаларды басқару стандарттарының негіздеріне үйрете бастады. Қазіргі уақытта Қазақстанда жыл сайын бірнеше

мың маман менеджмент бойынша біліктілігін арттырады олардың 20-30% - ы негізінен PMI PMBOK(r) және IPMA бойынша сертификатталған (Ұлттық биотехнология орталығы, 2011).

Негізгі мәселе бойынша Қазақстандағы биотехнологияны жобалық басқаруға көшіру керек екендігі жайында сөз қозғайтын болсақ, ең алдымен оның себептерін(кесте 1) қарастыру керек.

Жобалық менеджментке көшу ең алдымен әлемдік нарықта бәсекеге қабілеттілікті арттырады.

Биотехнология саласындағы жобаларды жобалық басқаруға көшу қажеттілігі	
Қазақстандағы даму деңгейін көтеру	Әлемдік нарықтағы бәсекелестікті арттыру
Барлық континент бойынша бірегей өнімдер	Жұмыстық әлдеқайда қымбатқа түсуі
Сапаның жоғары болуын қатаң талап ету	Жұмыстың үлкен көлемі
Биотехнология саласының жобалық тарауының төмен деңгейі	Комерциялық жобалардың аздығы

Кесте – 1. Биотехнология саласындағы жобаларды жобалық басқаруға көшу қажеттілігі

*Биотехнология саласындағы жобаларды басқарудағы бағдарламалық жасақтаманың маңызы.*

Алдыңғы бөлікте Қазақстанның алдында әлемдік нарықта өз орнын табу үшін биотехнология саласын қолданып, 4 бағытта жұмыс істеу керек екендігін атап көрсеттік. Бұл бағыттарды дұрыс әрі сауатты түрде іске асыру үшін міндетті түрде жобаларды басқаруды енгізу қажет. Ал жобаларды басқару өз уақытында ақпараттық технологиясыз іске асырылмайды. Осы себептерден биотехнология саласындағы жобаларды басқарудың бағдарламалық жасақтамасы өте маңызды.

Биотехнология саласындағы жобадағы ақпараттық технологиялар — жобада пайдаланылатын ақпаратты жинау, сақтау, іздеу, өңдеу, көрсету және беру процестерінің жиынтығы. Биотехнологияда жобаны басқару үшін бірыңғай ақпараттық жүйені құру қажет, өйткені жобаны басқару процесінде басқарудың әртүрлі деңгейлерінде ақпарат алмасу жүреді.

Ақпараттық жүйе өз кезегінде басқару процесінің барлық функциялары үшін және олардың құзыреті мен жауапкершілігіне сәйкес жобаның барлық қатысушыларының мүдделері үшін барлық кезеңдерде жобаларды басқару нәтижесінде алынған ақпаратты жинауға, сақтауға, өңдеуге және таратуға арналған құралдар мен технологияларды қамтиды. Жобаны басқарудың ақпараттық жүйесінің басқалардан, мысалы, корпоративтік ақпараттық жүйелерден түбегейлі айырмашылығы-корпоративтік ақпараттық жүйелердің көпшілігі жеке функцияларды қолдау үшін әзірленеді. Мұндай жүйелер компанияның бөлімшелері бойынша құрылымдалған, ал жобаны басқарудың ақпараттық жүйесі белгілі бір жобаға қатысты әртүрлі бөлімшелер мен ұйымдардың деректерін біріктіреді.

Жобаның орындалу барысын жоспарлау және бақылау, сондай-ақ жоба бойынша шешім қабылдаушыларды қажетті және жеткілікті ақпаратпен



қамтамасыз ету үшін жобаның ақпараттық моделін әзірлеу және өзекті жағдайда қолдау қажет, ол мыналарды қамтамасыз етеді:

- жобаның күнтізбелік жоспарлары, ресурстары, құндық және басқа да көрсеткіштері бойынша ақпаратты орталықтандырылған сақтау;
- жұмыс жоспарындағы, ресурстық, қаржылық және басқа да қамтамасыз ету түрлеріндегі өзгерістердің жобаның түпкілікті нәтижелері мен көрсеткіштеріне әсерін жылдам талдау мүмкіндіктері;
- ақпараттық жүйенің жұмыс істеуінің желілік режимінде таратылған қолдау және деректерді жанарту мүмкіндігі;
- есептер мен графикалық диаграммаларды автоматтандырылған генерациялау, жоба бойынша құжаттаманы әзірлеу, сондай-ақ Жобаны басқарудың басқа да міндеттері мен рәсімдерін шешу мүмкіндіктері.

Биотехнология саласындағы жобаның ұйымдастырушылық құрылымында мамандандырылған ақпараттық қолдауды қажет ететін басқарудың кем дегенде үш деңгейі бөлінуі мүмкін:

- жобаны басқарудың стратегиялық деңгейі;
- жеке жобаны басқару деңгейі;
- жоба жұмыстарының орындалу деңгейі.

Биотехнология саласындағы жобаны басқару деңгейінде ақпараттық жүйе жұмыстар кешенін жоспарлауды, жұмыстардың орындалуын ұйымдастыруды және бақылауды, жобаның орындалу барысын талдау мен реттеуді және жобаның жабылуын қамтамасыз етеді және қолдайды. Басшылықтың бұл деңгейі, ең алдымен, басқарудың басқа деңгейлерімен деректер алмасуды және талдау мен жедел басқару үшін есептер алуды қамтамасыз ететін әртүрлі кіріс параметрлері бар модельді есептеуді қолдайтын жұмыс кешендері мен жоба ресурстарының барабар ақпараттық моделін құруға мүмкіндік беретін қуатты құралдарға қызығушылық танытады.

Биотехнологияда жобаның тапсырмаларын (жұмыс пакеттерін) орындау деңгейінде жұмыстардың орындалуын реттейтін және қамтамасыз ететін егжей-тегжейлі ақпарат қажет, себебі бұл салада жаңалықтар өте ауқымды қаралады. Бұл ақпарат жобаны басқару деңгейінен және функционалды бөлімшелерден келеді. Сонымен қатар, осы деңгейде жұмыстардың барысы мен ресурстарды пайдалану туралы нақты деректер жиналып, жоғары тұрған басшылыққа беріледі. Қазіргі уақытта жергілікті және жаһандық желілерде жобаға қатысушылар арасындағы байланысты қамтамасыз ететін заманауи технологиялардың пайда болуының арқасында жобаны басқарудың (немесе жобалар кешенінің) таратылған интеграцияланған жүйесі тұжырымдамасын іске асыру, нақты уақыт режимінде өзекті ақпаратты жинау және тарату мүмкін болды.

Биотехнология саласындағы жобаны басқарудың ақпараттық жүйесін енгізу нәтижесінде мынадай міндеттерді шешу қамтамасыз етіледі:

Биотехнологиялық маңызы бар жобаның аралық нәтижелеріне қол жеткізудің құрылымы, жұмыс көлемі және директивалық мерзімдері бойынша талдау және шешім қабылдау рәсімдерін қолдау;

Әлемдік нарыққа шығу үшін қаржыландыру мәселесін қамтитын жобаның және оның жекелеген фазаларын іске асыруға қажетті уақыт пен шығындарды анықтау, сондай-ақ жобаның қаржылық қажеттіліктерін уақыт бойынша бөлу;

- жобаның жиынтық жоспарын әзірлеу және жобаның негізгі қатысушыларымен уақытша, көлемдік, құндық және басқа да параметрлер бойынша шешімдерді келісу рәсімдерін қолдау;

- жұмыстардың орындалу барысы және олардың сипаттамалары туралы нақты деректерді жинау (мысалы, орындалған көлемдер, қаржылық, еңбек, материалдық және басқа да ресурстар түрлерінің жұмсалған шығындары);

- өзара әрекеттесу және жұмысты үйлестіру процестерін қолдау, орындалуын бақылау және жедел басқару;

- бірыңғай жұмыс жоспарындағы өзгерістер бойынша шешімдерді қабылдау және келісу, келісілген шешімдерге сәйкес жасалған жұмыстарды орындау жоспарларын жедел түзету;

- жоба шеңберінде ішкі және сыртқы есептілікті қолдау.

Биотехнология саласындағы жобаны басқарудың заманауи автоматтандырылған ақпараттық жүйелері келесі функциялардың орындалуын қамтамасыз етуі керек:

- жұмыстарды күнтізбелік жоспарлау;

- ресурстарды жоспарлау;

- жоба операцияларын орындау уақытының сыни жолы мен резервтерін есептеу;

- жобаның қаржыландыру, материалдар мен жабдықтарға қажеттілігін есептеу;

- тәуекелдерді талдау және тәуекелдерді жоспарлау;

- жобаның орындалуы бойынша нақты деректерді есепке алу;

- есепті материалдарды дайындау;

- аумақтық қашықтағы пайдаланушыларға қол жеткізу;

- құжаттардың (деректердің) орталықтандырылған қоймасы-білім банкі;

- ұжымдық (бірлескен) жұмыс.

Жұмыс процесінің сызбасы 1-суретте көрсетілген. Дәстүрлі менеджменттің 5 кезеңі:

1 кезең. Бастама. Жоба менеджері мен команда жобаға қойылатын талаптарды анықтайды. Бұл кезеңде кездесулер

мен "ми шабуылдары" жиі өткізіледі, онда жобаның өнімі қандай болуы керек екендігі анықталады.

2 кезең. Жоспарлау. Бұл кезеңде команда алдыңғы кезеңде қойылған мақсатқа қалай жететінін шешеді. Бұл кезеңде команда жобаның мақсаттары мен нәтижелерін, сондай-ақ ол бойынша жұмыстардың құрамын нақтылайды және егжей-тегжейлі көрсетеді. Осы ақпарат негізінде команда күнтізбелік жоспар мен бюджетті қалыптастырады, тәуекелдерді бағалайды және мүдделі тараптарды анықтайды.

3 кезең. Даму. Бұл кезең барлық жобалар үшін жүзеге асырылмайды — әдетте бұл жоспарлау кезеңінің бөлігі болып табылады. Технологиялық жобаларға тән даму кезеңінде болашақ жобаның және/немесе өнімнің конфигурациясы және оған жетудің техникалық әдістері анықталады. Мысалы, ат жобаларында осы кезеңде бағдарламалау тілі таңдалады. (Отандық тәжірибеде бұл кезең әдетте ерекшеленбейді, ал "Даму" термині қолданылмайды — шамамен. транс.)

4 кезең. Іске асыру және тестілеу. Бұл кезеңде жобаның негізгі жұмысы-код жазу, ғимарат салу және т.б.

Әзірленген жоспарларға сүйене отырып, бұрын анықталған жобаның мазмұны құрыла бастайды, таңдалған көрсеткіштер бойынша бақылау жүргізіледі. Осы кезеңнің екінші бөлігінде өнім сыналады, ол Тапсырыс беруші мен мүдделі тараптардың талаптарына сәйкестігі тексеріледі. Тестілеу бөлігінде өнімнің кемшіліктері анықталады және түзетіледі.

5 кезең. Жобаны бақылау және аяқтау. Жобаға байланысты бұл кезең Тапсырыс берушіге жобаның нәтижелерін қарапайым түрде беруден немесе жобаны жақсарту және олардың қанағаттанушылығын арттыру және жоба нәтижелерін қолдау үшін клиенттермен өзара әрекеттесудің ұзақ процесінен тұруы мүмкін. Соңғысы клиенттерге қызмет көрсету және бағдарламалық қамтамасыз ету саласындағы жобаларға қатысты.



1-сурет. Жобалық басқарудың негізгі кезеңдері.

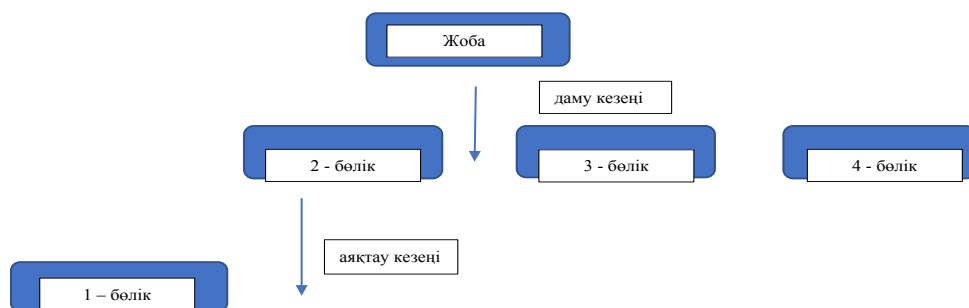
*Биотехнология саласындағы жобаларды басқарудағы тәсілдер, әдістер мен стандарттар.* Жоба саласындағы кәсіпқойлар мен практиктер қазірдің өзінде әртүрлі деңгейдегі жобаларға арналған көптеген әдістерді қалыптастырды. Ал биотехнология саласындағы арнайы жобаларды басқару әдістері болмасада, осы саладағы әдістерді қолдануға болады. Ең бастысы маңызды нәрсені түсіну: бюджет, мерзімдер, кезеңдерді сақтау және қызметтерді басқару – ол барлығына ортақ мәселе. Осы мәселелер қамтылса ғана кез келген жобаны тиімді басқаруға болады.

Бастау үшін ең танымал және тиімді басқару әдістерін қарастырайық:

- Agile
- Scrum
- PRINCE2

## 1. Agile

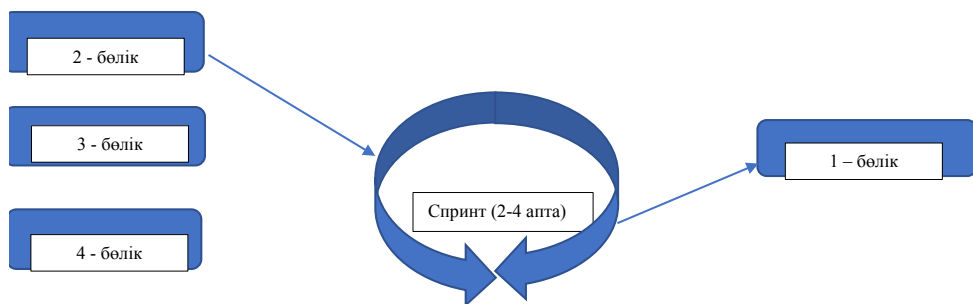
Егер сіздің биотехнология саласындағы жобада бір уақытта бірнеше қадамдарды орындау қажет болса онда Agile (2 – сурет) әдісін қолдану дұрыс. Жоспарлау және іске қосу бүкіл жоба үшін жүзеге асырылады, ал әзірлеу, тестілеу және т.б. әр бөлік үшін бөлек жүргізіледі. Бұл тәсіл шағын жобалардың нәтижелерін жедел ілгерілетуге және бюджетке үлкен шығынсыз түзетулер енгізуге және жобаның қалған бөліктеріне уақыт бөлуге мүмкіндік береді.



2 – сурет. Agile әдісінің кезеңдері

## 2. Scrum (3-сурет)

Классикалық менеджмент принциптері мен икемді тәсіл келісімдерін біріктіру қажет болса, жобаны басқарудың тамаша әдісі. Бұл тәсіл жобаға арналған тапсырмаларды құрылымдауда тепе-теңдікке және оларды өзгерту кезінде икемділікке ие. Сондай-ақ, Agile сияқты, бұл тәсіл бүкіл процесті бірден іске асыруға болатын шағын жобаларға бөледі. Менеджер қандай шағын жобалардың ең маңызды екенін шешеді және оларды шамамен 4 аптаға созылатын "Спринт" кезеңіне жібереді. "Спринттің" соңында клиентке бүкіл өнімнің жұмыс "бөлігі" беріледі. PR стратегиясын әзірлеу шеңберінде PR маманы стратегияны ұсынғанға және іске асырғанға дейін пайдалану үшін компания туралы материалдарды қалыптастырған кезде мысал келтіруге болады. Спринттің ұзақтығы тұрақты, бірақ команда оны жобаның басында, оның ерекшелігіне қарай таңдайды.



2 – сурет. Scrum әдісінің кезеңдері

### 3. PRINCE2

Ұлыбританиядан шыққан әдіс, ол "Projects IN Controlled Environments version 2" ("бақыланатын ортадағы жобалар, 2-нұсқа") дегенді білдіреді. Бұл тәсіл ауқымды мемлекеттік жобалар мен ірі ұйымдарға бағытталған. Ол әр кезеңде жобаның барлық қатысушылары арасында рөлдерді бөлуге негізделген. PRINCE2 қағидаттарының бірін – ерекше жағдайларды басқару қағидатын жүзеге асырады. Менеджер жобаның мерзімдеріндегі, бюджетіндегі, жалпы бағамындағы өзгерістерді қадағалайды және егер бұл оның өкілеттігінде болмаса, басқарушы Комитетке есеп беріп, түзету нұсқаларын ұсынуы керек.

Бұл процесс жоба менеджері мен топ менеджерінің бір немесе бірнеше тапсырмаларды орындау бойынша өзара әрекеттесуін қамтиды. Жоба менеджері тапсырмаларды орындауды топ менеджеріне тапсырады және одан жұмыс нәтижелерін қабылдайды.

Қазіргі таңда тек биотехнология саласына арнайы жасалған бағдарламалық жасақтама жоқ, бірақ келесідей жүйелер қолданыста Microsoft Office Project, Primavera және Project Expert.

Microsoft Office Project - бұл әмбебап күнтізбелік жоспарлау және жобаларды басқару жүйесі. Жеке жоспарлауға және жобаларды басқаруға арналған жұмыс үстелі қосымшасы. Microsoft Project пайдаланушыларға орталықтандырылған параметрлерді, бірыңғай ресурстар пулын, жоба қатысушыларының бірлесіп жұмыс істеуі үшін веб-интерфейсті қамтамасыз етеді және OLAP талдау және жоба портфолиосын модельдеу құралдарын қамтиды.

Project Management корпоративтік ақпараттық жүйенің құрамында да, автономды түрде де пайдалануға арналған, күнтізбелік-желілік жоспарлау, сыни жолды есептеу, ресурстарды теңестіру, what-if талдау және жобаларды, жобалар топтарын, портфельдер мен бағдарламаларды модельдеудің басқа да міндеттерін шешуді қамтамасыз етеді. Methodology Management жобаны басқару бойынша компанияның білім базасын қалыптастыруды және оған қол жеткізуді қамтамасыз етеді.

Primavera бағдарламалық жасақтамасы (Primavera Systems, Inc.) PMI, IPMA және ISO стандарттарына сәйкес жобаларды басқару процестерін автоматтандыруға арналған (Шетелдегі ғылым, 2011).

Бүгінгі таңда ең танымал бағдарламаларды қарастырайық:

1. Projecto
2. Trello
3. Asana
4. Битрикс24
5. Microsoft Project
6. Plan.io

	Projecto	Trello	Asana	Битрикс24	Microsoft Project	Plan.io
Ұйымдастыру құрылымы	+	-	+	+	-	-
Бақылау тақтасы және виджеттер	+	+	+	+	-	-

Күнтізбе	+	-	-	+	+	+
Жұмыс уақытын салыстыру	+	-	-	+	-	-
Міндеттерді қою және бақылау	+	+	+	+	+	+
Басшының атынан міндеттер құру	+	-	-	-	-	-
Жобаларды жүргізу	+	+	+	+	+	+
Құжаттар мұрағаты	+	+	+	+	-	-
Құжат айналымы және құжаттарды келісу	+	+	+	+	-	-
Хаттарды тіркеу	+	+	-	+	-	-
Жазбалар бойынша бірлескен жұмыс	+	-	+	-	-	-
Ортақ файлдар	+	+	+	+	+	+
Жұмыстан шығарылған қызметкерлердің деректерін тасымалдау	+	-	-	-	-	-
Қазақ тілі	-	-	-	-	-	-

2 – кесте. Нарықтағы танымал бағдарламалық жасақтамалардың салыстырмалы кестесі.

### **Қорытынды**

Көріп отырғанымыздай бұл бағдарламалардың барлығында қазақ тілі орнатылмаған, бұл шешілу керек мәселелердің бірі деп ойлаймын. Биотехнология саласы туралы көптеген ақпараттар таралған. Алайда, бұл ақпараттар жеткілікті түрде жинақталып, жүйеленбеген, яғни жобалық деңгейге көтеріліп, жобаларды басқару игерілмей қалған. Бұл жағдайда құрылымдалмаған білім саласын көруге болады. Осыған байланысты келесідей қажеттілік туындайды: биотехнология саласындағы жобаларды басқаруды ауқымды түрде жүргізу үшін бағдарламалық қамтамасыз ету қарастыру. Осындай қажеттіліктер мамандар, ғылыми қызметкерлер, білім саласы үшін таптырмас құрал және әдіс жиынтығы болады. Жеке және мемлекеттік кәсіпкерлер үшін биотехнология саласындағы инвестицияға, дамытуға алып келеді. Сонымен қатар Қазақстанның алдында әлемдік нарықта өз орнын табу үшін биотехнология саласын қолданып, 4 бағытта жұмыс істеу керек екендігін атап көрсеттік. Бұл бағыттарды дұрыс әрі сауатты түрде іске асыру үшін міндетті түрде жобаларды басқаруды енгізу қажет. Ал жобаларды басқару өз уақытында ақпараттық технологиясыз іске асырылмайды. Осы себептерден биотехнология саласындағы жобаларды басқарудың бағдарламалық жасақтамасы өте маңызды. Келісінген жүйе бойынша биотехнология саласына арнайы жасалынған жобаларды басқарудың бағдарламалық жасақтамасы дайындалуы қажет.

### **ӘДЕБИЕТТЕР**

- Handbook of Genetics & Biotechnology, Revised and Expanded 2nd Edition, 2014.
- "Ауыл шаруашылығы және тамақ өнімдерін өндіру саласындағы биотехнологиялық өнім (трансгенді қоспағанда) нарығының әлеуетін айқындау" жобасы бойынша есеп. (2007) - "BARS Consulting Group" ЖШС. – 2007.
- Tomar S. (2014) Handbook of Genetics Biotechnology -Handbook, 2014.
- Әлемде және Қазақстан Республикасында инновациялардың даму үрдістері туралы есеп. (2011) - Астана: "НИФ"АҚ. – 2011.

БҰҰ-ның (2012)"Қазақстанның инновациялық дамуына шолу"шолуы.- Женева. – 2012. – 211 б.

Блинов Н.П. (1995) Биотехнология негіздері. "Ғылым" баспа фирмасы, Санкт-Петербург, 2015 ж.-11 б.

Додонов В.Ю (2011). Әлемдік экономика мен Қазақстанның 2030 жылға дейінгі дамуының негізгі тенденциялары. -Еуразиялық экономикалық комиссияның ресми сайты. [Электрондық ресурс]. URL: <http://www.eurasiancommission.org/ru> Алматы: ҚР Президенті жанындағы ҚСЗИ, 2011. - 56 б.

Жобаларды басқару бойынша білім жинағына Нұсқаулық (PMBOK® нұсқаулығы) - бесінші басылым / PMI Standards Committee;

Microsoft 365 ресми сайты [Электрондық ресурс]: URL: <https://www.microsoft.com/ru-ru/microsoft-365/project/project-management-software?market=kz>

Primavera - жобаны басқару бағдарламалық құралы [Электрондық ресурс]: URL: <https://ru.wikipedia.org/wiki/Primavera>

Раманқұлов Е.М. (2008). Биотехнология. Әлемдегі және Қазақстандағы үрдістер. – Vestnik\_14\_2008.

Трофимов В.В. (2012). Жобаларды басқару: Оқу құралы / В.В. Трофимов. - 2-ші шығарылым Ұлттық биотехнология орталығы (2011) [Электрондық ресурс]: <https://www.biocenter.kz/kz/biz-turaly/ubo-turalymalimetter>

Штетелдегі ғылым. "Биотехнология және генетика. Ауыл шаруашылығы, тамақ және химия өнеркәсібі". – М.: РҒА ғылымды дамыту мәселелері институты. – № 1, 2011. – 29 б.

#### REFERENCES

Report on the project "Determination of the market potential of biotechnological products (except transgenic) in the field of agriculture and food production" (2007). – BARS Consulting Group LLP. – 2007.

Tomar S. (2014). Handbook of Genetics Biotechnology -Handbook, 2014.

Report on trends in the development of innovations in the world and in the Republic of Kazakhstan. (2011) – Astana: JSC "NIF". – 2011.

UN Review (2012). "Review of innovative development of Kazakhstan". – Geneva. - 2012. – 211 p.

Blinov N.P. (1995). Fundamentals of biotechnology. Publishing company "Nauka", St. Petersburg, 1995.-11 p.

Dodonov V.Yu. (2011). The main trends in the development of the world economy and Kazakhstan until 2030. – Almaty:

The official website of the Eurasian Economic Commission. [electronic resource]. (2011). URL: <http://www.eurasiancommission.org/ru> (accessed: 12/21/2021).KISIunder the President of the Republic of Kazakhstan, 2011. – 56 p.

Project Management Knowledge Guide (PMBOK® Guide) — fifth Edition / PMI Standards Committee;

Official Microsoft 365 website [Electronic resource]: URL: <https://www.microsoft.com/ru-ru/microsoft365/project/project-management-software?market=kz>

Primavera — project management software [Electronic resource]: URL: <https://ru.wikipedia.org/wiki/Primavera>

Ramankulov E.M. (2008). Biotechnology. Trends in the world and Kazakhstan. – Vestnik\_14\_2008.

Trofimov V.V. (2012). Project management: textbook / V.V. Trofimov. – 2nd National Center of Biotechnology [Electronic resource]: <https://www.biocenter.kz/kz/biz-turaly/ubo-turaly-malimetter>

Science abroad. The heading "Biotechnology and genetics. Agriculture, food and chemical industry"(2011). – М.: Institute of Problems of Development of Science of the Russian Academy of Sciences. – No. 1, 2011. – p.29

© **A.A. Taurbekova**<sup>1\*</sup>, **O.Zh. Mamyrbayev**<sup>2</sup>, **B.T. Karymsakova**<sup>3</sup>,  
**B. Zh. Zhumazhanov**<sup>4</sup>, 2023

<sup>1</sup>Kazakh National Research Technical University named after K.I. Satpayev,  
Almaty, Kazakhstan;

<sup>2</sup>Institute of Information and Computing Technologies, Almaty, Kazakhstan;

<sup>3</sup>Taraz Regional University named after M.H. Dulaty, Taraz, Kazakhstan;

<sup>4</sup>Institute of Information and Computing Technologies, Almaty, Kazakhstan.

E-mail: [ainura\\_071@mail.ru](mailto:ainura_071@mail.ru)

## INVESTIGATIONS OF MAGMA OUTPUT PROCESS

**Mamyrbayev Orken Zhumazhanovich** — PhD, Institute of Information and Computational Technologies, Almaty, Kazakhstan

E-mail: [morkenj@mail.ru](mailto:morkenj@mail.ru); <https://orcid.org/0000-0002-8627-1949>;

**Taurbekova Ainur Adilgaziievna** — doctoral student of Kazakh National Research Technical University named after K. I. Satpayev, Almaty, Kazakhstan

E-mail: [ainura\\_071@mail.ru](mailto:ainura_071@mail.ru), <https://orcid.org/0000-0001-5433-8347>;

**Karymsakova Balnur Tleetaevna** — lecturer of Taraz Regional University named after M. H. Dulaty, Taraz, Kazakhstan

E-mail: [tz-tek1@mail.ru](mailto:tz-tek1@mail.ru); <https://orcid.org/0009-0008-9863-7654>;

**Zhumazhanov Bagashar Zhumazhanovich** — candidate of technical sciences, Institute of Information and Computational Technologies, Almaty, Kazakhstan

E-mail: [bagasharj@mail.ru](mailto:bagasharj@mail.ru); <https://orcid.org/0000-0002-5035-9076>.

**Abstract.** The article presents methodological tools for the study of numerical simulation of the process of magma outflow coming through a "narrow channel" from the Earth's interior. The result of a numerical experiment for various parameters of outflowing magma from a "narrow channel" is shown. Based on the results of the experiment, conclusions were drawn about the process of raising heated magma from the lower layers of the Earth. This emphasis makes it possible to reveal a certain originality and scientific novelty in the development of methods for studying seismic shock processes. And the development of methods, in turn, made it possible to obtain new scientific results, which are also presented in this article.

**Keywords:** complex process, asthenosphere, lithosphere, mantle flow, magma, Navier-Stokes equations, Reynolds number, mathematical model, numerical experiment, plume, seismic shock processes



© А.Ә. Таурбекова<sup>1</sup>, Ө.Ж. Мамырбаев<sup>2</sup>, Б.Т. Қарымсақова<sup>3</sup>,  
Б.Ж. Жұмажанов<sup>4</sup>, 2023

<sup>1</sup>Қ.И. Сәтбаев атындағы Қазақ ұлттық техникалық зерттеу университеті,  
Алматы, Қазақстан;

<sup>2</sup>Ақпараттық және есептеуіш технологиялар институты, Алматы, Қазақстан;

<sup>3</sup>М.Х. Дулати атындағы Тараз өңірлік университеті оқытушысы,  
Тараз, Қазақстан;

<sup>4</sup>Ақпараттық және есептеуіш технологиялар институты, Алматы, Қазақстан.  
E-mail: ainura\_071@mail.ru

## МАГМАНЫҢ ШЫҒУ ПРОЦЕСІН ЗЕРТТЕУ

**Мамырбаев Өркен Жұмажанұлы** – PhD, Ақпараттық және есептеу технологиялары институты, Алматы, Қазақстан

E-mail: morkenj@mail.ru ; <https://orcid.org/0000-0002-8627-1949>;

Таурбекова Айнұр Әділғазықызы Қ.И. Сәтбаев атындағы Қазақ ұлттық техникалық зерттеу университетінің докторанты, Алматы, Қазақстан

E-mail: ainura\_071@mail.ru , <https://orcid.org/0000-0001-5433-8347>;

**Қарымсақова Балнұр Тлетайқызы** – М. Х. Дулати атындағы Тараз өңірлік университетінің оқытушысы, Тараз, Қазақстан

E-mail: tz-tek1@mail.ru. <https://orcid.org/0009-0008-9863-7654>;

**Жұмажанов Бағашар Жұмажанұлыт** – техника ғылымдарының кандидаты, Ақпараттық және есептеу технологиялары институты, Алматы, Қазақстан

E-mail: bagasharj@mail.ru; <https://orcid.org/0000-0002-5035-9076> .

**Аннотация.** Мақалада Жердің ішкі бөлігінен «тар арна» арқылы келетін магманың шығу процесін сандық модельдеуді зерттеудің әдістемелік құралдары ұсынылған. «Тар арнадан» ағып жатқан магманың әртүрлі параметрлері бойынша сандық тәжірибенің нәтижесі көрсетілген. Тәжірибе нәтижелері бойынша Жердің төменгі қабаттарынан қыздырылған магманың көтерілу процесі туралы қорытындылар жасалды. Бұл екпін сейсмикалық сілкініс процестерін зерттеу әдістерін әзірлеуде белгілі бір ерекшелік пен ғылыми жаңалықты ашуға мүмкіндік береді. Ал әдістердің дамуы, өз кезегінде, осы мақалада да берілген жаңа ғылыми нәтижелерді алуға мүмкіндік берді.

**Түйін сөздер:** күрделі процесс, астеносфера, литосфера, мантия ағыны, магма, Навье-Стокс теңдеулері, Рейнольдс саны, математикалық модель, сандық тәжірибе, шлейф, сейсмикалық соққы процестері

© А.А. Таурбекова<sup>1\*</sup>, О.Ж. Мамырбаев<sup>2</sup>, Б.Т. Карымсакова<sup>3</sup>,  
Б.Ж. Жумажанов<sup>4</sup>, 2023

<sup>1</sup>Казахский национальный исследовательский технический университет  
имени К.И. Сатпаева, Алматы, Казахстан;

<sup>2</sup>«Институт информационных и вычислительных технологий»,  
Алматы, Казахстан;

<sup>3</sup>Таразский региональный университет им. М. Х. Дулати, Тараз, Казахстан;

<sup>4</sup>«Институт информационных и вычислительных технологий»,  
Алматы, Казахстан.

E-mail: ainura\_071@mail.ru

## ИССЛЕДОВАНИЯ ПРОЦЕССА ИСТЕЧЕНИЯ МАГМЫ

**Мамырбаев Оркен Жумажанович** — PhD, Институт информационных и вычислительных технологий, Алматы, Казахстан

E-mail: morkenj@mail.ru, <https://orcid.org/0000-0002-8627-1949>;

**Таурбекова Айнура Адилгазыевна** — докторант Казахского национального исследовательского технического университета имени К. И. Сатпаева, Алматы, Казахстан

E-mail: ainura\_071@mail.ru, <https://orcid.org/0000-0001-5433-8347>;

**Карымсакова Балнур Глетаевна** — преподаватель Таразского регионального университета имени М.Х. Дулати, Тараз, Казахстан

E-mail: tz-tek1@mail.ru, <https://orcid.org/0009-0008-9863-7654>;

**Жумажанов Багашар Жумажанович** — кандидат технических наук, Институт информационных и вычислительных технологий, Алматы, Казахстан

E-mail: bagasharj@mail.ru, <https://orcid.org/0000-0002-5035-9076>.

**Аннотация.** В статье приведен методический инструментарий исследования, численное моделирование процесса истечения магмы, поступающих по «узкому каналу» из недр Земли. Показан результат численного эксперимента для различных параметров истекающей магмы из «узкого канала». По результатам эксперимента сделаны выводы о процессе поднятия разогретой магмы из нижних слоев Земли. Такой акцент позволяет выявить определенную оригинальность и научную новизну в развитии методов исследования процессов сейсмического толчка. А развитие методов, в свою очередь, позволило получить новые научные результаты, которые также приводятся в настоящей статье.

**Ключевые слова:** сложный процесс, мантийный поток, магма, уравнения Навье - Стокса, число Рейнольдса, математическая модель, процессов сейсмического толчка

### Введение

Известно, что одной из проблем в научных исследованиях является изучение сложных процессов, происходящих в земных недрах. Многообразие этих процессов и причин их возникновения требуют использования различных

способов их исследования. Однако, во многих случаях экспериментальные или аналитические исследования связаны с известными трудностями из-за сложности рассматриваемых явлений.

«Общепризнанно, что при изучении многих сложных явлений нельзя ограничиваться экспериментальными и аналитическими исследованиями. Применение компьютерной техники и технологии для решения сложных задач, особенно когда они не имеют простых решений, предполагает поиск численных методов и создания эффективных алгоритмов и программ для их реализации. Использование численных моделей позволяет избавиться от побочных факторов и сконцентрироваться на главном, что бывает нелегко достичь в реальном эксперименте» (Белоцерковский и др., 2008: 512). Поэтому для решения задач, получаемых в результате моделирования рассматриваемого здесь процесса истечения сильновязкой жидкости, может быть использован численный метод.

Предполагается, что движение рассматриваемой здесь жидкости описывается уравнениями Навье - Стокса при очень малых числах Рейнольдса. Для решения задачи, сформулированной для полученных уравнений, используется конечно-разностный метод, применительно для данной задачи.

Из геологических исследований известно, что одной из главных причин структурообразования в недрах Земли является процесс поднятия магмы из нижних слоев Земли. По данным этих исследований (Кузьмин и др., 2016: 11–30; Добрецов и др., 2001: 409), в результате истечения мантийных веществ из нижележащей мантии в астеносфере возникает т.н. плюмы. «Плюмы на поверхности Земли проявляются в виде излияния расплавленной магмы переменного, но обычно щелочного состава» (Добрецов и др., 2001: 409). Возникновение плюма под литосферой является причиной многих явлений, происходящих в земной коре, как вулканизм, сейсмичность и других, оказывающих влияние на структурообразование в верхних слоях Земли (Kyung и др., 2010: 2539–2550).

В связи с этим исследования по направлению изучения явлений, связанных с данным процессом поднятия мантийных веществ и их излияния, являются *актуальными* и имеют важное теоретическое и практическое значение для изучения таких катастрофических явлений как землетрясения, вулканизм и другие.

Сложность изучения этих явлений и недоступность большинства достоверных сведений о процессах, происходящих в глубинных недрах Земли, заставляют исследователей использовать методы математического и компьютерного моделирования.

На основе полученной математической модели поставлена следующая математическая задача, в результате решения которой должны быть определены:

- закон изменения области, образованной вытекающей из «узкого» канала жидкостью;

- движения веществ, составляющих астеносферу;
- напряженное состояние вышележащей литосферы.

*Постановка задачи.* Пусть под некоторой горизонтальной поверхностью имеется «источник», конец «узкого канала», из которого поступает на эту поверхность вещества, имеющие свойства сильновязкой жидкости. «Источник» представляет собой некоторую щель, из которой вытекает сильновязкая жидкость с некоторой заданной скоростью.

Требуется определить движение этой жидкости после истечения из щели, а затем определить область, которую она будет занимать с течением времени.

«Источник» может быть любой формы. С целью упрощения задачи можно считать, что она является круглой с определенным радиусом  $r$ , а также рассматривать задачу в двумерной системе координат  $xOz$ , в которой ось  $z$  направлена вверх, обратно направлению вектора силы тяжести  $\vec{g}$  (Рисунок 1). Начало системы координат  $O$  находится в середине (центра) «источника».

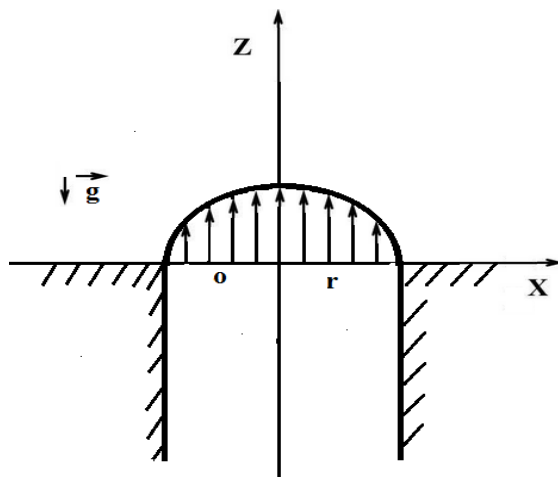


Рисунок 1 – Система координат и конец «узкого канала»

В результате математического моделирования процесса истечения магматических веществ из «узкого канала» в верхней мантии Земли, получена математическая задача о решении следующего квазилинейного неоднородного уравнения параболического типа (Лобковский, 2016: 11–30; Куралбаев, 2008: 212): и конец

$$\frac{\partial u}{\partial t} = \frac{ER}{3} \frac{\partial}{\partial x} \left( u \frac{\partial u}{\partial x} \right) + \begin{cases} v(t)(1-x^2), & \text{если } -1 \leq x \leq 1; \\ 0, & \text{если } -\infty < x < -1, 1 < x < +\infty \end{cases} \quad (1)$$

Здесь:

$x$  - горизонтальная координата;

$t$  – время;

$u = u(x, t)$  – искомая функция, описывающая свободную поверхность области, занимаемой вытекающей из «канала» жидкостью;

$v(t)$  – скорость вытекающей из «канала» жидкости в центре сечения;  
 $ER = \frac{\rho g H^3}{\eta u \tau}$  – безразмерный параметр, зависящий от характерных величин:  
 $H$  – горизонтальный и  $\tau$  – вертикальный размеры рассматриваемой области,  
 $u$  – характерная скорость,  $\rho$  – плотность,  $\eta$  – динамический коэффициент вязкости жидкости,  $g$  – ускорение силы тяжести.

В уравнении (1) все величины являются безразмерными. Для решения его используются условия, вытекающие из сущности постановки задачи.

Начальное условие (при  $t=0$ ) имеет следующий вид:

$$t = 0, u(x, 0) = 0. \quad (2)$$

Граничные условия могут быть заданы с учетом симметрии относительно вертикали:

$$x = 0, \frac{\partial u(0,t)}{\partial x} = 0, \quad (3)$$

$$x = \rho(t), u(\rho(t), 0) = 0. \quad (4)$$

Здесь:  $\rho(t)$  – неизвестная функция, описывающая изменение правой границы области, занимаемой жидкостью. Эта функция является неизвестной, она может быть определена следующей формулой:

$$\int_0^{\rho(t)} u(x,t) dx = \frac{3}{2} \int_0^t v(t) dt. \quad (5)$$

Если будет задана скорость  $v(t)$ , то правая часть этой формулы (5) может быть определена интегрированием. Следует учесть, что в начальный момент времени,

$$\text{при } t = 0, \rho(0) = 0. \quad (6)$$

Итак, здесь требуется решить задачу (1) - (4) для уравнения параболического типа с учетом формул (5) и (6). В данном случае необходимо решить фактически две задачи:

- найти функцию  $\rho(t)$ , определяющую границу области, занимаемой жидкостью;
- найти функцию  $u(x,t)$ , удовлетворяющую уравнению (1) и условиям (2) - (4).

*Метод решения.* Идея метода решения данной задачи заключается в следующем. Для некоторого момента времени находится решение задачи (1) - (4) методом конечных разностей. Для перехода к следующему значению времени  $t+\Delta t$  должно быть найдено значение функции  $\rho(t+\Delta t)$ . Затем решается задача (1) - (4) для нового значения  $\rho(t+\Delta t)$  и т.д.

Пусть вначале рассматривается расчетная схема для решения задачи (1) - (4) методом конечных разностей. Предполагая, что значение функции

(t) для конкретного значения аргумента t фиксированным, используется неявная нелинейная расчетная схема (Тихонов и др., 1977: 736). Выбрав шаги разбиения по аргументам: по t –  $\tau$ , а по x – h, можно записать следующие формулы, заменяющие производные искомой функции конечно – разностными соотношениями:

$$\frac{\partial u}{\partial t} \approx \frac{u_{ij+1} - u_{ij}}{\tau}; \tag{7}$$

$$\frac{\partial}{\partial x} \left( u \frac{\partial u}{\partial x} \right) \approx \frac{1}{h^2} \left\{ \left( \frac{u_{ij+1} + u_{i+1j+1}}{2} \right)^3 \cdot (u_{i+1j+1} - u_{ij+1}) - \left( \frac{u_{i-1j+1} + u_{ij+1}}{2} \right)^3 \cdot (u_{ij+1} - u_{i-1j+1}) \right\} \tag{8}$$

Формула, имеющаяся в правой части уравнения (1), будет заменена следующим соотношением;

$$v(t)(1 - x^2) \approx v_{j+1} (1 - x_i^2). \tag{9}$$

В формулах (7) – (9) введены следующие обозначения:

i - номер точки по координатной оси x;

j - номер, определяющий момент времени t, т.е.  $t = j \cdot \tau$ ;  $j = 0, 1, 2, \dots, m$ ;

$u_{ij} = u(x_i, t_j)$ ;

$x_i = h \cdot i$ ;  $i = 0, 1, 2, \dots, n$ ;

$v_j = v(t_j)$  – известные (заданные) величины.

*Расчетная схема.* Подставляя формулы (7) – (9) в уравнение (1), можно получить следующую дискретную формулу относительно неизвестных значений  $u_{ij}$  искомой функции:

$$u_{ij+1} = u_{ij} + \frac{ER \cdot \tau}{3 \cdot h^2} \left\{ \left( \frac{u_{ij+1} + u_{i+1j+1}}{2} \right)^3 \cdot (u_{i+1j+1} - u_{ij+1}) - \left( \frac{u_{i-1j+1} + u_{ij+1}}{2} \right)^3 \cdot (u_{ij+1} - u_{i-1j+1}) \right\} + \{ v_{j+1} (1 - x_i^2), \text{ если } 0 \leq i \leq k; 0, \text{ если } k + 1 < i < n; \} \tag{10}$$

Здесь  $k = \frac{1}{h}$  - количество точек, где функция отлична от нуля;  $j = 0, 1, 2, \dots, m$ .

Следует отметить, что параметр n является переменной величиной с течением времени.

Формула (10) для конкретного значения j является системой алгебраических уравнений относительно дискретных значений искомой функции  $u_{ij}$ .

Пусть теперь рассматривается формула (5).

Если будет задана функция v(t), т.е. максимальная скорость вытекающей из канала жидкости считается заданной что правую часть формулы (5) можно считать известной.

Пусть предполагается, что

$$f(t) = \frac{2}{3} \int_0^t v(t) dt \quad (11)$$

Эта функция является монотонно возрастающей функцией, т.е.

$$f(t + \Delta t) \geq f(t), \quad \Delta t > 0.$$

Интеграл в левой части формулы (5) может быть записан в виде:

$$\int_0^{p(t)} u(x, t) dx = \int_0^{p(t+\Delta t)} u(x, t) dx - \int_p^{p(t+\Delta t)} u(x, t) dx. \quad (12)$$

При малом значении  $\Delta t$  можно использовать теорему о среднем (Куралбаев, 2008):

$$\int_{p(t)}^{p(t+\Delta t)} u(x, t) dx \approx \underline{u}(t) \cdot [p(t + \Delta t) - p(t)],$$

$$\text{где } \underline{u}(t) = \frac{u(x, t+\Delta t) + u(x, t)}{2}.$$

Из формулы (12) следует, что

$$p(t + \Delta t) - p(t) = \frac{\int_0^{p(t+\Delta t)} u(x, t) dx - \int_0^{p(t)} u(x, t) dx}{\underline{u}(t)}$$

или

$$p(t + \Delta t) - p(t) = \frac{f(t+\Delta t) - f(t)}{\underline{u}(t)}, \quad (13)$$

$$\text{т.к. } \int_0^{p(t+\Delta t)} u(x, t) dx = f(t + \Delta t), \quad \int_0^{p(t)} u(x, t) dx = f(t).$$

Если считать, что  $\Delta t = \tau$  то из формулы (13) следует

$$P_{j+1} = P_j + \frac{2(f_{j+1} - f_j)}{u_{ij+1} + u_{ij}}, \quad (14)$$

где  $P_j = p(t_j)$ ,  $j = 0, 1, 2, \dots, m$ .

Итак, полученные дискретные формулы (10) и (14) определяют расчетную схему для определения значений искомых функций  $u_{ij}$  и  $P_j$ .

*Метод прогонки.* Формула (10) представляет собой систему алгебраических уравнений, для решения которой целесообразно использовать метод итераций, так как уравнения являются нелинейными. С этой целью система уравнений (10) должна быть преобразована в следующий стандартный вид:

$$A_i \cdot y_{i-1} - C_i \cdot y_i + A_{i+1} \cdot y_{i+1} = -F_i. \quad (15)$$

где  $i = 1, 2, \dots, n-1$ ; причем  $n$  может быть переменной. В этой формуле (15) введены следующие обозначения:

$$u_{i-1j-1} = y_{i-1} \quad u_{ij+1} = y_i \quad u_{i+1j+1} = y_{i+1}$$

$$A_i = \frac{ER}{3} \frac{\tau}{h^2} \left(\frac{y_{i-1} + y_i}{2}\right)^3, \quad A_{i+1} = \frac{ER}{3} \frac{\tau}{h^2} \left(\frac{y_i + y_{i+1}}{2}\right)^3; \quad (16)$$

$$C_i = 1 + A_i + A_{i+1}; \quad F_i = -y_i - \{V_{j+1}(1 - x_i^2), \text{ если } 0 \leq i \leq k, 0, \\ \text{если } k + 1 \leq i \leq n$$

Коэффициенты системы уравнений (15)  $A_i, C_{i+1}, F_i$  являются выражениями от значений искомой функции в предыдущей итерации, а в уравнениях (15)  $y_i$  является значением следующей итерации.

В каждой итерации система алгебраических уравнений (15) решается методом прогонки. Согласно сущности данного метода, решение системы уравнений (15) ищется в следующем виде:

$$y_i = \alpha_{i+1} \cdot y_{i+1} + \beta_{i+1}. \quad (17)$$

Так как система (15) представляет с собой систему алгебраических уравнений с треугольной матрицей. Здесь  $\alpha_i, \beta_i$  являются неизвестными коэффициентами, называемые прогоночными.

Метод прогонки состоит из двух этапов: прямая прогонка используется для вычисления прогоночных коэффициентов  $\alpha_i$  и  $\beta_i$ , а в обратной прогонке вычисляются искомые значения функции  $y_i$ .

Для определения первых прогоночных коэффициентов  $\alpha_1$  и  $\beta_1$  используется граничное условие (3); Сравнение следующих формул:

$$y_0 = y_1;$$

$$y_0 = \alpha_1 y_1 + \beta_1$$

позволяет определить

$$\alpha_1 = 1; \quad \beta_1 = 0, \quad (18)$$

Теперь используется следующая преобразованная формула (17):

$$y_{i-1} = \alpha_i y_i + \beta_i. \quad (17^*)$$

Подставляя формулу (17\*) в формулу (15), можно получить следующую формулу:

$$y_i = \frac{A_{i+1}}{C_i - A_i \alpha_i} y_{i+1} + \frac{F_i + A_i \beta_i}{C_i - A_i \alpha_i}. \quad (19)$$

Сравнивая формулы (19) и (17), можно получить следующие рекуррентные формулы:

$$\alpha_{i+1} = \frac{A_{i+1}}{C_i - A_i \alpha_i}; \quad \beta_{i+1} = \frac{F_i + A_i \beta_i}{C_i - A_i \alpha_i}. \quad (20)$$

*Прямая прогонка:* используя формулы (18) и (20), можно найти значения прогоночных коэффициентов.

Из граничного условия (4) следует, что

$$y_n = 0 \quad (21)$$

*Обратная прогонка:* формулы (21) и (17) позволяют определить искомые значения функций  $y_i, i = n, n-1, \dots, 1$ .

Совокупность формул (18), (20), (21) и (17) образуют численную модель



поставленной задачи. По этим формулам могут быть выполнены вычисления для одной итерации, когда заданы значения искомой функции в предыдущей итерации.

*Метод итераций.* Теперь необходимо определить формулы для выполнения итерационного процесса.

Пусть вводятся следующие обозначения:

$Z_i$  – значение искомой функции в предыдущей итерации;

$y_i$  – значение этой же функции в вычисляемой итерации.

Тогда коэффициент (15) будут записаны в следующем виде:

$$A_i = a \cdot \left(\frac{Z_{i-1} + Z_i}{2}\right)^3; A_{i+1} = a \cdot \left(\frac{Z_i + Z_{i+1}}{2}\right)^3 \quad (22)$$

$$F_i = -Z_i - \{V_{j+1}(1 - x_i^2), \text{ если } 0 \leq i \leq k; 0, \text{ если } k + 1 < i \leq n;$$

где  $a = \frac{ER}{3} \cdot \frac{\square}{h^2}$  постоянная величина.

Алгоритм итерационного процесса будет иметь следующие этапы:

1<sup>0</sup>. Ввод начальных значений искомой функции  $y_i$  в момент времени  $t$ , то есть при  $j$  (нулевая итерация); присваивание  $q_i = y_i$  (значения в предыдущий момент времени, при  $j$ ):

2<sup>0</sup>. Цикл по параметру от нуля до  $n$  для присваивания  $Z_i = y_i$ .

3<sup>0</sup>. Цикл по параметру для вычисления коэффициентов системы уравнений (15) по формулам (22).

4<sup>0</sup>. Прямая прогонка:

4.1. Вычисление по формулам (18) первых прогоночных коэффициентов;

4.2. Цикл по параметру  $i$  для вычисления прогоночных коэффициентов по формулам (20).

5<sup>0</sup>. Обратная прогонка:

5.1. Вычисление  $y_n$  по формуле (21);

5.2. Цикл по параметру  $i$  для вычисления значений  $y_i$ ,  $i = n-1, n-2, \dots, 1$  по формуле (17).

6<sup>0</sup>. Вычисление абсолютной разности значений искомой функции для двух итераций с помощью цикла по  $i$ :

$$\gamma_i = |y_i - z_i|, \quad i = 0, 1, 2, \dots, n.$$

7<sup>0</sup>. Определения наибольшего значения массива  $\gamma$ ,  $i = 0, 1, 2, \dots, n$ .

8<sup>0</sup>. Проверка условия точности вычислений:

$$y_i < \varepsilon$$

8.1. если данное условие не выполняется, то осуществляется переход в пункт 2<sup>0</sup> для продолжения итерационного процесса;

8.2. если данное условие выполняется, то считается, что требуемая точность достигнута и завершается итерационный процесс.

В результате выполнения данного итерационного процесса будут определены значения искомой функции  $y$ , удовлетворяющие требуемой точности, для момента  $t$ , то есть для  $j + 1$ .

Точность вычисления задается в виде следующего условия:

$$\max \{ |y_i - z_i| \} < \varepsilon, \tag{23}$$

где  $\varepsilon$  – достаточное малое положительное число; оно должно быть задано.

Учет «подвижности» границы области, занимаемой вытекающей жидкостью. С течением времени, по мере истечения жидкости из «канала», правая граница  $x = \rho(t)$  увеличивается. Выше была определена рекуррентная формула (14) для определения величина  $\rho(t)$ . С учетом обозначений (16), формула (14) может быть записана в следующем виде:

$$P_{j+1} = P_j + \frac{z(f_{j+1} - f_j)}{y_i + q_i}, \tag{24}$$

где  $q_i$  - значение искомой функции в момент  $t$ , то есть при  $j$ :  $q_i = u_{ij}$ , а  $y_i = u_{ij+1}$ .

Тогда общий порядок вычислений будет следующим:

- 1<sup>0</sup>. Ввод начальных значений:  $P_0 = 1$ ;  $y_i = 0$ ,  $i = 0, 1, 2, \dots, n$ ;  $n = \frac{1}{h}$ ;
- 2<sup>0</sup>.  $j=0$ ; начало внешнего цикла; присваивание  $q_i = y_i$ .
- 3<sup>0</sup>. Алгоритм итерационного процесса, приведенный выше.
- 4<sup>0</sup>. Организация выводов промежуточных значений искомой функции  $y_i$  и значения  $P_j$  для определенного момента времени  $t$  или  $j$ .
- 5<sup>0</sup>.  $j = j + 1$ ; переход к следующему моменту времени; вычисление  $P_{j+1}$  по формуле (24).
- 6<sup>0</sup>. Если  $j > m$ , то осуществляется переход к пункту 3<sup>0</sup> для вычисления значений искомой функции  $y_i$  для следующих моментов времени  $t$  или  $j$ .

Если  $j > m$ , то процесс завершается и осуществляется выход из программы.

*Численный эксперимент.* На основе данного алгоритма решения задачи была разработана компьютерная программа, с помощью которой проведен численный эксперимент.

*Целью* данного численного эксперимента является исследование поведения вытекающей из щели сильновязких жидкостей, имеющих различные механические свойства, в частности, различные коэффициенты динамической вязкости.

*Результаты расчетов.* Численный эксперимент был проведен для различных значений безразмерного параметра ER; в данном случае рассматривались следующие его значения: ER=0.1, ER=0.5, ER=1 и ER=10. Полученные численные результаты были представлены в виде таблиц и графиков.

Таблица 1- Наибольшие значения функции  $z = \xi(x, t)$

t	ER=1	ER=10	ER=0.5	ER=0.1
0.50	0.4958	0.4704	0.4977	0.4994
1.00	0.9178	0.7771	0.9497	0.9868
1.05	0.9520	0.7966	0.9889	1.0336
1.10	0.9847	0.8120	1.0269	1.0797
1.25	1.2078	0.8494	1.3330	1.2144
1.50	1.3042	0.8969	1.4421	1.4238

1.80	1.3929	0.9429	1.5563	1.6471
1.85	1.4137	0.9495	1.5725	1.6813
1.90	1.4316	0.9560	1.5877	1.7146
2.00	1.4581	0.9694	1.6156	1.7789
2.20	1.5061	0.9937	1.6812	2.1379
2.25	1.5162	0.9992	1.6953	2.1595
2.50	1.5665	1.0254	1.7658	2.2596
3.00	1.6467	1.0725	1.8651	2.4183

Сравнительное представление графиков функции, определяющей границу вытекшей вязкой жидкости, для различных моментов времени показаны на рисунках: рисунок 2 – положения границ при  $t = 1$ , рисунок 3 - положения границ при  $t = 1,5$ , рисунок 4 – положения границ при  $t = 2$ , рисунок 5 – положения границ при  $t = 3$ .

На рисунках показана правая часть границы, т.е. для  $x \geq 0$ , а левая часть границы (для  $x \leq 0$ ) является зеркальным отображением правой части.

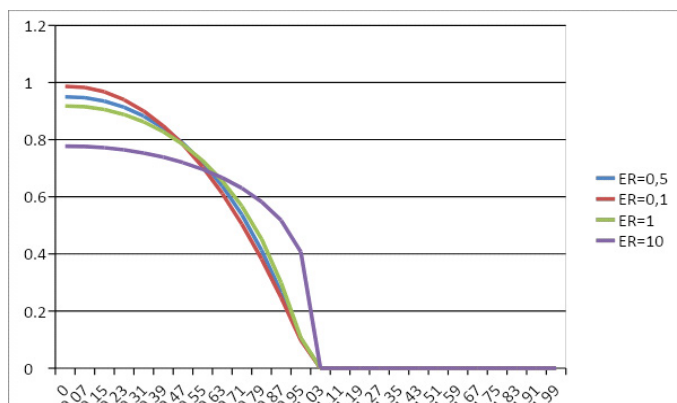


Рисунок 2 – Положения границ для момента  $t = 1$

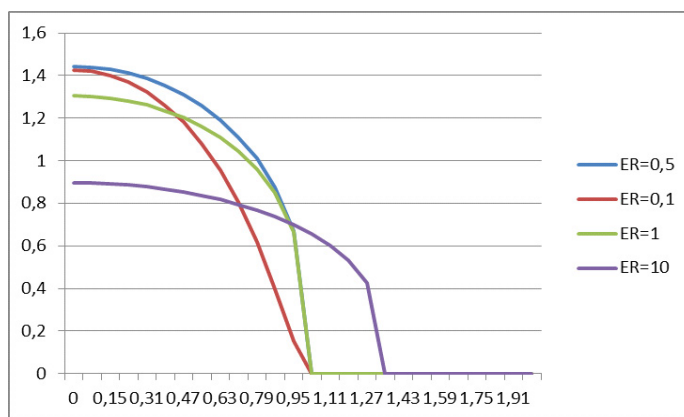


Рисунок 3 – Положения границ для момента  $t = 1,5$

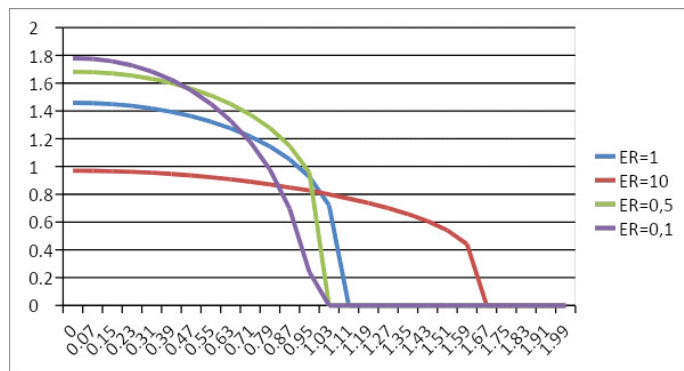


Рисунок 4 – Положения границ для момента  $t = 2$

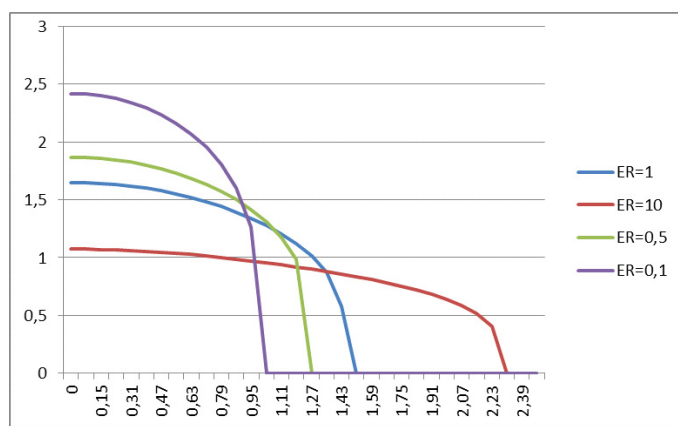


Рисунок 5 – Положения границ для момента  $t = 3$

На рисунках 6-9 показана динамика изменения границы вытекшей жидкости с течением времени: на рисунке 6 – для случая, когда число ER равно 10, на рисунке 7 – для ER=1, на рисунке 8 – для ER=0,5 и на рисунке 9 – для ER=0,1.

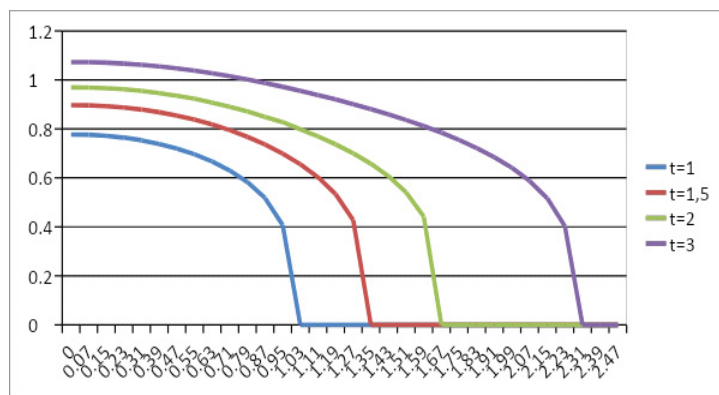


Рисунок 6 – Положения границы для случая ER=10

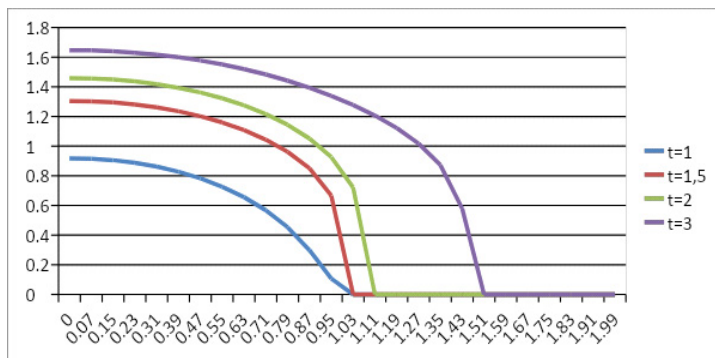


Рисунок 7 - Положения границы для случая ER=1

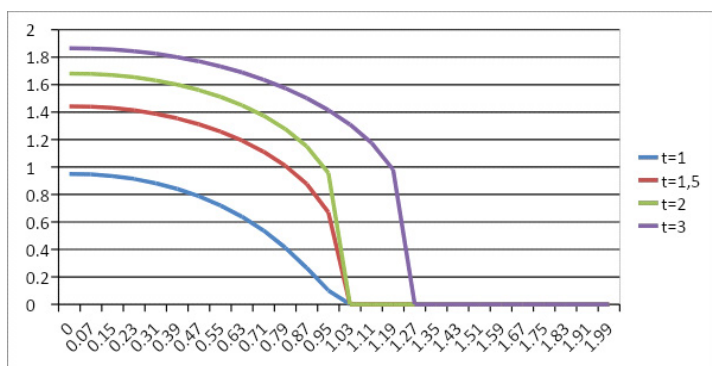


Рисунок 8 - Положения границы для случая ER=0,5

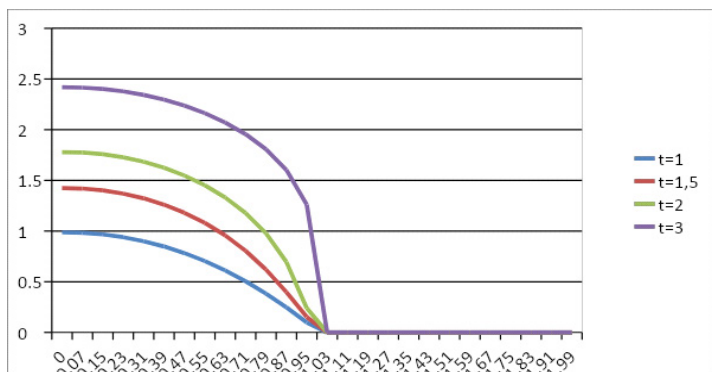


Рисунок 9 - Положения границы для случая ER=0,1

Доказательством того, что жидкость с меньшей вязкостью растекается в горизонтальном направлении быстрее, чем жидкость с большей вязкостью, являются следующие данные (Таблица 2), которые показывают положение границы вытекшей жидкости в горизонтальном направлении для различных значений числа ER. В данной таблице 2, первоначальная граница считается, что она совпадает с границей щели, т.е. равна 1.

Таблица 2 – Положение пересечения границы жидкости с горизонтальной осью

t	ER=1	ER=10	ER=0.5	ER=0.1
0.50	1.00	1.00	1.00	1.00
1.00	1.00	1.00	1.00	1.00
1.05	1.00	1.03	1.00	1.00
1.10	1.00	1.05	1.00	1.00
1.25	1.00	1.15	1.00	1.00
1.50	1.00	1.33	1.00	1.00
1.80	1.01	1.51	1.00	1.00
1.85	1.05	1.55	1.00	1.00
1.90	1.06	1.59	1.00	1.00
2.00	1.07	1.65	1.00	1.00
2.20	1.15	1.79	1.03	1.00
2.25	1.17	1.81	1.03	1.00
2.50	1.27	1.97	1.11	1.00
3.00	1.45	2.27	1.27	1.00

Здесь приведены результаты расчетов для промежутка времени  $0 \leq t \leq 3$ . В случае, когда  $ER=0.1$ , т.е. когда динамический коэффициент вязкости жидкости является большим числом, горизонтальная граница вытекшей жидкости остается неизменной. Для  $ER=0.5$  изменение этой границы начинается после  $t \geq 2,20$ , для  $ER=1$  -  $t \geq 1,80$  и для  $ER=10$  после  $t \geq 1,05$ .

Положение точки пересечения поверхности вытекшей жидкости с горизонтальной осью для различных моментов времени для различных значений числа ER показано в следующей таблице 3.

Таблица 3 - Положения точки пересечения границы вытекшей жидкости с горизонтальной осью

t	1,0	1,5	2,0	3,0
ER=10	1,03	1,35	1,67	2,31
ER=1	1,03	1,03	1,11	1,51
ER=0,5	1,01	1,01	1,01	1,27
ER=0,1	1,00	1,00	1/00	1,03

Процессы перехода горных пород перед землетрясениями из ненарушенного состояния в состояние динамического разрушения, в условиях ограниченности объема очаговой зоны и сжатия на больших глубинах, происходят замедленно. Такое замедление может быть использовано в прогностических целях (Ким и др., 2022: 69-84).

### **Выводы**

Анализ результатов, полученных в численном эксперименте, позволяют сделать следующее заключение.

1. Сравнение результатов расчетов для различных значений числа ER (Таблица 1) показало, что максимальное поднятие поверхности вытекшей жидкости  $z = \xi(x, t)$  при малых значениях числа ER больше чем при больших

его значениях. Следует напомнить, что число  $ER$  обратно пропорционально коэффициенту динамической вязкости вытекающей жидкости. Это означает, что вытекающая жидкость, динамический коэффициент которой является сравнительно большим, поднимается выше и меньше растекается в горизонтальном направлении, чем жидкость, у которой динамический коэффициент вязкости меньше.

2. Данные, приведенные в таблицах 2 и 3 показывают, что сильновязкая жидкость с меньшим коэффициентом динамической вязкости растекается в горизонтальном направлении более шире нежели жидкость с большим коэффициентом вязкости. Вязкая жидкость с большим коэффициентом вязкости (при малых числах  $ER$ ) в начальные моменты времени ведет себя как твердое тело.

3. Результаты расчетов компьютерные моделирование показали, что предложенная здесь модель движения сильновязкой жидкости может быть использована для описания процесса истечения магматических веществ из нижележащих недр Земли по узкому каналу, образованным в тектоносфере и приводит к возбуждению температурной аномалии в мантии. Такая аномалия развивается и вызывает подъем вещества из глубоких слоев наружу и погружение изначально сформировавшихся слоев земной коры вглубь мантии.

Результате моделирование процесса извержения магмы представляет ценную информацию для наблюдающих за изменением структуры Земной коры, быстрого обнаружения сильных землетрясений и предупреждений пользователям как землетрясение достигнет до их местоположения.

#### ЛИТЕРАТУРА

Белоцерковский О.М., Опарин А.М., Чететкин В.М. (2008). Численное моделирование гидродинамических неустойчивостей и турбулентности. В книге «Будущее прикладной математики. лекции для молодых исследователей. От идей до технологии». Под ред. Г.Г. Малинецкого. – М.: КомКнига, 512 с.

Добрецов Н.Л., Кирдяшкин А.Г., Кирдяшкин А.А. (2001). Глубинная геодинамика. – Новосибирск, 409 с.

Ким А.С., Шпади Ю.Р., Литвинов Ю.Г. (2022). Математическое моделирование нестационарных процессов в сейсмоактивной зоне. Динамические процессы в геосферах, т. 14. № 1, 69–84

Кузьмин М.И., Ярмолюк Б.Б. (2016). Тектоника плит и мантийные плюмы – основа эндогенной тектонической активности Земли последние 2 млрд лет. Новосибирск: Изд-во СО РАН, филиал «Гео». Геология и геофизика. Т.57, №1. 11–30.

Куралбаев З.К. (2008). Модельное исследование тектонических движений в системе «литосфера-астеносфера». Алматы, 212 с.

Лобковский Л.И. (2016). Тектоника деформируемых литосферных плит и модель региональной геодинамики применительно к Арктике и Северо-Восточной Азии. Новосибирск: Изд-во СО РАН, филиал «Гео». Геология и геофизика. Т.57, №3, 11–30.

Тихонов А.Н., Самарский А.А. (1977). Уравнения математической физики. М.: Наука, 736 с.

Kyung H.M., Chongyoup K. (2010). Simulation of Particle Migration in Free-Surface Flows. AIChE Journal. N.10. Vol. 56, 2539–2550.

**REFERENCES**

- Belotserkovsky O.M., Oparin A.M., Chechetkin V.M. (2008). Numerical simulation of instabilities and turbulence gidrodinmicheskikh // In the book "The Future of Applied Mathematics. lectures for young researchers. From ideas to technology ". Ed. G.G.Malinetskogo. -M.: KomKniga, p. 512 (in Rus.)
- Dobretsov N.L., Kirdyashkin A.G., Kirdyashkin A.A. (2001). Deep geodynamics. Novosibirsk, 409 p. (in Rus.)
- Kim A.S., Shpadi Yu. R., Litvinov Yu. G. (2022). Mathematical modeling of non-stationary processes in a seismically active zone. Dynamic Processes in Geospheres. Vol. 14, № 1, 69–84. (in Rus.)
- Kuralbayev Z.K. (2008). Model study tectonic movements in the "lithosphere-asthenosphere." Almaty, p. 212. (in Rus.)
- Kuzmin M.I., Yarmolyuk B.B. (2016). Tectonics of plates and mantle plumes - the basis of endogenous tectonic activity of the Earth for the last 2 billion years. Novosibirsk: Izdatel'stvo SB RAS, Geo branch. Geology and geophysics. Vol.57. No.1. 11–30. (in Rus.)
- Kyung H.M., Chongyoun K. (2010), Simulation of Particle Migration in Free-Surface Flows. AIChE Journal. N.10. Vol.56 2539–2550. (in Eng.)
- Lobkovsky L.I. (2016). Tectonics of deformable lithospheric plates and a model of regional geodynamics applied to the Arctic and North-East Asia. Novosibirsk: Izdatel'stvo SB RAS, branch "Geo". Geology and geophysics. Vol.57. No.3, 11–30. (in Rus.)
- Tikhonov A.A. Samarskii A.A. (1977). The equations of mathematical physics. M.: Nauka, 736 p. (in Rus.).



NEWS OF THE NATIONAL ACADEMY OF SCIENCES OF THE REPUBLIC OF  
KAZAKHSTAN  
PHYSICO-MATHEMATICAL SERIES  
ISSN 1991-346X  
Volume 3. Number 347 (2023). 193–209  
<https://doi.org/10.32014/2023.2518-1726.213>

UDC 681.518.3

© **G.S. Shaimerdenova\***, **R.A. Sarkulakova**, **M.M. Turganbekova**,  
**B.O. Tastanbekova**, **M.T. Baizhanova**, 2023

M. Auezov South Kazakhstan University, Shymkent, Kazakhstan.  
E-mail: [danel01kz@gmail.com](mailto:danel01kz@gmail.com)

### **ADVANCEMENTS IN MOBILE AND ONLINE BANKING: A COMPREHENSIVE ANALYSIS OF TECHNOLOGIES AND INNOVATIONS**

**Shaimerdenova G.S.** — PhD. M.Auezov South Kazakhstan University. Department of Information and communication technologies. 160000. Shymkent, Kazakhstan

E-mail: [danel01kz@gmail.com](mailto:danel01kz@gmail.com). ORCID: 0000-0001-8685-7125;

**Sarkulakova R.A.** — Senior Lecturer. M.Auezov South Kazakhstan University. Department of Information and communication technologies. 160000. Shymkent, Kazakhstan

E-mail: [rsarkulakova@inbox.ru](mailto:rsarkulakova@inbox.ru). ORCID: 0000-0002-2853-2207;

**Turганbekova M.M.** — Senior Lecturer. M.Auezov South Kazakhstan University. Department of Information and communication technologies. 160000. Shymkent, Kazakhstan

E-mail: [madina\\_turganbek@mail.ru](mailto:madina_turganbek@mail.ru). ORCID: 0009-0004-8582-8276;

**Tastanbekova B.O.** — Senior Lecturer. M.Auezov South Kazakhstan University. Department of Information and communication technologies. 160000. Shymkent, Kazakhstan

E-mail: [baya2013@inbox.ru](mailto:baya2013@inbox.ru). ORCID: 0000-0002-4549-0344.

**Baizhanova M.T.** — Senior Lecturer. M.Auezov South Kazakhstan University. Department of Information and communication technologies. 160000. Shymkent, Kazakhstan

E-mail: [madya\\_bayzhanova@mail.ru](mailto:madya_bayzhanova@mail.ru). ORCID: 0000-0002-6022-1820.

**Abstract.** The article provides an in-depth review of the rapid progression of digital banking driven by technological advancements and breakthrough innovations. The manuscript explores the journey from the inception of online banking to the rise of mobile banking, and further to the advent of digital-only banks. It elucidates how these digital channels have revolutionized the financial landscape, enabling unprecedented convenience, speed, and accessibility in banking services. The paper also discusses the transformational impact of technologies such as mobile applications, biometric authentication, artificial intelligence (AI), machine learning (ML), and blockchain in reshaping the banking sector. The crucial role of AI and ML in enhancing customer service, fraud detection, credit risk assessment, and personalization of product recommendations is comprehensively examined. Moreover, the manuscript delves into the potential of blockchain technology in

ensuring secure, transparent, and efficient transactions. Furthermore, it underscores the role of regulatory frameworks and compliance in the digital banking ecosystem. It brings to light the challenges and risks associated with digital banking, including cybersecurity threats and the digital divide, while suggesting mitigation strategies. Looking ahead, the manuscript identifies future trends and opportunities such as the integration of digital banking with emerging financial technologies (e.g., digital currencies, robo-advisors), the role of 5G and Internet of Things (IoT) in digital banking, and the importance of creating personalized, customer-centric banking experiences. The paper concludes with implications for the banking industry and policymakers, and directions for future research, making it an indispensable resource for anyone interested in understanding the evolution and future of digital banking.

**Keywords:** digital banking, financial innovation, cybersecurity, regulatory framework, emerging technologies

© Г.С. Шаймерденова\*, Р.А. Саркулакова, М.М. Тұрғанбекова,  
Б.Ө. Тастанбекова, М.Т. Байжанова, 2023

М. Әуезов атындағы Оңтүстік Қазақстан университеті, Шымкент, Қазақстан.  
E-mail: danel01kz@gmail.com

## **МОБИЛЬДІ ЖӘНЕ ОНЛАЙН-БАНКИНГТЕГІ ЖЕТІСТІКТЕР: ТЕХНОЛОГИЯЛАР МЕН ИННОВАЦИЯЛАРДЫ КЕШЕНДІ ТАЛДАУ**

**Шаймерденова Г.С.** — PhD. М.Әуезов атындағы Оңтүстік Қазақстан университеті. Ақпараттық-коммуникациялық технологиялар кафедрасы. 160000. Шымкент, Қазақстан  
E-mail: danel01kz@gmail.com. ORCID: 0000-0001-8685-7125;

**Саркулакова Р.А.** — Аға оқытушы. М.Әуезов атындағы Оңтүстік Қазақстан университеті. Ақпараттық-коммуникациялық технологиялар кафедрасы. 160000. Шымкент, Қазақстан  
E-mail: rsarkulakova@inbox.ru. ORCID: 0000-0002-2853-2207;

**Тұрғанбекова М.М.** — Аға оқытушы. М.Әуезов атындағы Оңтүстік Қазақстан университеті. Ақпараттық-коммуникациялық технологиялар кафедрасы. 160000. Шымкент, Қазақстан  
E-mail: madina\_turganbek@mail.ru. ORCID: 0009-0004-8582-8276;

**Тастанбекова Б.Ө.** — Аға оқытушы. М.Әуезов атындағы Оңтүстік Қазақстан университеті. Ақпараттық-коммуникациялық технологиялар кафедрасы. 160000. Шымкент, Қазақстан  
E-mail: baya2013@inbox.ru. ORCID: 0000-0002-4549-0344.

**Байжанова М.Т.** — Аға оқытушы. М.Әуезов атындағы Оңтүстік Қазақстан университеті. Ақпараттық-коммуникациялық технологиялар кафедрасы. 160000. Шымкент, Қазақстан  
E-mail: madyna\_bayzhanova@mail.ru. ORCID: 0000-0002-6022-1820.

**Аннотация.** Мақалада технологиялық жетістіктер мен серпінді инновацияларға негізделген цифрлық банкінгтің қарқынды дамуына егжей-тегжейлі шолу берілген. Қолжазба онлайн-банкінгтің пайда болуынан мобильді банкінгтің көтерілуіне дейінгі және тек цифрлық форматта жұмыс істейтін банктер пайда болғанға дейінгі жолды зерттейді. Онда бұл цифрлық арналар бұрын-соңды болмаған банктік ыңғайлылықты, жылдамдықты және қолжетімділікті қамтамасыз ету арқылы қаржылық ландшафтты қалай

өзгерткені көрсетілген. Мақалада сонымен қатар мобильді қосымшалар, биометриялық аутентификация, жасанды интеллект (AI), машиналық оқыту (МО) және блокчейн сияқты технологиялардың банк секторының өзгеруіне трансформациялық әсерін талқылайды. Жасанды интеллект пен машиналық оқытудың тұтынушыларға қызмет көрсетуді жақсартудағы, алаяқтықты анықтаудағы, несиелік тәуекелді бағалаудағы және өнім ұсыныстарын жекелендірудегі шешуші рөлі жан-жақты зерттеледі. Сонымен қатар, қолжазба қауіпсіз, мөлдір және тиімді транзакцияларды қамтамасыз етудегі блокчейн технологиясының әлеуетін қарастырады. Сонымен қатар, ол цифрлық банкинг экокүйесіндегі нормативтік-құқықтық базаның және сәйкестіктің рөлін көрсетеді. Ол цифрлық банкингке қатысты мәселелер мен тәуекелдерді, соның ішінде киберқауіпсіздік қатерлерін және цифрлық алшақтықты анықтайды және азайту стратегияларын ұсынады. Қолжазбада болашақ тенденциялар мен мүмкіндіктер, мысалы, цифрлық банкингті жаңа қаржылық технологиялармен біріктіру (мысалы, сандық валюталар, робо-кеңесшілер), цифрлық банкингтегі 5G және Заттар интернетінің (IoT) рөлі және жеке, клиентке бағытталған банктік қызметті құрудың маңыздылығы анықталады. Мақалада банк саласы мен саясаткерлер үшін салдармен, сондай-ақ болашақ зерттеу бағыттарымен аяқталады, бұл оны цифрлық банкингтің эволюциясы мен болашағына қызығушылық танытқандар үшін таптырмас ресурс етеді.

**Түйін сөздер:** цифрлық банкинг, қаржылық инновациялар, киберқауіпсіздік, нормативтік база, жаңа технологиялар

© Г.С. Шаймерденова\*, Р.А. Саркулакова, М.М. Турганбекова,  
Б.О. Тастанбекова, М.Т. Байжанова, 2023

Южно-Казахстанский университет имени М.Ауэзова, Астана, Казахстан.

E-mail: danel01kz@gmail.com

## ДОСТИЖЕНИЯ В МОБИЛЬНОМ И ОНЛАЙН-БАНКИНГЕ: КОМПЛЕКСНЫЙ АНАЛИЗ ТЕХНОЛОГИЙ И ИННОВАЦИЙ

**Шаймерденова Г.С.** — PhD. Южно-Казахстанский университет имени М.Ауэзова. Кафедра информационно-коммуникационных технологий. 160000. Шымкент, Казахстан.

E-mail: danel01kz@gmail.com. ORCID: 0000-0001-8685-7125;

**Саркулакова Р.А.** — Старший преподаватель. Южно-Казахстанский университет имени М.Ауэзова. Кафедра информационно-коммуникационных технологий. 160000. Шымкент, Казахстан.

E-mail: rsarkulakova@inbox.ru. ORCID: 0000-0002-2853-2207;

**Турганбекова М.М.** — Старший преподаватель. Южно-Казахстанский университет имени М.Ауэзова. Кафедра информационно-коммуникационных технологий. 160000. Шымкент, Казахстан.

E-mail: madina\_turganbek@mail.ru. ORCID: 0009-0004-8582-8276;

**Тастанбекова Б.О.** — Старший преподаватель. Южно-Казахстанский университет имени М.Ауэзова. Кафедра информационно-коммуникационных технологий. 160000. Шымкент, Казахстан.

E-mail: baya2013@inbox.ru. ORCID: 0000-0002-4549-0344.

**Байжанова М.Т.** — Старший преподаватель. Южно-Казахстанский университет имени М.Ауэзова. Кафедра информационно-коммуникационных технологий. 160000. Шымкент, Казахстан.  
E-mail: madyna\_bayzhanova@mail.ru. ORCID: 0000-0002-6022-1820.

**Аннотация.** В статье представлен подробный обзор быстрого развития цифрового банкинга, обусловленного технологическими достижениями и прорывными инновациями. В рукописи исследуется путь от зарождения онлайн-банкинга до подъёма мобильного банкинга и далее до появления банков, работающих только в цифровом формате. В нем показано, как эти цифровые каналы произвели революцию в финансовом ландшафте, обеспечив беспрецедентное удобство, скорость и доступность банковских услуг. В документе также обсуждается трансформационное влияние таких технологий, как мобильные приложения, биометрическая аутентификация, искусственный интеллект (ИИ), машинное обучение (МО) и блокчейн на изменение банковского сектора. Всесторонне исследуется решающая роль искусственного интеллекта и машинного обучения в улучшении обслуживания клиентов, обнаружении мошенничества, оценке кредитного риска и персонализации рекомендаций по продуктам. Кроме того, в рукописи рассматривается потенциал технологии блокчейн в обеспечении безопасных, прозрачных и эффективных транзакций. Кроме того, в нем подчеркивается роль нормативно-правовой базы и соответствия требованиям в экосистеме цифрового банкинга. В нем выявляются проблемы и риски, связанные с цифровым банкингом, включая угрозы кибербезопасности и цифровой разрыв, а также предлагаются стратегии смягчения последствий. Забегая вперед раскрою, что в рукописи определяются будущие тенденции и возможности, такие как интеграция цифрового банкинга с новыми финансовыми технологиями (например, цифровые валюты, роботы-консультанты), роль 5G и Интернета вещей (IoT) в цифровом банкинге, а также важность создания персонализированного, клиентоориентированного банковского обслуживания. Статья завершается последствиями для банковской отрасли и политиков, а также направлениями будущих исследований, что делает его незаменимым ресурсом для всех, кто интересуется эволюцией и будущим цифрового банкинга.

**Ключевые слова:** цифровой банкинг, финансовые инновации, кибербезопасность, нормативная база, новые технологии

### **Introduction**

Mobile and online banking have revolutionized financial management, offering unparalleled convenience, speed, and accessibility. These digital banking channels evolved with advancements in information and communication technologies, playing a critical role in shaping the modern financial landscape. Online banking emerged in the early 1980s when banks began offering limited electronic services via telephone lines and dedicated terminals (Sanli, 2015). With the internet's

advent in the 1990s, banks started providing online banking services through websites, allowing customers to access accounts, transfer funds, pay bills, and view transaction history from personal computers.

As mobile phones advanced and became widespread, banks saw potential in offering banking services through mobile devices. Mobile banking debuted in the early 2000s with basic SMS-based services for account balance checks and alerts. The late 2000s saw the launch of smartphones and mobile applications, as banks developed dedicated mobile apps offering comprehensive features like account management, remote check deposit, and peer-to-peer payments (Prisc, 2021). The banking landscape further evolved with the rise of digital-only banks, or neobanks, which operate exclusively online without physical branches. These banks focus on user experience, competitive fees, and innovative features, making them an attractive alternative to traditional banks.

Mobile and online banking have profoundly impacted the banking sector, driving competition, improving customer experiences, and reducing operational costs. Customers now expect seamless, secure, and efficient banking experiences accessible anytime, anywhere. Consequently, banks invest heavily in digital transformation initiatives to meet demands and remain competitive in an increasingly digital financial landscape. Mobile and online banking have come a long way since their early beginnings, becoming integral components of the modern banking experience. With rapid advancements in technology and changing customer preferences, digital banking channels will continue to evolve and reshape the financial industry.

Digital banking has become a cornerstone of the modern financial landscape due to its transformative impact on the banking sector and the numerous advantages it offers (Arrighi, 2014). It provides convenience and accessibility, enabling customers to access financial services 24/7 from anywhere with an internet connection, saving time and resources. Leveraging digital channels allows banks to reduce operational costs and pass on savings to customers in the form of lower fees and better interest rates. Improved customer experience, real-time services, and enhanced security are other benefits of digital banking. User-friendly interfaces and personalized services lead to increased customer satisfaction and loyalty. Real-time transaction processing and instant access to account information keep customers informed about their financial status. Advanced security measures protect sensitive financial data. Digital banking promotes financial inclusion by reaching unbanked and underbanked populations, fostering economic growth and reducing income inequality in underserved communities. The rise of digital banking spurs innovation and competition in the financial industry, ultimately benefiting customers through new features and improved services (de Lis, 2018). Additionally, digital banking contributes to environmental sustainability by reducing the need for physical branches and paper-based transactions.

Data-driven decision-making and integration with emerging technologies are other key aspects of digital banking. Analyzing customer data helps banks tailor

their products and services, improve risk management, and make better-informed decisions. Digital banking platforms can easily integrate with technologies such as artificial intelligence, machine learning, blockchain, and the Internet of Things (IoT) to enhance offerings and streamline operations.

Digital banking plays a vital role in the modern financial landscape by driving innovation, improving customer experiences, and promoting financial inclusion. As technology continues to advance and customer preferences shift, the importance of digital banking will only grow, shaping the future of the banking industry (Liu, 2020).

The scope of this review is to examine the historical development, current state, and future trends of mobile and online banking, with a focus on the technologies and innovations that have shaped this transformation. The review will provide a comprehensive analysis of the digital banking landscape, highlighting its impact on customers, banks, and the broader financial industry.

By achieving this aim, the review will offer a comprehensive understanding of the evolution of mobile and online banking, providing valuable insights for academics, industry professionals, and policymakers interested in the ongoing digital transformation of the banking sector.

### **Methodology**

To provide a comprehensive and well-structured analysis of the evolution of mobile and online banking, the following methodology was employed:

#### 1. Literature Search Strategy:

A systematic search of relevant literature was conducted using various electronic databases, such as Google Scholar, IEEE Xplore, ScienceDirect, and JSTOR. The search focused on academic articles, conference proceedings, whitepapers, and industry reports published in English. To identify relevant publications, a combination of keywords and phrases was used, including "mobile banking," "online banking," "digital banking," "technological advancements," "innovations," "challenges," and "future trends."

#### 2. Inclusion and Exclusion Criteria:

To ensure the quality and relevance of the selected literature, specific inclusion and exclusion criteria were applied. Inclusion criteria consisted of factors such as:

Publication date: Focusing on sources published within the last 10-15 years to capture recent developments and trends.

Relevance: Articles must have had a primary focus on mobile and online banking technologies and innovations.

Methodological rigor: Preference was given to studies that demonstrated methodological rigor and provided clear, well-supported conclusions.

Exclusion criteria included: Non-English publications; opinion pieces or editorials without substantial evidence or analysis; publications with a narrow focus that did not contribute to the broader understanding of the topic.

#### 3. Data Extraction and Analysis:

Once the relevant literature was identified, the following data were extracted:

Bibliographic information (author, title, publication date, source); study design and methodology; key findings, technologies, innovations, challenges, and trends.

The extracted data were synthesized and analyzed to address the review's objectives. This analysis involved: a chronological overview of the development of mobile and online banking, highlighting key milestones and technological advancements; a thematic analysis of the identified innovations, challenges, and trends in mobile and online banking; a discussion of the regulatory frameworks, compliance requirements, and their impact on digital banking; an exploration of future trends and opportunities, including the potential implications of emerging technologies.

By employing this methodology, the review manuscript offered a comprehensive and systematic analysis of the evolution of mobile and online banking, focusing on the technologies and innovations that drove this transformation. This approach ensured that the review was grounded in the existing literature, providing valuable insights for academics, industry professionals, and policymakers interested in the digital transformation of the banking sector.

### **Discussion**

The emergence of online banking systems can be traced back to the early 1980s, when financial institutions began experimenting with electronic banking services to meet the evolving needs of their customers. Early online banking systems were quite different from the sophisticated platforms we know today, as they were limited in functionality and accessibility.

Before the widespread adoption of the internet, banks offered limited electronic services using telephone lines and dedicated terminals. Some of the earliest electronic banking services were introduced by major banks in the United States, such as Bank of America and Citibank. These services typically allowed customers to access basic account information and perform simple transactions like transferring funds between accounts (Gonzalez, 2015).

During this period, home banking systems were also developed, allowing customers to access their accounts through a personal computer equipped with a modem. These early systems often used proprietary software provided by the bank and required customers to dial directly into the bank's computer systems (Shaikh, 2015). Despite offering limited features, these early home banking systems were considered innovative and paved the way for the development of more advanced online banking services.

With the advent of the internet in the 1990s, banks saw the potential for expanding their electronic banking services and started offering online banking through websites. These enabled customers to access their accounts, transfer funds, pay bills, and view transaction history from their personal computers. These early online banking websites were relatively basic in design and functionality compared to today's standards. They relied on static HTML pages, lacked personalization features, and often required the use of specific browsers for compatibility.

Security was also a significant concern in the early days of online banking.

Banks implemented various security measures, such as usernames, passwords, and security questions, to authenticate users and protect sensitive financial data. Over time, encryption technologies like Secure Socket Layer (SSL) were adopted to enhance the security of online transactions and communication between users and banking websites.

Despite the limitations of early online banking systems, they represented a significant breakthrough in the banking industry, offering customers increased convenience and accessibility to their financial information. As internet usage continued to grow and technology advanced, online banking systems evolved to offer more sophisticated features, better user experiences, and improved security measures.

Early online banking systems played a crucial role in shaping the digital banking landscape we know today. By providing customers with the ability to access and manage their finances electronically, these systems laid the foundation for the rapid development of mobile and online banking technologies and innovations. The emergence of mobile banking can be attributed to the rapid development of mobile technology and the widespread adoption of smartphones and mobile internet. As the use of smartphones became more prevalent, banks saw an opportunity to leverage these devices to offer financial services to their customers on the go, providing greater convenience and accessibility. This led to the birth of mobile banking.

In the early stages of mobile banking, banks primarily used text messages (SMS) to communicate with customers. Customers could receive account balance updates, transaction alerts, and even perform basic transactions by sending text messages to their banks. The convenience of SMS banking was a significant step forward, but the system was relatively simple and limited in functionality.

As mobile technology progressed, banks began developing mobile applications (apps) specifically designed for smartphones. These early mobile banking apps allowed customers to access their accounts, view transaction history, transfer funds, and pay bills directly from their smartphones. Mobile apps offered a more user-friendly and convenient experience compared to SMS banking, as they provided a richer interface and more advanced features.

The development of mobile banking apps was further accelerated by the emergence of app stores, such as Apple's App Store and Google's Play Store, which made it easier for customers to discover and download banking apps. This led to a surge in the adoption of mobile banking services, as customers increasingly embraced the convenience and flexibility provided by these apps.

In addition to traditional banks, the growth of mobile banking also spurred the creation of new financial institutions, such as digital-only banks or neobanks, that operate exclusively through mobile apps and online platforms (Norohna, 2023). These digital banks often offer innovative features, lower fees, and personalized services that cater to the needs of the tech-savvy, mobile-first generation.

The emergence of mobile banking also introduced new security concerns, as banks had to find ways to protect customer data and ensure the security of



transactions performed through mobile devices. Banks responded by implementing advanced security measures, such as biometric authentication, encryption, and secure mobile application development practices.

Over time, mobile banking has continued to evolve, with new technologies and innovations being integrated into the apps. Today, mobile banking offers a wide array of features, such as mobile check deposits, person-to-person (P2P) payments, budgeting tools, and even access to financial products like loans and investments.

Mobile banking has revolutionized the banking industry and customer interaction with financial institutions. Offering convenient, accessible, and user-friendly financial services via smartphones, it has become an integral component of the contemporary financial landscape.

Technological advancements have reshaped customer interaction with banks and financial management, enabling institutions to offer innovative services, enhance customer experience, and improve operational efficiency. Crucial developments include mobile apps that increase accessibility and enable customers to perform various banking tasks from their smartphones, contributing to mobile banking's growing popularity (Chou, 2004). Biometric authentication, including fingerprint, facial recognition, and voice recognition, provides a secure and convenient way for customers to access their accounts, enhancing digital banking security and streamlining user experience. Artificial intelligence (AI) and machine learning have improved customer service through AI-powered chatbots and virtual assistants, while machine learning algorithms enable better fraud detection, credit risk assessment, and personalized product recommendations (Manser Payne, 2018).

Open banking and application programming interfaces (APIs) have transformed the digital banking landscape, creating a more interconnected and collaborative financial ecosystem involving traditional banks, fintech companies, and other financial service providers (Fiedler, 2022). Blockchain technology offers potential solutions for secure, transparent, and efficient transactions, reducing transaction costs and streamlining cross-border payments. The growing popularity of digital wallets and contactless payments has also impacted digital banking, providing a fast and convenient alternative to traditional payment methods. These key technological advancements have revolutionized customer interaction with financial institutions and financial management, contributing to the ongoing digital transformation of the banking industry.

Mobile apps and responsive web design are two approaches used by businesses and organizations, including banks, to ensure that their digital services are accessible, user-friendly, and functional across a wide range of devices. These approaches aim to provide customers with seamless and engaging experiences, regardless of the type of device they use to access the services.

Mobile apps are standalone applications specifically designed for smartphones and other mobile devices. They are developed using native programming languages, such as Java for Android devices and Swift for iOS devices, and offer a tailored user experience optimized for the device's operating system. Mobile apps

typically provide a rich, interactive interface, making it easy for users to navigate through the app and perform various tasks. They can also take full advantage of the device's hardware features, such as the camera, GPS, and biometric sensors, to offer additional functionalities and a more personalized experience. Information about the most famous mobile banking apps and their features are shown in Table 1.

Table 1 – Most Famous Mobile Banking Apps and Their Features (Horton, 2023; Richter, 2018)

Mobile Banking App	Description
Chase Mobile	Chase Mobile is the mobile banking app developed by JPMorgan Chase, one of the largest banks in the United States. The app allows customers to view account balances, transaction history, deposit checks, pay bills, transfer funds, and locate ATMs and branches. It also offers features like personalized alerts, budgeting tools, and Zelle integration for quick P2P payments.
Bank of America	Bank of America's mobile banking app enables customers to access their accounts, transfer money, deposit checks, pay bills, and locate nearby ATMs and financial centers. The app also includes features like custom alerts, mobile credit card management, and integration with the bank's virtual financial assistant, Erica, which uses AI to provide personalized financial guidance.
Wells Fargo Mobile	Wells Fargo Mobile app allows customers to manage their accounts, pay bills, deposit checks, transfer funds, and find ATMs and branches. The app also offers additional features such as budgeting tools, custom alerts, and card management options, including the ability to temporarily freeze a lost or misplaced card.
Citi Mobile	Citi Mobile is the mobile banking app developed by Citibank, offering customers access to their account information, transaction history, bill payments, fund transfers, and mobile check deposits. The app also features Citi Mobile Snapshot, which allows users to view account balances and recent transactions without logging in, and integration with Citi's virtual assistant.
Capital One	Capital One's mobile banking app lets customers manage their accounts, pay bills, deposit checks, transfer funds, and locate nearby ATMs and branches. The app also integrates with Capital One's CreditWise feature, enabling users to monitor their credit score and receive personalized credit insights. The app offers additional security features like fingerprint login.

Kaspi.kz, a leading fintech company in Kazakhstan, offers a comprehensive mobile banking solution through its all-in-one mobile app (Ivashina, 2019). As a key player in the nation's digital banking landscape, Kaspi.kz provides innovative and user-friendly services, including account access, transaction history, bill payments, and seamless fund transfers. The app also supports instant peer-to-peer payments and various financial products like loans, insurance, and investments. Kaspi Gold, a loyalty program integrated into the app, rewards users with cashback for purchases made using Kaspi.kz cards. Security is a notable focus for Kaspi.kz, employing advanced measures like biometric authentication to protect user data and ensure transaction safety. Users can manage card settings, such as blocking and unblocking cards or setting transaction limits, to further enhance security.

The Kaspi.kz app also offers non-financial services, enabling users to book movie tickets, pay for utilities, and top up mobile phone balances within its ecosystem. This holistic approach to digital services makes Kaspi.kz an essential tool for users seeking to manage finances and access everyday services effortlessly.

Mobile apps can be installed directly from app stores, such as Google Play Store and Apple App Store, and often provide offline functionality, allowing users to access certain features even without an internet connection. This can be particularly useful for banking services, as customers may need to access their account information or perform transactions when they are not connected to the internet.

On the other hand, responsive web design is an approach to website development that ensures the website's layout, images, and other elements automatically adjust to fit the screen size and orientation of the device being used. This enables users to have a consistent experience across various devices, such as smartphones, tablets, and desktop computers. Responsive web design is achieved through the use of fluid grids, flexible images, and CSS media queries that detect the characteristics of the user's device and adapt the website accordingly.

Responsive web design is particularly important for online banking services, as customers may access their accounts using different devices throughout the day. By ensuring that the website's layout and content are displayed optimally on every device, banks can provide a more user-friendly and efficient experience for their customers.

Both mobile apps and responsive web design have their advantages and drawbacks. Mobile apps tend to offer a more engaging, feature-rich experience and can leverage the device's hardware capabilities. However, they require more development resources and maintenance, as separate versions of the app must be created and updated for each operating system. Responsive web design, in contrast, offers a more universal solution that can be accessed across various devices with minimal additional development effort. However, it may not provide the same level of interactivity and hardware integration as a dedicated mobile app.

Mobile apps and responsive web design are both essential components of a comprehensive digital strategy for banks and other organizations. By providing customers with accessible, user-friendly, and functional digital services across different devices, banks can enhance customer satisfaction, increase engagement, and ultimately strengthen their position in the competitive digital banking landscape.

Biometric authentication has emerged as an innovative and effective method for ensuring security in various industries, including digital banking. As traditional authentication methods such as passwords and PINs become increasingly vulnerable to security breaches and hacking attempts, biometric authentication offers a more reliable and user-friendly alternative.

Biometric authentication refers to the process of verifying an individual's identity based on unique physiological or behavioral characteristics. These characteristics include fingerprints, facial recognition, iris or retina scans, voice recognition, and even gait analysis (Michaels, 2022). By utilizing these unique identifiers, biometric

authentication can provide a higher level of security compared to traditional methods, as they are difficult to forge, hack, or replicate.

One of the primary advantages of biometric authentication in digital banking is the enhanced security it provides. By requiring users to verify their identity using their unique biometric traits, banks can significantly reduce the risk of unauthorized access to their customers' accounts. Biometric authentication is particularly useful in protecting against common cyber threats such as phishing, identity theft, and brute-force attacks, which often target weak or stolen passwords.

Another benefit of biometric authentication is the improved user experience it offers. Instead of requiring customers to remember and enter complex passwords or PINs, biometric authentication allows users to access their accounts quickly and conveniently using their fingerprints, facial features, or voice. This streamlined process not only enhances the overall customer experience but also encourages the adoption of digital banking services.

However, the implementation of biometric authentication also raises certain concerns related to privacy and data security. As biometric data is highly sensitive and unique to each individual, it is crucial that banks and other organizations employing biometric authentication take necessary measures to protect this information. This includes ensuring that biometric data is stored securely, encrypted, and not shared with unauthorized parties.

Moreover, biometric authentication systems are not immune to potential vulnerabilities or hacking attempts. For example, sophisticated attackers may try to bypass facial recognition systems using high-quality masks or 3D-printed replicas. To mitigate these risks, banks may employ multi-factor authentication methods, which combine biometrics with other security measures, such as passwords or one-time codes, to provide a more robust layer of protection.

The application of AI and ML technologies in mobile and online banking has led to significant advancements in the way financial institutions provide services, enhance customer experience, and improve operational efficiency (Donepudi, 2017). By leveraging AI and ML, banks can develop intelligent systems that can analyze vast amounts of data, make predictions, and adapt to new information. Here are some key applications of AI and ML in mobile and online banking (Figure 1):

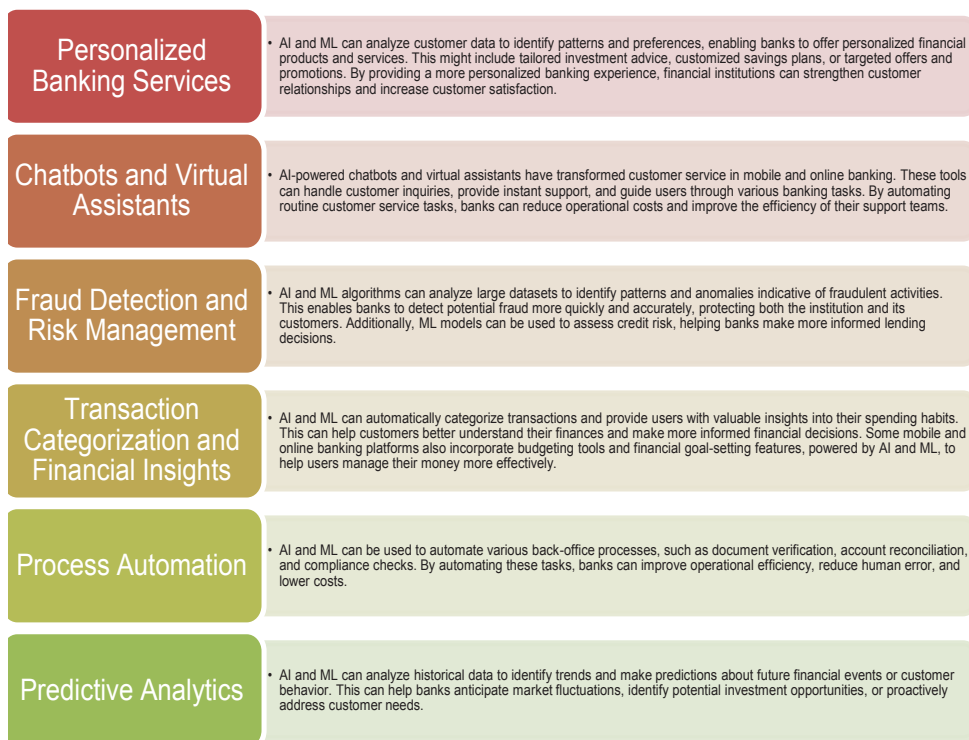


Fig 1. Essential uses of artificial intelligence (AI) and machine learning (ML) in mobile and online banking (Bhatore, 2020; Basdekis, 2022)

AI and ML have brought about significant advancements in mobile and online banking, allowing financial institutions to offer innovative services, enhance customer experience, and improve operational efficiency. By leveraging these technologies, banks can better understand and serve their customers, stay ahead of emerging trends, and remain competitive in the ever-evolving digital landscape.

Open banking refers to the practice of sharing customer financial data securely and with their consent among different financial institutions through open Application Programming Interfaces (APIs). This practice enables the development of innovative and customer-centric financial products and services that can enhance the overall banking experience. API integration is an essential aspect of open banking, allowing different systems and applications to communicate with each other seamlessly.

API integration enables banks to connect their systems and data with third-party applications and services, such as personal finance management tools, investment apps, or credit reporting services. This integration facilitates the sharing of data and enables the development of more customized and personalized financial products and services.

One of the main benefits of open banking and API integration is the enhanced

customer experience. With APIs, banks can offer a more holistic view of their customers' financial lives, allowing them to provide more personalized services that better meet their customers' needs. For example, a customer's financial data from multiple banks can be integrated into a single dashboard, enabling them to view all their accounts and transactions in one place.

Another advantage of open banking and API integration is the increased competition and innovation in the financial services industry. By enabling third-party developers to access customer data and build new products and services, open banking fosters a more competitive and dynamic market that benefits consumers.

However, there are also potential risks associated with open banking and API integration, particularly related to data privacy and security. Banks must ensure that customer data is only shared with authorized third-party providers and that appropriate security measures are in place to protect that data from unauthorized access or misuse.

Open banking and API integration offer significant opportunities for banks and financial services providers to enhance the customer experience and promote innovation in the industry. By leveraging APIs to share data securely, banks can offer more personalized services and enable third-party providers to develop new and innovative financial products and services. However, it is essential for banks to address potential privacy and security concerns to ensure the safety and trust of their customers.

Digital wallets and contactless payments have emerged as a popular and convenient method for consumers to make purchases both online and in-store. A digital wallet is a virtual wallet that stores payment and identification information securely and enables users to make payments via a mobile device. Contactless payments, on the other hand, allow users to make payments by simply tapping their card or mobile device on a payment terminal, without the need for a physical signature or PIN. One of the main benefits of digital wallets and contactless payments is the convenience they offer to consumers. By eliminating the need for physical cards and cash, these payment methods offer a quick and easy way to make purchases. Additionally, digital wallets can store multiple payment options and loyalty cards, allowing users to manage all their payment and rewards information in one place. Digital wallets and contactless payments also offer enhanced security compared to traditional payment methods. With features such as biometric authentication and tokenization, these payment methods provide an extra layer of protection against fraud and unauthorized access. Tokenization replaces sensitive payment information with a unique token, reducing the risk of data breaches and ensuring that customer information is kept secure (Kajdi, 2022).

Moreover, digital wallets and contactless payments have become increasingly popular during the COVID-19 pandemic, as consumers seek to minimize physical contact and reduce the spread of germs. Contactless payments, in particular, have seen a surge in adoption as retailers and businesses promote cashless payment options to minimize the need for physical contact between customers and employees.

However, there are also potential drawbacks to digital wallets and contactless payments. Some consumers may be concerned about the security and privacy of their personal and payment information when using these payment methods. Additionally, some retailers and businesses may not accept digital wallets or contactless payments, limiting their usefulness in certain situations. Digital wallets and contactless payments offer convenience and enhanced security to consumers, particularly in the age of COVID-19 (Hellemans, 2022). As these payment methods continue to evolve and become more widely accepted, they have the potential to transform the way people make purchases and manage their finances.

Blockchain and distributed ledger technologies (DLT) have gained increasing attention in recent years as potential game-changers in the financial industry. Blockchain is a type of DLT that allows for secure, transparent, and decentralized record-keeping of transactions without the need for a central authority. DLT refers to a broader category of technologies that enable the decentralized and secure sharing of information across a network of participants. One of the main benefits of blockchain and DLT is their potential to increase the efficiency and security of financial transactions. By eliminating the need for intermediaries and enabling direct peer-to-peer transactions, these technologies can reduce transaction times, lower costs, and improve transparency. Additionally, the decentralized nature of blockchain and DLT can enhance security by removing a single point of failure or attack. Blockchain and DLT also have the potential to transform various aspects of the financial industry. For example, they can enable the development of new financial products and services, such as decentralized exchanges, smart contracts, and peer-to-peer lending platforms. These technologies can also facilitate cross-border payments and remittances by reducing the need for intermediaries and simplifying compliance with regulations.

However, there are also challenges associated with blockchain and DLT adoption. One of the main challenges is scalability, as these technologies are currently limited in their capacity to handle a large number of transactions. Additionally, there are still concerns around the regulation and governance of blockchain and DLT networks, particularly in regards to data privacy and security.

Big data and analytics play an increasingly important role in the financial industry, enabling banks and financial institutions to analyze large amounts of data to gain insights, make predictions, and improve decision-making. Big data refers to the vast amount of data generated by various sources, such as customer transactions, social media, and internet searches. Analytics refers to the process of analyzing and interpreting this data to extract meaningful insights. One of the main benefits of big data and analytics is the ability to gain a deeper understanding of customers and their behavior. By analyzing customer data, financial institutions can identify patterns and trends, and gain insights into customer needs and preferences. This information can then be used to develop more targeted marketing campaigns, create personalized financial products and services, and improve customer engagement and retention.

Big data and analytics can also be used to detect and prevent fraudulent activities. By analyzing large amounts of data, banks can identify anomalies and patterns indicative of fraudulent transactions, enabling them to take proactive measures to prevent financial losses and protect their customers.

Moreover, big data and analytics can be used to improve operational efficiency and reduce costs. For example, banks can use data analytics to optimize their supply chain, forecast demand, and automate routine tasks, reducing the need for manual intervention and improving productivity.

However, there are also potential risks associated with the use of big data and analytics in the financial industry, particularly related to privacy and security. Banks must ensure that customer data is collected and analyzed in compliance with relevant regulations and that appropriate security measures are in place to protect that data from unauthorized access or misuse.

Big data and analytics have become essential tools for banks and financial institutions to gain insights, improve decision-making, and enhance customer experience. By leveraging the power of big data and analytics, financial institutions can develop more targeted and personalized financial products and services, prevent fraudulent activities, and optimize their operations. However, it is essential for banks to address potential privacy and security concerns to ensure the safety and trust of their customers.

### **Conclusion**

In conclusion, this systematic review of the evolution of mobile and online banking provides valuable insights into the key technologies and innovations shaping the future of digital finance. Our analysis reveals the growing importance of addressing cybersecurity threats, the digital divide, user experience and adoption barriers, as well as the potential impact of emerging technologies such as 5G, IoT, digital currencies, and robo-advisors on the banking industry.

The findings have important implications for both the banking industry and policymakers. Financial institutions must prioritize the development of secure, accessible, and customer-centric services to remain competitive in the digital landscape. This includes investing in robust cybersecurity measures, enhancing digital literacy, and promoting financial inclusion through innovative banking solutions. Policymakers, on the other hand, should focus on creating a supportive regulatory environment that encourages innovation while safeguarding consumer interests.

Future research directions in the field of mobile and online banking could explore the long-term impact of emerging technologies, such as AI and blockchain, on the financial industry. Moreover, researchers may investigate strategies for addressing the digital divide in developing countries, fostering financial inclusion, and optimizing user experience to ensure the successful adoption of digital banking services. Additionally, further studies could examine the role of public-private partnerships in driving innovation and promoting a secure, inclusive, and sustainable digital finance ecosystem.



By understanding the key drivers of change in mobile and online banking, stakeholders can better prepare for the future and harness the power of digital technologies to transform the financial landscape and promote economic growth worldwide.

#### REFERENCES

- Sanli B., Hobikoglu E. (2015). Development of Internet Banking as the Innovative Distribution Channel and Turkey Example. *Procedia - Social and Behavioral Sciences*, 195, 343–352. DOI: 10.1016/j.sbspro.2015.06.362
- Prise J. (2021). The History and Evolution of Mobile Banking. Shoutem. URL: <https://shoutem.com/blog/the-history-and-evolution-of-mobile-banking/>
- Arrighi M. (2014). Going digital: the banking transformation roadmap. Kearney. URL: <https://www.kearney.com/financial-services/article/-/insights/going-digital-the-banking-transformation-roadmap>
- de Lis F., Ortun P.U. (2018). Digital transformation and competition in the financial sector. BBVA Research, 19/02. Pp.1–14. URL: <https://www.bbvaesearch.com/wp-content/uploads/2019/01/Digital-transformation-and-competition-in-the-financial-sector.pdf>
- Liu P., Li H. (2020). Does bank competition spur firm innovation? *Journal of Applied Economics*, 23(1), 519–538. DOI: 10.1080/15140326.2020.1806001
- Gonzalez D. (2015). Currency and Campaigns. *Managing Online Risk*, 185–211. Butterworth-Heinemann. DOI: 10.1016/B978-0-12-420055-5.00008-6
- Norohna G. (2023). Neo Banking Vs Digital Banking. Fi. URL: <https://fi.money/blog/posts/neo-banking-vs-digital-banking>
- Chou Y., Lee C., Chung J. (2004). Understanding m-commerce payment systems through the analytic hierarchy process. *Journal of Business Research*, 57(12), 1423–1430. DOI: 10.1016/S0148-2963(02)00432-0
- Manser Payne E., Peltier J.W., Barger V.A. (2018). Mobile banking and AI-enabled mobile banking: The differential effects of technological and non-technological factors on digital natives' perceptions and behavior. *Journal of Research in Interactive Marketing*, 12(3), 328–346. DOI: 10.1108/JRIM-07-2018-0087
- Fiedler M., Mauerhoefer T., Soller H., Motsch N. (2022). What's new in banking API programs. McKinsey Digital. URL: <https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/tech-forward/whats-new-in-banking-api-programs>
- Horton C. (2023). Best Mobile Banking Apps of April 2023. *Forbes Advisor*. URL: <https://www.forbes.com/advisor/banking/best-mobile-banking-apps/>
- Richter B. (2018). Best mobile banking apps in the world: The top 100 ranked. *Retail Banker International*. URL: <https://www.retailbankerinternational.com/news/best-mobile-banking-apps-in-the-world/>
- Ivashina V., Cekin E. (2019). Kaspı.kz IPO. Harvard Business School Case 220-007. URL: <https://www.hbs.edu/faculty/Pages/item.aspx?num=56825>
- Michaels P. (2022). How Biometric Authentication Secures the Future of Digital Banking. *LuminDigital*. URL: <https://lumindigital.com/lumin-lab/how-biometric-authentication-secures-the-future-of-digital-banking/>
- Donepudi P. (2017). Machine Learning and Artificial Intelligence in Banking. *Engineering International*, 5(2), 83–86. DOI: 10.18034/ei.v5i2.490
- Bhatore S., Mohan L., Reddy Y.R. (2020). Machine learning techniques for credit risk evaluation: a systematic literature review. *J. Bank. Financ. Technol.*, 4, 111–138. DOI: 10.1007/s42786-020-00020-3
- Basdekis C., Christopoulos A., Katsampoxakis I. (2022). FinTech's rapid growth and its effect on the banking sector. *J Bank Financ, Technol*, 6, 159–176. DOI: 10.1007/s42786-022-00045-w
- Hellemans J., Willems K., Brengman M. (2022). Covid-19 and mobile payment in Belgium: Closing the digital divide or just for the young, social, and impulsive? *Electron Commer Res*. DOI: 10.1007/s10660-022-09655-4

UDC 004.8. 004.9

© **Y. Kuchin**<sup>1,2\*</sup>, **N. Yunicheva**<sup>1,3</sup>, **R.I. Mukhamediev**<sup>1,2</sup>,  
**E. Mukhamedieva**<sup>1</sup>, 2023

<sup>1</sup>Institute of Information and Computational Technologies MSHE RK, Almaty;

<sup>2</sup>Satbayev University (KazNITU), Almaty;

<sup>3</sup>Almaty University of Power Engineering and Telecommunications (AUPET),  
Kazakhstan, Almaty.

E-mail: [ykuchin@mail.ru](mailto:ykuchin@mail.ru)

## **ESTIMATION OF THE POSSIBILITY TO SELECT RESERVOIR OXIDATION ZONES BY MACHINE LEARNING METHODS**

**Kuchin** — PhD student, senior researcher of the Institute of Information and Computational Technologies of the National Academy of Sciences of the Republic of Kazakhstan

E-mail: [ykuchin@mail.ru](mailto:ykuchin@mail.ru). ORCID ID: <https://orcid.org/0000-0002-5271-9071>;

**Yunicheva Nadiya** — PhD, leading researcher of the Institute of High Technologies of the National Academy of Sciences of the Republic of Kazakhstan

E-mail: [naduni@mail.ru](mailto:naduni@mail.ru). ORCID ID: <https://orcid.org/0000-0001-6351-3450>;

**Mukhamediev Ravil** — PhD, professor, K.I. Satpayev KazNITU, professor

E-mail: [ravil.muhamedyev@gmail.com](mailto:ravil.muhamedyev@gmail.com), <https://orcid.org/0000-0002-3727-043X>;

**Mukhamedieva Elena** — is a researcher at the Institute of Information and Computational Technologies of the Ministry of Education and Science of the Republic of Kazakhstan

E-mail: [muhamedieva@gmail.com](mailto:muhamedieva@gmail.com). ORCID ID: <https://orcid.org/0000-0001-9596-4432>.

**Abstract.** Nuclear power, despite the environmental risks involved, remains one of the cleanest ways to meet the growing demand for energy without increasing greenhouse gas emissions. In in-situ leaching (ISL) uranium mining in Kazakhstan, uranium reserves are determined by converting the recorded natural gamma radiation of radium and other uranium decay products to uranium content. In zones of reservoir oxidation (ROZ, in the presence of pronounced gamma anomalies, uranium is completely absent, so they must be taken into account when calculating uranium reserves. Not accounting for reservoir oxidation zones (ROZ when planning uranium mining by in-situ leaching is one of the main reasons why estimated reserves cannot be confirmed and often leads to huge material losses, when entire mining blocks turn out to be empty. On the other hand, the selection of the LCP by the available methods is expensive and time-consuming, and is not always performed in a timely way. The paper considers the possibility of separating the ROSs using machine learning methods. For this purpose, the

wells were divided into three classes depending on the characteristics of the ROZ. RandomForestClassifier showed an average accuracy of 0.55, but the values varied greatly depending on the class of wells.

**Keywords:** uranium mining, well geophysical survey, machine learning, classification, boosting

**Acknowledgements.** *The work was carried out with the financial support of the Committee of Science of the Ministry of Science and Higher Education of the Republic of Kazakhstan (grant: № AP14869110 «Improving the accuracy of solving problems of interpretation of geophysical well research data on uranium deposits using machine learning methods»).*

© Я. Кучин<sup>1,2\*</sup>, Н. Юничева<sup>1,3</sup>, Р.И. Мухамедиев<sup>1,2</sup>, Е. Мухамедиева<sup>1</sup>, 2023

<sup>1</sup>Ақпараттық және есептеуіш технологиялар институты, Алматы;

<sup>2</sup>Сатбаев университеті, Алматы;

<sup>3</sup>Ф. Дәукеев атындағы Алматы энергетика және байланыс университеті, Алматы.

## МАШИНАЛЫҚ ОҚЫТУ ӘДІСТЕРІМЕН ҚАБАТТЫҢ ТОТЫҒУ АЙМАҚТАРЫН ОҚШАУЛАУ МҮМКІНДІГІН БАҒАЛАУ

**Кучин Ян Игоревич** — PhD студенті, ҚР БҒМ ҒК АЕТИ аға ғылыми қызметкері

E-mail: [ykuchin@mail.ru](mailto:ykuchin@mail.ru). ORCID идентификаторы: <https://orcid.org/0000-0002-5271-9071>;

**Юничева Надия Рафкатовна** — т. ғ. к., доцент, ҚР БҒМ ҒК АЕТИ жетекші ғылыми қызметкері

E-mail: [naduni@mail.ru](mailto:naduni@mail.ru). ORCID идентификаторы: <https://orcid.org/0000-0001-6351-3450>;

**Мұхамедиев Равиль Ильгизович** — инженерия ғылымдарының докторы, ҚазҰТУ профессоры

E-mail: [ravil.muhamedyev@gmail.com](mailto:ravil.muhamedyev@gmail.com), <https://orcid.org/0000-0002-3727-043X>;

**Мұхамедиева Елена** — ҚР БҒМ ҒК АЕТИ ғылыми қызметкері

E-mail: [muhamedijeva@gmail.com](mailto:muhamedijeva@gmail.com). ORCID/<https://orcid.org/0000-0001-9596-4432>.

**Аннотация.** Атом энергиясы, ілеспе экологиялық тәуекелдерге қарамастан, парниктік газдар шығарындыларын арттырмай, энергияға өсіп келе жатқан сұранысты қамтамасыз етудің ең таза әдістерінің бірі болып қала береді. Атом электр станцияларының жұмысын қамтамасыз ету үшін уран кенін өндіру қажет. Қазақстанда жерасты ұңғымалық шаймалау (ЖҰШ) әдісімен уран өндіру кезінде уран қорлары радийдің және уранның ыдырауының басқа да өнімдерінің тіркелген табиғи гамма-сәулеленуін уран құрамына қайта есептеу жолымен айқындалады. Қабаттың тотығу аймақтарында (ҚТА) айқын гамма-аномалиялары болған кезде уран мүлдем жоқ, сондықтан оларды уран қорларын есептеу кезінде ескеру қажет. Жерасты ұңғымаларын шаймалау әдісімен уран өндіруді жоспарлау кезінде қабаттың тотығу аймақтарын (ҚТА) есепке алмау есептік қорларды растай алмаудың негізгі себептерінің бірі болып табылады және көбінесе бүкіл тау-кен блоктары бос болған кезде үлкен материалдық шығындарға әкеледі. Сонымен қатар, ҚТА -ны қолданыстағы әдістермен бөлу қымбат және еңбекті қажет етеді, әрдайым уақтылы орындалмайды

Мақалада ҚТА-ны машиналық оқыту әдістерімен бөліп көрсету мүмкіндігі қарастырылады. Ол үшін ұңғымалар ҚТА сипаттамаларына байланысты үш сыныпқа бөлінді. RandomForestClassifier ҚТА 0.55 шығарудың орташа дәлдігін көрсетті, бірақ мәндер ұңғымалар класына байланысты айтарлықтай өзгерді.

**Түйін сөздер:** уран өндіру, ұңғымаларды геофизикалық зерттеу, машиналық оқыту, жіктеу, бустинг

*Алғыс. Жұмыс Қазақстан Республикасы Ғылым және жоғары білім министрлігі Ғылым комитетінің қаржылық қолдауымен орындалды (грант: № АР14869110 «Машиналық оқыту әдістерінің көмегімен уран кен орындарындағы ұңғымалардың геофизикалық зерттеулерінің деректерін түсіндіру міндеттерін шешу дәлдігін арттыру».*

© Я. Кучин<sup>1,2\*</sup>, Н. Юничева<sup>1,3</sup>, Р.И. Мухамедиев<sup>1,2</sup>, Е. Мухамедиева<sup>1</sup>, 2023

<sup>1</sup>Институт информационных и вычислительных технологий МНВО РК,  
Алматы;

<sup>2</sup> КазНИТУ им. К.И. Сатпаева, Алматы;

<sup>3</sup>Алматинский Университет энергетики и связи им. Г. Даукеева, Алматы.  
E-mail: ykuchin@mail.ru, naduni@mail.ru

## **ОЦЕНКА ВОЗМОЖНОСТИ ВЫДЕЛЕНИЯ ЗОН ПЛАСТОВОГО ОКИСЛЕНИЯ МЕТОДАМИ МАШИННОГО ОБУЧЕНИЯ**

**Кучин Ян Игоревич** — PhD студент, старший научный сотрудник ИИВТ КН МОН РК

E-mail: ykuchin@mail.ru. ORCID ID: <https://orcid.org/0000-0002-5271-9071>;

**Юничева Надия Рафкатовна** — к.т.н., доцент, ведущий научный сотрудник ИИВТ КН МОН

РК; E-mail: naduni@mail.ru. ORCID ID: <https://orcid.org/0000-0001-6351-3450>;

**Мухамедиев Равиль Ильгизович** — д. инж. н., профессор КазНИТУ им. К.И. Сатпаева,  
профессор

E-mail: ravil.muhamedyev@gmail.com, <https://orcid.org/0000-0002-3727-043X>;

**Мухамедиева Елена** — научный сотрудник ИИВТ МОН РК

E-mail: muhamedijeva@gmail.com. ORCID ID: <https://orcid.org/0000-0001-9596-4432>.

**Аннотация.** Ядерная энергия несмотря на сопутствующие экологические риски остается одним из наиболее чистых способов обеспечения растущего спроса на энергию без увеличения выбросов парниковых газов. Для обеспечения работы атомных электростанций необходима добыча урановой руды. При добыче урана методом подземного скважинного выщелачивания (ПСВ) в Казахстане, запасы урана определяются путем пересчета зарегистрированного естественного гамма-излучения радия и других продуктов распада урана в содержание урана. В зонах пластового окисления (ЗПО), при наличии ярко выраженных гамма-аномалий уран полностью отсутствует, поэтому их необходимо учитывать при подсчете запасов урана. Не учёт зон пластового окисления (ЗПО) при планировании добычи урана методом подземного скважинного выщелачивания является одной из основ-

ных причин невозможности подтвердить расчетные запасы и часто приводит к огромным материальным потерям, когда целые добычные блоки оказываются пустыми. В то же время, выделение ЗПО существующими методами является дорогим и трудозатратным и не всегда выполняется своевременно. В статье рассматривается возможность выделения ЗПО методами машинного обучения на основе данных каротажа. Для этого скважины были разделены на три класса в зависимости от характеристик ЗПО. RandomForestClassifier показал среднюю точность выделения ЗПО 0.55, однако значения сильно варьировались в зависимости от класса скважин.

**Ключевые слова:** добыча урана, геофизические исследования скважин, машинное обучение, классификация, бустинг

**Благодарности.** Работа выполнена при финансовой поддержке Комитета науки Министерства науки и высшего образования Республики Казахстан Республики Казахстан (грант: № AP14869110 "Повышение точности решения задач интерпретации данных геофизических исследований скважин на урановых месторождениях с использованием методов машинного обучения).

## Введение

Особенности формирования месторождений предопределяют их радиологическую обстановку и элементный состав. При этом, как правило, в рудной залежи наблюдается недостаток радия в сравнении с равновесным состоянием, а в ее обрамлении радиоактивное равновесие нарушено в сторону избытка радия, что является следствием образования т.н. «остаточных» и «диффузионных» ореолов радия. Это связано с тем, что, месторождения пластово-инфильтрационного типа формируются в осадочных проницаемых толщах горных пород на границе окислительно-восстановительного барьера. А поскольку в окислительной и восстановительной обстановках поведение подвижных форм урана и радия существенно различаются, в различных морфологических элементах рудных тел в результате процессов «выноса-привноса» «материнского» урана и «дочернего» радия возникают геохимические зоны, где соотношения массовых долей радия и урана отличаются от значений, соответствующих состоянию радиоактивного равновесия между ними.

Состояние радиоактивного равновесия между радием и ураном принято характеризовать коэффициентом нарушения радиоактивного равновесия (или просто коэффициентом радиоактивного равновесия)  $K_{pp}$ , который равен отношению массовых долей радия и урана.

Таким образом, значение величины  $K_{pp} = 1$  соответствует наличию радиоактивного равновесия, а отличие значений  $K_{pp}$  от единицы фиксирует наличие систем, не достигших равновесия либо претерпевших нарушения их замкнутости.

В разрезе рудное тело на гидрогенных месторождениях урана имеет вид

ролла, перемещающегося в направлении движения пластовых вод (рис. 1), а изменение  $K_{pp}$  подчиняется следующим основным закономерностям (Domarenko, 2011):

1) средние значения  $K_{pp}$  для различных морфологических элементов рудных тел, участков и геохимических зон месторождений (с содержанием урана более 0,01 %) меняются в достаточно широких пределах — от 0,60 до 1,0;

2) непосредственно за фронтом пластового окисления радиоактивное равновесие смещено в сторону избытка радия ( $K_{pp} = 1,5-2,5$  и более) - вплоть до практически полного отсутствия урана. Это так называемые «остаточные» радиевые ореолы;

3) по мере перехода в зону восстановления радиоактивное равновесие постепенно от равновесных руд (вблизи зоны пластового окисления — ЗПО) смещается в сторону избытка радия, формируя на границе рудных тел с содержанием урана 0,01 % и выше маломощные (0,2–0,4 м) области радиевых оторочек (т. н. «диффузионные» радиевые ореолы), окаймляющие рудное тело.

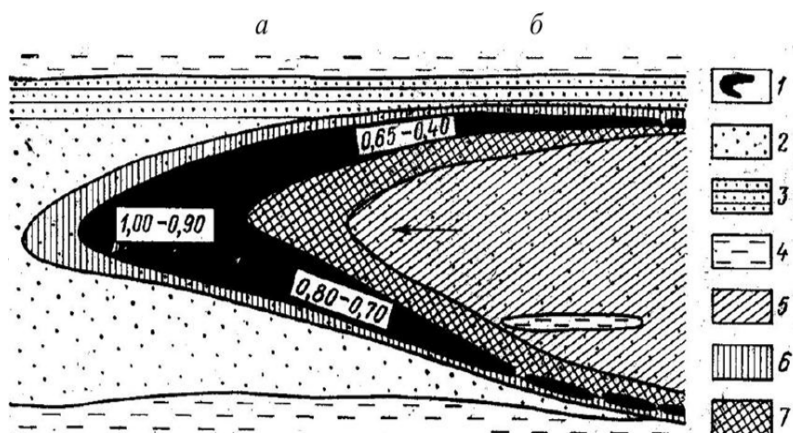


Рис.1. Схема радиологической зональности в разрезе рудовмещающего горизонта пластово-инфильтрационных месторождений урана.

Fig. 1. Schematic of radiological zoning in the cross section of the ore-bearing horizon of layer-infiltration uranium deposits.

На рисунке 1 стрелка указывает направление движения пластовых вод, числа - значение  $K_{pp}$ , а - зона восстановления, б - зона окисления: 1 - урановое рудное тело; 2 - пески; 3 - глинистые песчаники; 4 - глины, алевролиты; 5 - окисленные породы (зона пластового окисления); б - диффузионный ореол радия; 7 - остаточный ореол радия.

В настоящее время содержание урана определяется путем деления содержания радия, полученного в результате интерпретации гамма-каротажа (ГК), на  $K_{pp}$  (Инструкция, 2006).

При ГК регистрируется естественное гамма-излучение радия и других продуктов распада урана. Поскольку в зонах пластового окисления  $K_{pp}=\infty$ , т.е. при наличии ярко выраженных гамма-аномалий, уран полностью отсутствует, их необходимо учитывать при интерпретации гамма-каротажа. ЗПО можно выделять при анализе керна на этапе разведки, а затем строить геологические разрезы и протягивать их по геологическим разрезам, как показано на рис. 2.

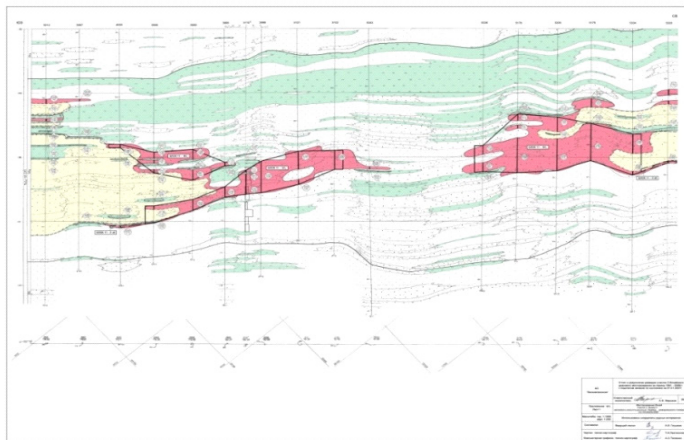


Рис. 2. Геологический разрез, зоны ЗПО (выделены желтым) справа и слева.

Fig. 2. Geologic cross-section, the ROZ zones (highlighted in yellow) on the right and left.

Однако отбор керна и проведение лабораторных исследований — долгий и дорогостоящий процесс. Кроме того, построение разрезов и экстраполяция выделенных ЗПО не всегда производится своевременно и корректно, требует высокой квалификации и значительного ручного труда. В результате, зачастую первоначальная интерпретация проводится без учета ЗПО, а потом приходится делать перерасчет уже с учетом ЗПО, что приводит к существенному уменьшению первоначально подсчитанных запасов урана, на которых было основано все планирование добычи.

Другим способом выделения ЗПО является проведение каротажа нейтронов деления (КНД), позволяющим определять содержание урана в естественном залегании, исключая ошибки, связанные с изменением коэффициента радиоактивного равновесия в процессе выщелачивания урана, а также дает возможность оценивать геотехнологические параметры (влажность, глинистость, пористость). Также этот метод позволяет непосредственно определять содержания урана, минуя стадию пересчета содержания радия в содержание урана через  $K_{pp}$ . В таких случаях ЗПО выделяются там, где рудным интервалам по гамма-каротажу не соответствуют рудные интервалы по КНД.

Пример переинтерпретации ГК после выделения ЗПО по результатам КНД приведен на рис. 3.

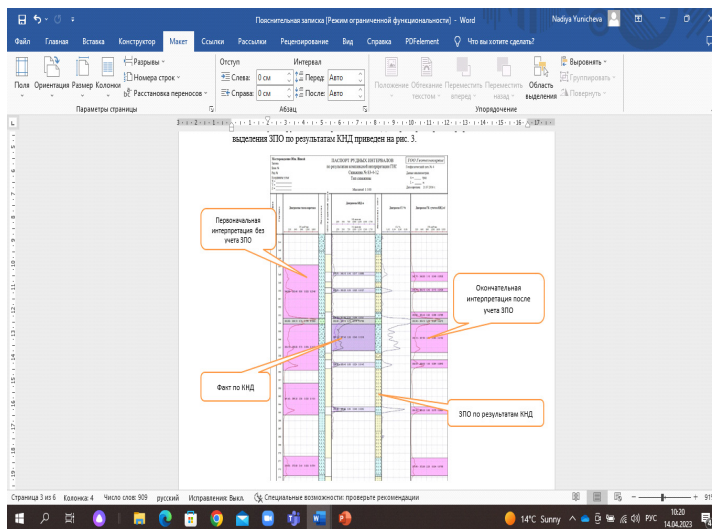


Рис. 3. Пример переинтерпретации ГК после выделения ЗПО по результатам КНД.  
Fig. 3. Example of re-interpretation of the GC after selecting an ROZ based on the results of the KND.

На рисунке видно, что фактические рудные интервалы по КНД оказались значительно меньше рассчитанных по радию. Однако проведение КНД стоит дорого, скорость каротажа составляет не более 50 м/ч (скорость проведения ГК составляет до 1000 м/ч), но главное ограничение состоит в том, что ресурс используемой трубки генератора нейтронов крайне ограничен. Поэтому КНД заказывают не на всех месторождениях и в небольших объемах (5–10 %) от общего количества скважин.

### Методы и материалы

Неучет ЗПО является одной из основных причин неподтверждения запасов и часто приводит к существенным материальным потерям, когда целые блоки оказываются пустыми. В то же время на данный момент нет быстрого и надежного способа выделения ЗПО для корректной интерпретации ГК. Создание формализованного алгоритма выделения ЗПО также представляется невозможным.

В этой связи одним из перспективных направлений решения данной проблемы может стать машинное обучение (МО), когда предсказание расположения ЗПО будет осуществляться путем анализа всей совокупности имеющейся геологической и геофизической информации, путем выявления закономерностей, не замечаемых людьми. Авторам не удалось найти публикации на данную тему, хотя в целом, машинное обучение широко применяется в области анализа геологических данных (Mukhamediev, 2020; Mukhamediev, 2022; Nguyen, 2019; Mukhamediev, 2021; Cracknell, 2014; Kumar, 2020; Chengxiang, 2017; Sun, 2020; Merembayev, 2019; Amirgaliev, 2013; Mukhamediev, 2014; Kuchin, 2019; Kuchin, 2020).



Для оценки возможности применения МО для выявления ЗПО предложено провести целый ряд исследований, возможная схема которых приведена на рис 4.

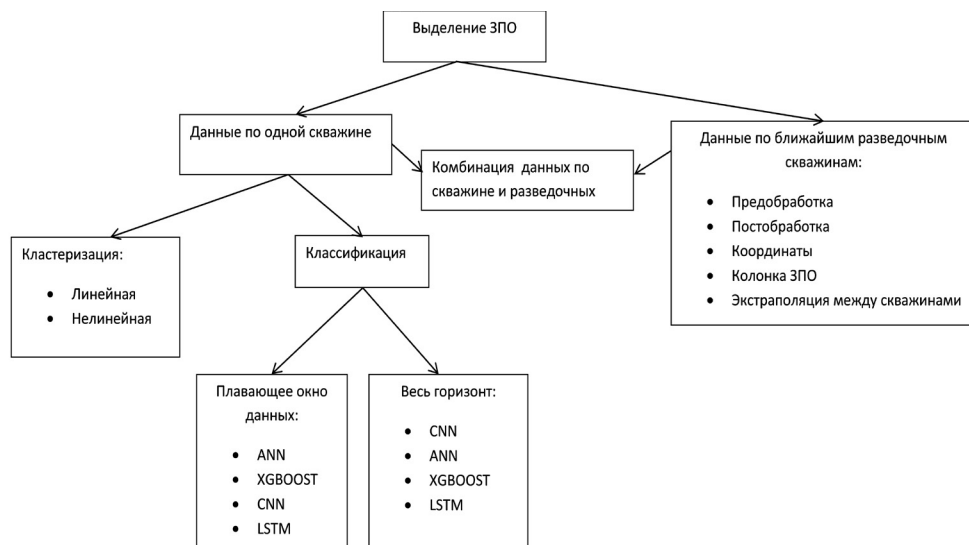


Рис. 4. Схема исследования применимости методов МО для выделения ЗПО  
Fig. 4. The research scheme on the applicability of ML methods for the selection of the ROZ

Таким образом, методологическая схема исследования состоит из следующих этапов:

Создание специального набора данных с разделением скважин на три класса: 42 скважины без ЗПО (LOW\_ZPO), 84 скважины с долей ЗПО 5–50 % от рудовмещающего горизонта (MEDIUM ZPO) и 42 скважины с долей ЗПО более 50 % рудовмещающего горизонта (HI\_ZPO).

В качестве входных параметров использовались данные гамма-каротажа, электрокаротажа (КС, ПС), литология (поскольку ЗПО может находиться только в проницаемых породах), глубина и координаты устья скважины, а также данные по ближайшей соседней скважине.

Разработка методов предобработки данных;

Применение методов машинного обучения.

В качестве классификатора использовался RandomForestClassifier (RFC), который представляет собой ансамбль деревьев решений (Ho, 1995) (рис. 5). Это один из самых гибких и простых в использовании алгоритмов. RFC создает деревья решений для случайно выбранных семплов данных, получает прогноз от каждого дерева и выбирает наилучшее решение посредством голосования. Он также предоставляет довольно эффективный критерий важности показателей (признаков).

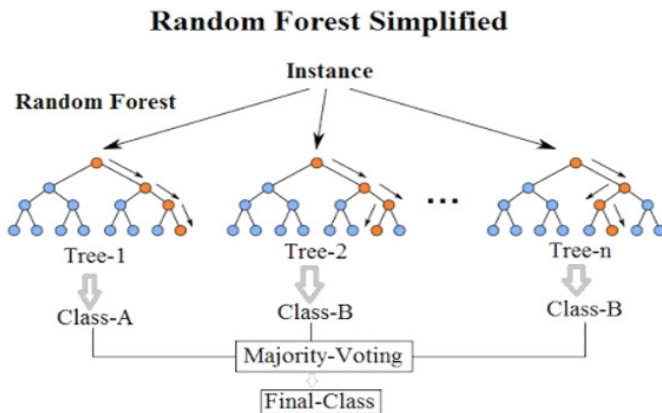


Рис. 5. Классификатор RandomForestClassifier  
 Fig. 5. RandomForestClassifier

### Обсуждение и результаты

На данный момент проведены предварительные эксперименты применения модели RandomForestClassifier на плавающем окне данных. В ходе экспериментов, точность определения ЗПО сильно варьировалась в зависимости от класса скважины, но в среднем составила 0.55, что позволяет сделать вывод о принципиальной возможности решения задачи выделения ЗПО методами МО. Для оценки работы классификатора был разработан простейший визуализатор, который показывает фактические (синие) и предсказанные (красные) трех классов: 1-проницаемые породы, 2-непроницаемые породы, 8 – ЗПО, при этом ось Y представляет собой глубину в пределах рудовмещающего горизонта с шагом 10см. Примеры результатов выделения ЗПО с помощью вышеупомянутой модели на разных классах скважин (LOW\_ZPO, MEDIUM\_ZPO и HI\_ZPO) приведены на рис. 6, 7 и 8.

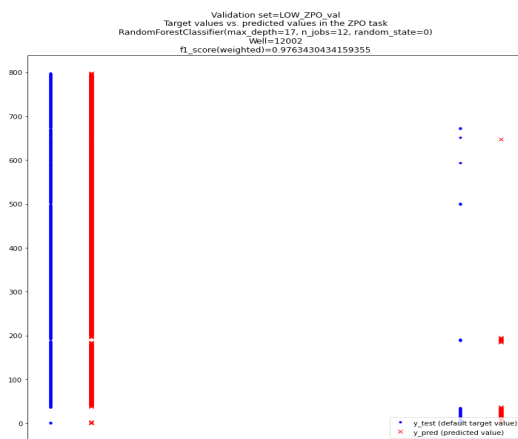


Рис. 6. Пример фактических и предсказанных классов пород для скважины LOW\_ZPO  
 Fig. 6. Example of real and forecasted rock classes for well LOW\_ZPO

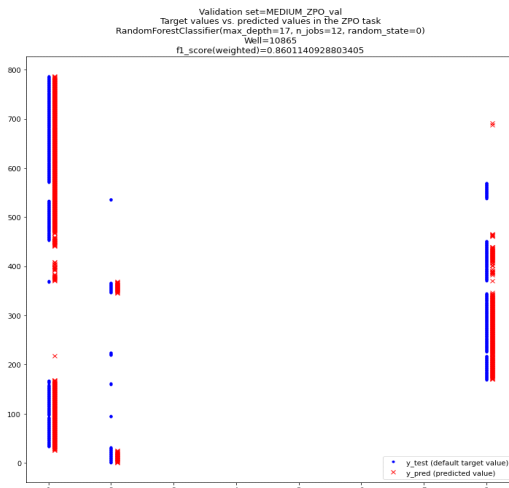


Рис. 7. Пример фактических и предсказанных классов пород для скважины MEDIUM\_ZPO  
Fig. 7. Example of real and forecasted rock classes for well MEDIUM\_ZPO

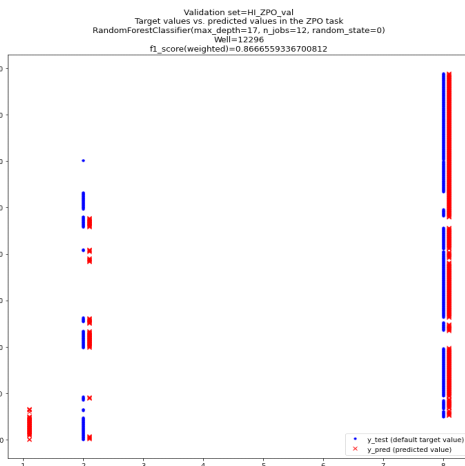


Рис. 8. Пример фактических и предсказанных классов пород для скважины HI\_ZPO  
Fig. 8. Example of real and forecasted rock classes for well HI\_ZPO

**Заклучение**

Выделение ЗПО требует интенсивной работы разных специалистов и требует значительного времени. Зачастую геологические разрезы, содержащие ЗПО не делаются вовремя и недоступны на этапе интерпретации гамма-каротажа и подсчета запасов. Альтернативный метод, каротаж нейтронов деления (КНД-м), ввиду технической сложности и дороговизны выполняется лишь на 1–5 % скважин. Игнорирование ЗПО может привести к неправильной оценке запасов, т.е. при фактическом отсутствии урана, запасы, подсчитанные по принятой методике по гамма-активности его продуктов распада, могут показаться значительными. Это приведет к бурению и обустройству десятков

или даже сотен «пустых» скважин, финансовые потери могут составить от 5 до 10 млн долларов в год.

В целом, неточности в определении литологического состава, фильтрационных характеристик и ЗПО приводят к ошибкам в технологическом процессе установки фильтров, ошибкам в определении запасов руды и в конечном счете существенным финансовым потерям. Существующие методики решения перечисленных проблем разработаны начиная с 70-х годов прошлого века, не поддаются существенной модернизации и требуют большое количество высокопрофессионального ручного труда.

Таким образом, для преодоления вышеперечисленных трудностей, в данной статье исследован вопрос возможности выделения зон пластового окисления методами машинного обучения.

Для проверки справедливости данной гипотезы создан специальный набор данных с разделением скважин на три класса: скважины без ЗПО, скважины с долей ЗПО 5–50 % от рудовмещающего горизонта и скважины с долей ЗПО более 50% рудовмещающего горизонта.

В качестве входных параметров использовались данные гамма-каротажа, электрокаротажа (КС, ПС), литология (поскольку ЗПО может находиться только в проницаемых породах), глубина и координаты устья скважины, а также данные по ближайшей соседней скважине. Далее проведена процедура разработки методов предобработки данных и применены методы машинного обучения.

В результате экспериментов можно сделать вывод о возможности применения методов машинного обучения для выделения ЗПО. Тем не менее, учитывая важность и ответственность выделения ЗПО при подсчете запасов, для оценки возможности промышленного применения алгоритма выделения ЗПО методами МО, необходимо провести все исследования согласно схеме на рис. 4, а также дать ответы следующие вопросы:

- Какова минимальная требуемая точность при выделении ЗПО.
- Можно ли достичь требуемую точность?
- Нужны ли для этого дополнительные данные?
- Достижима ли требуемая точность, если использовать только данные по одной скважине, без учета соседних?
- Transfer Learning. Можно ли перенести модель, разработанную на данных одного месторождения на другие месторождения?

Для ответа на перечисленные вопросы исследования по данному направлению будут продолжены до окончания срока реализации проекта в следующем году.

#### **ЛИТЕРАТУРА**

Амиргалиев Е., Исаков С., Кучин Я., Мухамедиев Р. (2013). Методы машинного обучения в задачах распознавания пород на урановых месторождениях // Известия НАН РК: Серия физико-математических наук. -2013. - № 3. – С.82–88.

Домаренко В.А. (2011). Рациональные методы поисков и геолого-экономической оценки

месторождений руд редких и радиоактивных элементов. Часть 1. Прогнозирование, поиски и оценка: Учебное пособие. 220с. (на русском).

Инструкция (методические рекомендации) по подземному выщелачиванию урана (2006). 315 с. (на русском).

Крекнелл М. (2014). Машинное обучение для геологического картирования: Алгоритмы и приложения". Докторская диссертация, Школа наук о Земле, Университет Тасмании, Хобарт, Австралия.

Кумар С., Чаттерджи С., Ооммен Т., Гуха А. (2020). Автоматизированное литологическое картирование путем интеграции методов спектрального усиления и алгоритмов машинного обучения с использованием гиперспектральных данных AVIRIS-NG в золотоносных гранитно-зеленокаменных породах в Хатти, Индия // *Международный журнал прикладных наблюдений Земли. Геоинф.* – 2020. - № 86, статья 102006.

Кучин Я., Якунин К., Мухамедиева Е., Мухамедиев Р. (2020). Проект по созданию классификатора литологических типов для урановых месторождений Казахстана // *Физика: Серия конференций.* – 2020. - № 1405(1). – С. 012001.

Кучин Я., Мухамедиев Р., Якунин К. (2019). Качество классификации данных при непоследовательных экспертных оценках // *Облако науки.* -2019. - № 6(1). -С.109–126. (На русском).

Мерембаев Т., Юнусов Р., Едилхан А. (2019). Алгоритмы машинного обучения для классификации стратиграфии на урановых месторождениях // *Материалы по компьютерным наукам.* – 2019. -№ 150.- С.46–52.

Мухамедиев Р., Амиргалиев Е., Исаков С., Кучин Я., Мухамедиева Е. (2014). Интеграция результатов алгоритмов распознавания на урановых месторождениях. *Журнал передового вычислительного интеллекта и интеллектуальной информатики*, 18(3):347–352.

Мухамедиев Р.И. и др. (2022). Оценка фильтрационных свойств вмещающих пород в урановых месторождениях песчаникового типа с использованием методов машинного обучения // *IEEE Access.* – 2022. - №10. – С.18855–18872. DOI: 10.1109/ACCESS.2022.3149625.

Мухамедиев Р.И., Кучин Я.А., Якунин К.О., Мухамедиева Е.Л., Костарев С.В. (2020). Предварительные результаты оценки литологических классификаторов для урановых месторождений инфильтрационного типа // *Облако науки.* -2020. -№7(2). -С.258–272. (На русском).

Мухамедиев Р.И., Сымагулов А., Кучин Ю., Якунин К. и Елис М. (2021). От классического машинного обучения к глубоким нейронным сетям: Упрощенный наукометрический обзор // *Прикладн. науки.* – 2021. -№ 11(12). -С.554.

Нгуен Г. (2019). Среды и библиотеки машинного обучения и глубокого обучения для крупномасштабного интеллектуального анализа данных: опрос // *Обзор искусственного интеллекта.* -2019. -№ 52(1). -С.77–124.

Сан Т., Х. Ли, К. Ву, Ф.Ч., З. Жу, З. Ху (2020). Моделирование перспективности полезных ископаемых на основе данных с использованием методов машинного обучения и глубокого обучения: Исследование на примере южной провинции Цзянси, Китай // *Минералы.* -2020. -№ 10(2). -С.102.

Чэнсян Д., Хэпин П., Синан Ф., Конате А., Руйдун Ц. (2017). Машина опорных векторов как альтернативный метод для классификации литологии кристаллических пород // *Журнал геофизики и инженерии.* -2017. -№ 14(2). – С. 341–349.

Кучин Я., Мухамедиев Р., Якунин К., Грундспенкис Дж. Сымагулов А. (2020). Оценка влияния экспертной маркировки обучающих данных на качество автоматической классификации литологических групп с помощью искусственных нейронных сетей // *Прикладные компьютерные системы.* -2020. -№ 25(2). -С145–152, <https://doi.org/10.2478/acss-2020-0016>.

Кучин Ю.И., Мухамедиев Р.И., Якунин К.О. (2020). Один из методов генерации синтетических данных для оценки верхней границы производительности алгоритмов машинного обучения // *Cogent Engineering.* – 2020. - № 7(1). -С.1718821.

Хо Т. (1995). Метод случайного леса. 3-я международная конференция по анализу и распознаванию документов IEEE". -1995. -№ 1. -С.278–282.

#### REFERENCES

Amirgaliev E., Iskakov S., Kuchin Y., Mukhamediev R. (2013). Machine Learning Methods in the Problems of Rock Recognition at Uranium Deposits //News of the NAS RK. Physics and Mathematics Series, 3, 82–88.

Chengxiang D., Heping P., Sinan F., Konaté A., Ruidong Q. (2017). Support vector machine as an alternative method for lithology classification of crystalline rocks //Journal of Geophysics and Engineering, 14(2), 341–349.

Cracknell M. (2014). Machine learning for geological mapping: Algorithms and applications." Ph.D. dissertation, School of Earth Science, University of Tasmania, Hobart, Australia.

Domarenko V.A. (2011). Rational methods of search and geological-economic evaluation of deposits of ores of rare and radioactive elements. Part 1. Forecasting, prospecting and evaluation: a training manual. 220p. (In Russ.).

Instruction (methodical recommendations) on underground uranium in-situ leaching" (2006).315p. (In Russ.).

Kuchin Y., Yakunin K., Mukhamedyeva E., Mukhamedyev R. (2019). Project on creating a classifier of lithological types for uranium deposits in Kazakhstan. Journal of Physics: Conference Series, 1405(1), 012001.

Kuchin Y., Mukhamediev R., Yakunin K., Grundspenkis J., Symagulov A. (2020). Assessing the Impact of Expert Labelling of Training Data on the Quality of Automatic Classification of Lithological Groups Using Artificial Neural Networks // Applied Computer Systems, 25(2), 145–152, <https://doi.org/10.2478/acss-2020-0016>.

Kumar C., Chatterjee S., Oommen T., Guha A. (2020). Automated lithological mapping by integrating spectral enhancement techniques and machine learning algorithms using AVIRIS-NG hyperspectral data in Gold-bearing granite-greenstone rocks in Hutti, India //Int. J. Appl. Earth Obs. Geoinf., 86: art. no. 102006.

Merembayev T., Yunussov R., Yedilkhan A. (2019). Machine learning algorithms for stratigraphy classification on uranium deposits //Procedia Computer Science, 150, 46–52.

Muhamediyev R., Amirgaliev E., Iskakov S., Kuchin Y., Muhamedyeva E. (2014). Integration of Results of Recognition Algorithms at the Uranium Deposits //Journal of Advanced Computational Intelligence and Intelligent Informatics, 18(3), 347–352.

Mukhamediev R.I. et al. (2022). Estimation of Filtration Properties of Host Rocks in Sandstone-type Uranium Deposits Using Machine Learning Methods //IEEE Access, 10,18855–18872. DOI: 10.1109/ACCESS.2022.3149625

Mukhamediev R.I., Kuchin Y.A., Yakunin K.O., Mukhamedieva E.L., Kostarev S.V. (2020). Preliminary results of the assessment of lithological classifiers for uranium deposits of the infiltration type // Cloud of Science, 7(2), 258–272. (In Russ.).

Mukhamediev R.I., Symagulov A., Kuchin Y., Yakunin K., Yelis M. (2021). From Classical Machine Learning to Deep Neural Networks: A Simplified Scientometric Review //Appl. Sci., 11(12), 554.

Nguyen G. et al. (2019). Machine Learning and Deep Learning frameworks and libraries for large-scale data mining: a survey // Artificial Intelligence Review, 52(1), 77–124.

Sun T., Li H., Wu K., Chen F., Zhu Z., Hu Z. (2020). Data-Driven Predictive Modelling of Mineral Prospectivity Using Machine Learning and Deep Learning Methods: A Case Study from Southern Jiangxi Province, China. Minerals, 10(2), 102.

Ho T.K. (1995). Random decision forests // 3rd international conference on document analysis and recognition IEEE", 1, 278–282.

## МАЗМҰНЫ

<b>Г. Әбдіқалық, Ә. Мұқанова, А. Назырова</b> CRF ЖӘНЕ RANDOM FOREST МОДЕЛДЕРІНІҢ КӨМЕГІМЕН ҚАЗАҚ ТІЛІНДЕ АТАЛҒАН ОБЪЕКТІЛЕРДІ ТАҢУ: САЛЫСТЫРМАЛЫ ЗЕРТТЕУ.....	7
<b>Г.Б. Абдикеримова, М.Б. Есенова, Т.Т. Оспанова, У.Ж. Айтимова, М. Айтимов</b> ҒАРЫШТЫҚ КЕСКІНДЕРДІ ӨНДЕУДЕ АҚПАРАТТЫҚ ТЕКСТУРАЛЫҚ ЛАВС МАСКАЛАР ӘДІСТЕРІН ҚОЛДАНУ.....	18
<b>Б.У. Асанова, Б.Б. Оразбаев, Ж.Ж. Молдашева, Г.Ж. Шүйтенов, Э.М. Дюсембина</b> ТҮРЛІ СИПАТТАҒЫ ҚОЛ ЖЕТІМДІ АҚПАРАТТАР НЕГІЗІНДЕ БАЯУ КОКСТЕУ ҚОНДЫРҒЫСЫНЫҢ ӨЗАРА БАЙЛАНЫСҚАН ТЕХНОЛОГИЯЛЫҚ АГРЕГАТТАРЫ МОДЕЛЬДЕРІН ҚҰРУ ӘДІСТЕМЕСІ.....	28
<b>Г.Б. Бахадирова, Н. Тасболатұлы, А.С. Муканова, Ш. Тураев</b> MATLAB SIMULINK-ТЕ СЫЗЫҚТЫҚ ЕМЕС ЖҮЙЕ ҮШІН КЕРІ БАЙЛАНЫСТЫ СЫЗЫҚТЫҚ БАСҚАРУДЫ ЖОБАЛАУ.....	44
<b>Е.С. Голенко, А.А. Исмаилова</b> ПРЕДСКАЗАНИЕ ФУНКЦИЙ БЕЛКА С ИСПОЛЬЗОВАНИЕМ КОМБИНАЦИИ VILSTM И АЛГОРИТМА САМОВНИМАНИЯ.....	62
<b>Л.З. Жолшиева, Т.К. Жукабаева, Ш. Тураев, М.А. Бердиева</b> CNN НЕГІЗІНДЕ ҚАЗАҚ ҒЫМ ТІЛІН ТАҢУ.....	76
<b>К.К. Кадиркулов, А.А. Исмаилова, Ә.Б. Бейсегұл</b> ЛАБОРАТОРИЯЛЫҚ ЗЕРТТЕУ НӘТИЖЕЛЕРІН ТАЛДАУ ҮШІН МАШИНАЛЫҚ ОҚЫТУДЫҢ МОДЕЛІН ТАҢДАУ.....	88
<b>А. Муканова, А. Муханова, Т. Оспанова, А. Бакиева, В. Махатова</b> ҚҰЗЫРЕТТІК ТӘСІЛДЕР НЕГІЗІНДЕГІ БІЛІМ БЕРУ БАҒДАРЛАМАЛАРЫН ӨЗІРЛЕУДІҢ МАҢЫЗДЫ АСПЕКТІЛЕРІ.....	99
<b>Ш.Ж. Мусиралиева, М.А. Болатбек, М. Сағынай, Ж.Ы. Елтай, К.Б. Багитова</b> ЭКСТРЕМИСТІК МӘЛІМЕТТЕР ТҮСІНІГІ ЖӘНЕ ЭКСТРЕМИЗМГЕ ҚАРСЫ КҮРЕС ЖОБАЛАРЫНА ЖҮЙЕЛІК ШОЛУ.....	112
<b>Д. Оралбекова, О. Мамырбаев, А. Жунусова, Б. Жұмажанов</b> КҮРДЕЛІ МОРФОЛОГИЯЛЫҚ ҚҰРЫЛЫМЫ БАР ТІЛГЕ АРНАЛҒАН ЗАМАНАУИ ТІЛДІК МОДЕЛЬДЕУ ӘДІСТЕРІН ЗЕРТТЕУ.....	131
<b>Б.Т. Рзаев, Ж.Т. Бельдеубаева, И.М. Увалиева</b> СТЕКИНГ ӘДІСІН ҚОЛДАНУ АРҚЫЛЫ АҚПАРАТТЫҚ ЖЕЛІДЕГІ ЗИЯНДЫ ДЕРЕКТЕРДІ АНЫҚТАУ.....	147
<b>Н.С. Баймулдина, Г.Н. Скабаева, А.Д. Жақсыбаева</b> БИОТЕХНОЛОГИЯ САЛАСЫНДАҒЫ ЖОБАЛАРДЫ БАСҚАРУДЫҢ БАҒДАРЛАМАЛЫҚ ҚАМТАМАСЫЗ ЕТУІ.....	161
<b>А.Ә. Таурбекова, Ө.Ж. Мамырбаев, Б. Т. Қарымсақова, Б. Ж. Жұмажанов</b> МАГМАНЫҢ ШЫҒУ ПРОЦЕСІН ЗЕРТТЕУ.....	176
<b>Г.С. Шаймерденова, Р.А. Саркулакова, М.М. Тұрғанбекова, Б.Ө. Тастанбекова, М.Т. Байжанова,</b> МОБИЛЬДІ ЖӘНЕ ОНЛАЙН-БАНКИНГТЕГІ ЖЕТІСТІКТЕР: ТЕХНОЛОГИЯЛАР МЕН ИННОВАЦИЯЛАРДЫ КЕШЕНДІ ТАЛДАУ.....	193
<b>Я. Кучин, Н. Юничева, Р.И. Мухамедиев, Е. Мухамедиева</b> МАШИНАЛЫҚ ОҚЫТУ ӘДІСТЕРІМЕН ҚАБАТТЫҢ ТОТЫҒУ АЙМАҚТАРЫН ОҚШАУЛАУ МҮМКІНДІГІН БАҒАЛАУ.....	210

## СОДЕРЖАНИЕ

<b>Г. Абдикалык, А. Муканова, А. Назырова</b> РАСПОЗНАВАНИЕ ИМЕНОВАННЫХ ИМЕНОВАННЫХ ОБЪЕКТОВ В КАЗАХСКОМ ЯЗЫКЕ С ПОМОЩЬЮ МОДЕЛЕЙ CRF И RANDOM FOREST: СРАВНИТЕЛЬНОЕ ИССЛЕДОВАНИЕ.....	7
<b>Г.Б. Абдикеримова, М.Б. Есенова, Т.Т. Оспанова, У.Ж. Айтимова, М. Айтимов</b> ИСПОЛЬЗОВАНИЕ МЕТОДОВ ИНФОРМАТИВНОЙ ТЕКСТУРНОЙ МАСОК ЛАВСА ПРИ ОБРАБОТКЕ КОСМИЧЕСКИХ ИЗОБРАЖЕНИЙ.....	18
<b>Б.У. Асанова, Б.Б. Оразбаев, Ж.Ж. Молдашева, Г.Ж. Шуйтенов, Э.М. Дюсембина</b> МЕТОДИКА РАЗРАБОТКИ МОДЕЛЕЙ ВЗАИМОСВЯЗАННЫХ ТЕХНОЛОГИЧЕСКИХ АГРЕГАТОВ УСТАНОВКИ ЗАМЕДЛЕННОГО КОКСОВАНИЯ НА ОСНОВЕ ДОСТУПНОЙ ИНФОРМАЦИИ РАЗЛИЧНОГО ХАРАКТЕРА.....	28
<b>Г.Б. Бахадирова, Н. Тасболатұлы, А.С. Муканова, Ш.Тураев</b> ПРОЕКТИРОВАНИЕ ЛИНЕЙНОГО УПРАВЛЕНИЯ С ОБРАТНОЙ СВЯЗЬЮ ДЛЯ НЕЛИНЕЙНОЙ СИСТЕМЫ В MATLAB SIMULINK.....	44
<b>Е.С. Голенко, А.А. Исмаилова</b> ПРЕДСКАЗАНИЕ ФУНКЦИЙ БЕЛКА С ИСПОЛЬЗОВАНИЕМ КОМБИНАЦИИ VILSTM И АЛГОРИТМА САМОВНИМАНИЯ.....	62
<b>Л.З. Жолшиева, Т.К. Жукабаева, Ш. Тураев, М.А. Бердиева</b> РАСПОЗНАВАНИЕ КАЗАХСКОГО ЖЕСТОВОГО ЯЗЫКА НА ОСНОВЕ CNN.....	76
<b>К.К. Кадиркулов, А.А. Исмаилова, Ә.Б. Бейсегұл</b> ВЫБОР МОДЕЛИ МАШИННОГО ОБУЧЕНИЯ ПО ИНТЕРПРЕТАЦИИ РЕЗУЛЬТАТОВ ЛАБОРАТОРНЫХ ИССЛЕДОВАНИЙ.....	88
<b>А. Мукашова, А. Муханова, Т. Оспанова, А. Бакиева, В. Махагова</b> ВАЖНЫЕ АСПЕКТЫ РАЗРАБОТКИ ОБРАЗОВАТЕЛЬНЫХ ПРОГРАММ, ОСНОВАННЫХ НА КОМПЕТЕНТНОСТНОМ ПОДХОДЕ.....	99
<b>Ш.Ж. Мусиралиева, М.А. Болатбек, М. Сағынай, Ж.Ы. Елтай, К.Б. Багитова</b> ПОНЯТИЕ ЭКСТРЕМИСТСКИХ ДАННЫХ И СИСТЕМНЫЙ ОБЗОР ПРОЕКТОВ ПО БОРЬБЕ С ЭКСТРЕМИЗМОМ.....	112
<b>Д. Оралбекова, О. Мамырбаев, А. Жунусова, Б. Жумажанов</b> ИССЛЕДОВАНИЕ СОВРЕМЕННЫХ МЕТОДОВ ЯЗЫКОВОГО МОДЕЛИРОВАНИЯ ДЛЯ ЯЗЫКА СО СЛОЖНОЙ МОРФОЛОГИЧЕСКОЙ СТРУКТУРОЙ.....	131
<b>Б.Т. Рзаев, Ж.Т. Бельдеубаева, И.М. Увалнева</b> ИДЕНТИФИКАЦИЯ ВРЕДОНОСНЫХ ДАННЫХ В ИНФОРМАЦИОННОЙ СЕТИ С ИСПОЛЬЗОВАНИЕМ МЕТОДА СТЕКИНГА.....	147
<b>Н.С. Баймулдина, Г.Н. Скабаева, А.Д. Жақсыбаева</b> ПРОГРАММНОЕ ОБЕСПЕЧЕНИЕ ДЛЯ УПРАВЛЕНИЯ ПРОЕКТАМИ В ОБЛАСТИ БИОТЕХНОЛОГИИ.....	161
<b>А.А. Таурбекова, О.Ж. Мамырбаев, Б.Т. Карымсакова, Б.Ж. Жумажанов</b> ИССЛЕДОВАНИЯ ПРОЦЕССА ИСТЕЧЕНИЯ МАГМЫ.....	176
<b>Г.С. Шаймерденова, Р.А. Саркулакова, М.М. Турганбекова, Б.О. Тастанбекова, М.Т. Байжанова</b> ДОСТИЖЕНИЯ В МОБИЛЬНОМ И ОНЛАЙН-БАНКИНГЕ: КОМПЛЕКСНЫЙ АНАЛИЗ ТЕХНОЛОГИЙ И ИННОВАЦИЙ.....	193
<b>Я. Кучин, Н. Юничева, Р.И. Мухамедиев, Е. Мухамедиева</b> ОЦЕНКА ВОЗМОЖНОСТИ ВЫДЕЛЕНИЯ ЗОН ПЛАСТОВОГО ОКИСЛЕНИЯ МЕТОДАМИ МАШИННОГО ОБУЧЕНИЯ.....	210



## CONTENTS

<b>G. Abdikalyk, A. Mukanova, A. Nazyrova</b> NAMED ENTITY RECOGNITION FOR KAZAKH LANGUAGE USING CRF AND RANDOM FOREST MODELS: A COMPARATIVE STUDY.....	7
<b>G.B. Abdikerimova, M.B. Yessenova, T.T. Ospanova, U.Zh Aitimova, M. Murat</b> USE OF INFORMATION TEXTURE LAWS MASK METHODS IN SPACE IMAGE PROCESSING.....	18
<b>B. Assanova, B. Orazbayev, Zh. Moldasheva, G. Shuitenov, E. Dyussemina</b> METHODOLOGY FOR DEVELOPING MODELS OF INTERRELATED TECHNOLOGICAL UNITS OF A DELAYED COKING UNIT ON THE BASIS OF AVAILABLE INFORMATION OF A DIFFERENT NATURE.....	28
<b>G.B. Bahadirova, H. Tasbolatuly, A.S. Mukanova, Sh. Turaev</b> DESIGNING LINEAR FEEDBACK CONTROL FOR A NONLINEAR SYSTEM IN MATLAB SIMULINK.....	44
<b>Y.S. Golenko, A.A. Ismailova</b> PROTEIN FUNCTION PREDICTION USING THE COMBINATION OF BILSTM AND SELF-ATTENTION ALGORITHM.....	62
<b>L. Zholshiyeva, T. Zhukabayeva, Sh. Turaev, M. Berdieva</b> KAZAKH SIGN LANGUAGE RECOGNITION BASED ON CNN.....	76
<b>K. Kadirkulov, A. Ismailova, A. Beissegul</b> SELECTION OF A MACHINE LEARNING MODEL FOR INTERPRETING LABORATORY RESULTS.....	88
<b>A. Mukashova, A. Mukanova, T. Ospanova, A. Bakiyeva, V. Makhatova</b> IMPORTANT ASPECTS OF DEVELOPING EDUCATIONAL PROGRAMS BASED ON THE COMPETENCY-BASED APPROACH.....	99
<b>Sh. Mussiraliyeva, M. Bolatbek, M. Sagynay, Zh. Yeltay, K. Bagitova</b> THE CONCEPT OF EXTREMIST DATA AND A SYSTEMATIC REVIEW OF ANTI-EXTREMISM PROJECTS.....	112
<b>D. Oralbekova, O. Mamyrbayev, A. Zhunussova, B. Zhumazhanov</b> STUDY OF MODERN METHODS OF LANGUAGE MODELING FOR A LANGUAGE WITH A COMPLEX MORPHOLOGICAL STRUCTURE.....	131
<b>B. Rzayev, Zh. Beldeubayeva, I. Uvaliyeva</b> IDENTIFICATION OF MALICIOUS DATA IN THE INFORMATION NETWORK BY USING THE STACKING METHOD.....	147
<b>N.S. Baimuldina, G.N. Skabayeva, A. Zhaksybayeva</b> PROJECT MANAGEMENT SOFTWARE IN THE FIELD OF BIOTECHNOLOGY.....	161
<b>A.A. Taurbekova, O.Zh. Mamyrbaev, B.T. Karymsakova, B.Zh. Zhumazhanov</b> INVESTIGATIONS OF MAGMA OUTPUT PROCESS.....	176
<b>G.S. Shaimerdenova, R.A. Sarkulakova, M.M. Turganbekova, B.O. Tastanbekova, M.T. Baizhanova</b> ADVANCEMENTS IN MOBILE AND ONLINE BANKING: A COMPREHENSIVE ANALYSIS OF TECHNOLOGIES AND INNOVATIONS.....	193
<b>Y. Kuchin, N. Yunicheva, R.I. Mukhamediev, E. Mukhamedieva</b> ESTIMATION OF THE POSSIBILITY TO SELECT RESERVOIR OXIDATION ZONES BY MACHINE LEARNING METHODS.....	210

**Publication Ethics and Publication Malpractice  
the journals of the National Academy of Sciences of the Republic of Kazakhstan**

For information on Ethics in publishing and Ethical guidelines for journal publication see <http://www.elsevier.com/publishingethics> and <http://www.elsevier.com/journal-authors/ethics>.

Submission of an article to the National Academy of Sciences of the Republic of Kazakhstan implies that the described work has not been published previously (except in the form of an abstract or as part of a published lecture or academic thesis or as an electronic preprint, see <http://www.elsevier.com/postingpolicy>), that it is not under consideration for publication elsewhere, that its publication is approved by all authors and tacitly or explicitly by the responsible authorities where the work was carried out, and that, if accepted, it will not be published elsewhere in the same form, in English or in any other language, including electronically without the written consent of the copyright-holder. In particular, translations into English of papers already published in another language are not accepted.

No other forms of scientific misconduct are allowed, such as plagiarism, falsification, fraudulent data, incorrect interpretation of other works, incorrect citations, etc. The National Academy of Sciences of the Republic of Kazakhstan follows the Code of Conduct of the Committee on Publication Ethics (COPE), and follows the COPE Flowcharts for Resolving Cases of Suspected Misconduct ([http://publicationethics.org/files/u2/New\\_Code.pdf](http://publicationethics.org/files/u2/New_Code.pdf)). To verify originality, your article may be checked by the Cross Check originality detection service <http://www.elsevier.com/editors/plagdetect>.

The authors are obliged to participate in peer review process and be ready to provide corrections, clarifications, retractions and apologies when needed. All authors of a paper should have significantly contributed to the research.

The reviewers should provide objective judgments and should point out relevant published works which are not yet cited. Reviewed articles should be treated confidentially. The reviewers will be chosen in such a way that there is no conflict of interests with respect to the research, the authors and/or the research funders.

The editors have complete responsibility and authority to reject or accept a paper, and they will only accept a paper when reasonably certain. They will preserve anonymity of reviewers and promote publication of corrections, clarifications, retractions and apologies when needed. The acceptance of a paper automatically implies the copyright transfer to the National Academy of Sciences of the Republic of Kazakhstan.

The Editorial Board of the National Academy of Sciences of the Republic of Kazakhstan will monitor and safeguard publishing ethics.

Правила оформления статьи для публикации в журнале смотреть на сайтах:

**[www.nauka-nanrk.kz](http://www.nauka-nanrk.kz)**

**<http://physics-mathematics.kz/index.php/en/archive>**

**ISSN 2518-1726 (Online),**

**ISSN 1991-346X (Print)**

Подписано в печать 28.09.2023.

Формат 60x881/8. Бумага офсетная. Печать – ризограф.

18,0 п.л. Тираж 300. Заказ 3.