

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

ҚАЗАҚСТАН РЕСПУБЛИКАСЫ  
ҰЛТТЫҚ ҒЫЛЫМ АКАДЕМИЯСЫ

әл-Фараби атындағы Қазақ ұлттық университетінің

# Х А Б А Р Л А Р Ы

---

---

## ИЗВЕСТИЯ

НАЦИОНАЛЬНОЙ АКАДЕМИИ  
НАУК РЕСПУБЛИКИ КАЗАХСТАН  
Қазақстан Республикасының  
Ғылым Академиясының  
Әл-Фараби атындағы  
Қазақ ұлттық университеті

## NEWS

OF THE ACADEMY OF SCIENCES  
OF THE REPUBLIC OF  
KAZAKHSTAN  
al-Farabi Kazakh National University

### PHYSICO-MATHEMATICAL SERIES

#### 4 (344)

OCTOBER – DECEMBER 2022

PUBLISHED SINCE JANUARY 1963

PUBLISHED 4 TIMES A YEAR

ALMATY, NAS RK

## БАС РЕДАКТОР:

**МУТАНОВ Ғалымқайыр Мұтанұлы**, техника ғылымдарының докторы, профессор, ҚР ҰҒА академигі, ҚР БҒМ ҚҰО ақпараттық және есептеу технологиялар институтының бас директорының м.а. (Алматы, Қазақстан), **Н=5**

## РЕДАКЦИЯ АЛҚАСЫ:

**КАЛИМОЛДАЕВ Мақсат Нұрәділұлы** (бас редактордың орынбасары), физика-математика ғылымдарының докторы, профессор, ҚР ҰҒА академигі, ҚР БҒМ ҚҰО ақпараттық және есептеу технологиялар институты бас директорының кеңесшісі, зертхана меңгерушісі (Алматы, Қазақстан), **Н=7**

**МАМЫРБАЕВ Өркен Жұмажанұлы** (ғалым хатшы), Ақпараттық жүйелер саласындағы техника ғылымдарының (PhD) докторы, ҚР БҒМ ҚҰО ақпараттық және есептеу технологиялар институты директорының ғылым жөніндегі орынбасары (Алматы, Қазақстан), **Н=5**

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

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

**СМОЛАРЖ Анджей**, Люблин политехникалық университетінің электроника факультетінің доценті (Люблин, Польша), **Н=17**

**ӘМІРҒАЛИЕВ Еділхан Несіпханұлы**, техника ғылымдарының докторы, профессор, ҚР ҰҒА академигі, Жасанды интеллект және робототехника зертханасының меңгерушісі (Алматы, Қазақстан), **Н=12**

**КИЛАН Әлімхан**, техника ғылымдарының докторы, профессор (ғылым докторы (Жапония), ҚР БҒМ ҚҰО ақпараттық және есептеу технологиялар институтының бас ғылыми қызметкері (Алматы, Қазақстан), **Н=6**

**ХАЙРОВА Нина**, техника ғылымдарының докторы, профессор, ҚР БҒМ ҚҰО ақпараттық және есептеу технологиялар институтының бас ғылыми қызметкері (Алматы, Қазақстан), **Н=4**

**ОТМАН Мохаммед**, PhD, Информатика, коммуникациялық технологиялар және желілер кафедрасының профессоры, Путра университеті (Селангор, Малайзия), **Н=23**

**НЫСАНБАЕВА Сауле Еркебұланқызы**, техника ғылымдарының докторы, доцент, ҚР БҒМ ҚҰО ақпараттық және есептеу технологиялар институтының аға ғылыми қызметкері (Алматы, Қазақстан), **Н=3**

**БИЯШЕВ Рустам Гакашевич**, техника ғылымдарының докторы, профессор, Информатика және басқару мәселелері институты директорының орынбасары, Ақпараттық қауіпсіздік зертханасының меңгерушісі (Қазақстан), **Н=3**

**КАПАЛОВА Нұрсұлу Алдажарқызы**, техника ғылымдарының кандидаты, ҚР БҒМ ҚҰО ақпараттық және есептеу технологиялар институтының киберқауіпсіздік зертханасының меңгерушісі (Алматы, Қазақстан), **Н=3**

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

**МИХАЛЕВИЧ Александр Александрович**, техника ғылымдарының докторы, профессор, Беларусь Ұлттық Ғылым академиясының академигі (Минск, Беларусь), **Н=2**

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

«ҚР ҰҒА Хабарлары. Физика-математикалық 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-математикалық 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/>*

© Қазақстан Республикасының Ұлттық ғылым академиясы, 2022  
Типографияның мекен-жайы: «Аруна» ЖК, Алматы қ., Мұратбаев көш., 75.

## Главный редактор:

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

## Редакционная коллегия:

**КАЛИМОЛДАЕВ Максат Нурадилович**, (заместитель главного редактора), доктор физико-математических наук, профессор, академик НАН РК, советник генерального директора «Института информационных и вычислительных технологий» КН МНВО РК, заведующий лабораторией (Алматы, Казахстан), **Н=7**

**МАМЫРБАЕВ Оркен Жумажанович**, (ученый секретарь), доктор философии (PhD) по специальности «Информационные системы», заместитель директора по науке РГП «Институт информационных и вычислительных технологий» Комитета науки МНВО РК (Алматы, Казахстан), **Н=5**

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

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

**СМОЛАРЖ Анджей**, доцент факультета электроники Люблинского политехнического университета (Люблин, Польша), **Н=17**

**АМИРГАЛИЕВ Едилхан Несипханович**, доктор технических наук, профессор, академик Национальной инженерной академии РК, заведующий лабораторией «Искусственного интеллекта и робототехники» (Алматы, Казахстан), **Н=12**

**КЕЙЛАН Алимхан**, доктор технических наук, профессор (Doctor of science (Japan)), главный научный сотрудник РГП «Института информационных и вычислительных технологий» КН МНВО РК (Алматы, Казахстан), **Н=6**

**ХАЙРОВА Нина**, доктор технических наук, профессор, главный научный сотрудник РГП «Института информационных и вычислительных технологий» КН МНВО РК (Алматы, Казахстан), **Н=4**

**ОТМАН Мохамед**, доктор философии, профессор компьютерных наук, Департамент коммуникационных технологий и сетей, Университет Путра Малайзия (Селангор, Малайзия), **Н=23**

**НЫСАНБАЕВА Сауле Еркебулановна**, доктор технических наук, доцент, старший научный сотрудник РГП «Института информационных и вычислительных технологий» КН МНВО РК (Алматы, Казахстан), **Н=3**

**БИЯШЕВ Рустам Гакашевич**, доктор технических наук, профессор, заместитель директора Института проблем информатики и управления, заведующий лабораторией информационной безопасности (Казахстан), **Н=3**

**КАПАЛОВА Нурсулу Алдажаровна**, кандидат технических наук, заведующий лабораторией кибербезопасности РГП «Института информационных и вычислительных технологий» КН МНВО РК (Алматы, Казахстан), **Н=3**

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

**МИХАЛЕВИЧ Александр Александрович**, доктор технических наук, профессор, академик НАН Беларуси (Минск, Беларусь), **Н=2**

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

«Известия НАН РК. Серия физико-математическая».

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-математическая.*

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/>*

© Национальная академия наук Республики Казахстан, 2022  
Адрес типографии: ИП «Аруна», г. Алматы, ул. Муратбаева, 75.

### Chief Editor:

**MUTANOV Galimkair Mutanovich**, doctor of technical sciences, professor, academician of NAS RK, acting General Director of the Institute of Information and Computing Technologies CS MES RK (Almaty, Kazakhstan), **H=5**

### EDITORIAL BOARD:

**KALIMOLDAYEV Maksat Nuradilovich**, (Deputy Editor-in-Chief), Doctor of Physical and Mathematical Sciences, Professor, Academician of NAS RK, Advisor to the General Director of the Institute of Information and Computing Technologies of the CS MES RK, Head of the Laboratory (Almaty, Kazakhstan), **H = 7**

**Mamyrbayev Orken Zhumazhanovich**, (Academic Secretary), PhD in Information Systems, Deputy Director for Science of the Institute of Information and Computing Technologies CS MES RK (Almaty, Kazakhstan), **H = 5**

**BAIGUNCHEKOV Zhumadil Zhanabaevich**, 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**

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

**SMOLARJ Andrej**, Associate Professor Faculty of Electronics, Lublin polytechnic university (Lublin, Poland), **H= 17**

**AMIRGALIEV Edilkhan Nesipkhanovich**, Doctor of Technical Sciences, Professor, Academician of NAS RK, Head of the Laboratory of Artificial Intelligence and Robotics (Almaty, Kazakhstan), **H= 12**

**KEILAN Alimkhan**, Doctor of Technical Sciences, Professor (Doctor of science (Japan)), chief researcher of Institute of Information and Computational Technologies CS MES RK (Almaty, Kazakhstan), **H= 6**

**KHAIROVA Nina**, Doctor of Technical Sciences, Professor, Chief Researcher of the Institute of Information and Computational Technologies CS MES RK (Almaty, Kazakhstan), **H= 4**

**OTMAN Mohamed**, PhD, Professor of Computer Science Department of Communication Technology and Networks, Putra University Malaysia (Selangor, Malaysia), **H= 23**

**NYSANBAYEVA Saule Yerkebulanovna**, Doctor of Technical Sciences, Associate Professor, Senior Researcher of the Institute of Information and Computing Technologies CS MES RK (Almaty, Kazakhstan), **H= 3**

**BIYASHEV Rustam Gakashevich**, doctor of technical sciences, professor, Deputy Director of the Institute for Informatics and Management Problems, Head of the Information Security Laboratory (Kazakhstan), **H= 3**

**KAPALOVA Nursulu Aldazharovna**, Candidate of Technical Sciences, Head of the Laboratory cyber-security, Institute of Information and Computing Technologies CS MES RK (Almaty, Kazakhstan), **H=3**

**KOVALYOV Alexander Mikhailovich**, Doctor of Physical and Mathematical Sciences, Academician of the National Academy of Sciences of Ukraine, Institute of Applied Mathematics and Mechanics (Donetsk, Ukraine), **H=5**

**MIKHALEVICH Alexander Alexandrovich**, Doctor of Technical Sciences, Professor, Academician of the National Academy of Sciences of Belarus (Minsk, Belarus), **H=2**

**TIGHINEANU Ion Mihailovich**, Doctor of Physical and Mathematical Sciences, Academician, President of the Academy of Sciences of Moldova, Technical University of Moldova (Chisinau, Moldova), **H=42**

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

#### Physico-mathematical series.

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-Ж, issued 14.02.2018

Thematic scope: *physical-mathematical series*.

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/>*

© National Academy of Sciences of the Republic of Kazakhstan, 2022

Address of printing house: ST «Aruna», 75, Muratbayev str, Almaty.

NEWS OF THE NATIONAL ACADEMY OF SCIENCES  
OF THE REPUBLIC OF KAZAKHSTAN  
PHYSICO-MATHEMATICAL SERIES  
ISSN 1991-346X

Volume 4, Number 344 (2022), 5-15

<https://doi.org/10.32014/2022.2518-1726.152>

УДК 28.23.29

МРНТИ 28.17.2

**А.С. Баймаханова<sup>1</sup>, А.Ж. Сейтмуратов<sup>2\*</sup>**

<sup>1</sup>Қожа Ахмет Ясауи атындағы Халықаралық қазақ-түрік университеті,  
Түркістан, Қазақстан;

<sup>2</sup>Қорқыт Ата атындағы Қызылорда университеті, Қызылорда, Қазақстан.  
E-mail: *angisin\_@mail.ru*

### **DEEP LEARNING АЛГОРИТМІН ҚОЛДАНУ НЕГІЗІНДЕ ЦИФРЛЫҚ ҚҰЖАТТАРДЫ ЖІКТЕУ**

**Аннотация.** Цифрлық құжаттарды сақтауда көптеген ұйымдар үшін құжаттарды жіктеу маңызды процесс болып табылады. Зерттеу барысында Deep Learning алгоритмдерін қолдануды дамыту технологиясы жасалады. Құжаттар алдымен визуалды түрде, онан соң, бірінші кезеңде жіктелмейтін құжаттар табиғи тілде өңдеу (NLP) арқылы жіктеледі. Оның ішінде топтар саны 20 шамасында, ал құжаттар саны 10 000-дай болады. Зерттеулер бір беттік құжаттан бастап жүргізіледі. Бірінші кезеңде құжаттар көрнекі түрде жіктеледі. Екінші кезеңде жіктелмейтін құжаттарды табиғи тілді өңдеу (NLP) арқылы жіктеуге тырысты. Жаттығулар жинағы шамамен 20 классификация арқылы жасалды құжаттар қолмен. Зерттеулер бір бетте жүргізілді құжаттар. Орындаулар жасау арқылы AlexNet көмегімен құжат толық өлшенеді,  $\frac{1}{2}$  және  $\frac{1}{3}$  бөлігінде тәжірибелер кескін классификациясы үшін, сөздер пакеті және LSTM алгоритмі үшін NLP алынады. Зерттеудің көрсеткіштері сәйкес өлшенді әртүрлі опцияларға жіктеледі. Жетістік деңгейі 75% және 85% болғанда, тиісінше бірінші және екінші кезеңдердің аяқталуы қажет. AlexNet архитектурасы өзгертілгенде жіктеуге келмейтін құжаттардағы кейбір мәтіндерді қолмен енгізуімізге тура келеді. Зерттеу жұмысында алынған нәтижелер, CNN-де құжат тобы белгілі бір ықтималдықпен анықталады. Басқа сөздер менайтқанда, ықтималдықтың белгілі бір мәнін қанағаттандыратын құжаттар тобы қажет. Тақырыптың өзектілігі цифрлық мұрағаттарда көптеген құжаттарды сақтау ұйымдары үшін құжаттарды жіктеу маңызды процесс болып табылады. Жаңартылған AlexNet көмегімен құжат кескінінің

классификациясы AlexNet архитектурасында сегіз зерттеуден қабаттардан тұрады, бес конволюционды және оған үш толық қосылады. Сонымен қатар жұмыста бастапқы AlexNet түпнұсқасының бірінші кіру деңгейін өзгертіп, ең жақсы көрсеткен нәтижелерді алуғадың мүмкін екенін көрсеттік .

**Түйін сөздер:** Deep Learning, AlexNet, NLP, CNN, цифрлық құжаттар.

**А.С. Баймаханова<sup>1</sup>, А.Ж. Сейтмуратов<sup>2\*</sup>**

<sup>1</sup>Международный казахско-турецкий университет имени Ходжи Ахмеда Ясави, Туркистан, Казахстан;

<sup>2</sup>Кызылординский университет имени Коркыт Ата,  
Кызылорда, Казахстан.  
E-mail: *angisin\_@mail.ru*

## **КЛАССИФИКАЦИЯ ЦИФРОВЫХ ДОКУМЕНТОВ НА ОСНОВЕ ИСПОЛЬЗОВАНИЯ АЛГОРИТМА DEEP LEARNING**

**Аннотация.** Классификация документов является важным процессом для многих организаций при хранении цифровых документов. В ходе исследования создается технология отработки использования алгоритмов глубокого обучения. Документы сначала классифицируются визуально, затем документы, не классифицированные на первом этапе, классифицируются с использованием обработки естественного языка (NLP). Среди них количество групп около 20, а количество документов около 10 000. Исследования ведутся, начиная с одностраничного документа. На первом этапе документы классифицируются визуально, на втором этапе неклассифицируемые документы пытались классифицировать с помощью обработки естественного языка (NLP). Комплект упражнений был создан с использованием около 20 классификационных документов вручную. Исследования проводились на одностраничных документах. Выполняя исполнения, документ полностью измеряется с помощью AlexNet, в 1/2 и 1/3 экспериментов получается NLP для классификации изображений, пакета слов и алгоритма LSTM. Показатели исследования классифицируются по разным вариантам, которые соответственно измеряются. При успешности 75% и 85% должны быть выполнены первый и второй этапы соответственно. При изменении архитектуры AlexNet нам приходится вручную вводить некоторый текст в документы, которые невозможно классифицировать. По результатам, полученным в исследовательской работе, в CNN группа документов определяется с определенной вероятностью. Другими словами, нам нужна группа документов, удовлетворяющих определенному значению вероятности. Актуальность темы в цифровых архивах классификация документов является важным процессом для многих организаций по хранению документов. Классификация

изображений документов с обновленной AlexNet состоит из восьми исследовательских слоев в архитектуре AlexNet, пяти сверточных и трех полностью аддитивных. Кроме того, в работе мы показали, что можно изменить первый уровень доступа оригинального AlexNet и получить наилучшие результаты.

**Ключевые слова:** глубокое обучение, AlexNet, NLP, CNN, цифровые документы.

**A. Baimakhanova<sup>1</sup>, A. Seitmuratov<sup>2\*</sup>**

<sup>1</sup>Khoja Akhmet Yassawi International Kazakh-Turkish University,  
Turkistan, Kazakhstan;

<sup>2</sup>Korkyt Ata Kyzylorda State University, Kyzylorda, Kazakhstan.  
E-mail: *angisin\_@mail.ru*

## **CLASSIFICATION OF DIGITAL DOCUMENTS USING DEEP LEARNING ALGORITHM**

**Abstract.** Document classification is an important process for many organizations in digital document storage. In the course of the research, the technology of developing the use of Deep Learning algorithms is created. The documents are first classified visually, then the documents that are not classified in the first stage are classified using natural language processing (NLP). Among them, the number of groups is about 20, and the number of documents is about 10,000. Research is conducted starting from a one-page document. In the first stage, documents are visually classified. In the second stage, unclassifiable documents were tried to be classified using natural language processing (NLP). A set of exercises was created using about 20 classification documents manually. Studies were conducted on one page documents. By making the executions, the document is fully measured using AlexNet, in  $\frac{1}{2}$  and  $\frac{1}{3}$  of the experiments, NLP is obtained for image classification, word pack and LSTM algorithm. The indicators of the study are classified into different options, which are measured accordingly. With a success rate of 75% and 85%, the first and second stages should be completed respectively. When changing the architecture of AlexNet, we have to manually enter some text into documents that cannot be classified. According to the results obtained in the research work, in CNN a group of documents is determined with a certain probability. In other words, we need a group of documents that satisfy a certain probability value. Relevance of the topic In digital archives, document classification is an important process for many document preservation organizations. The updated AlexNet document image classification consists of eight exploratory layers in the AlexNet architecture, five convolutional layers, and three fully additive layers. In addition, in the work we have shown that it is possible to change the first access level of the original AlexNet and get the best results.

**Key words:** Deep Learning, AlexNet, NLP, CNN, digital documents.

**Кіріспе.** Қазақстан Республикасы Үкіметінің 2017 жылғы 12 желтоқсанда №827 Қаулысымен «Цифрлық Қазақстан» мемлекеттік бағдарламасы бекітілді. Жолдауа еліміздің ұзақ мерзімді мақсаттарды және аз уақыт ішіндешілуі қажет міндеттер нақты анықталған. Сол міндеттердің ішіндегі елімізде IT-саласындамыту мәселесі туралы: «Біз цифрлық технологияларға негізделген болашағызор бағыттарға айрықша мән беретін боламыз» - деп, IT-саласын дамытуды ерекше бақылауда ұстауымыз қажеттілігін атап айтты. («Цифрлық Қазақстан»-2019). Атап айтқанда, медицина қаржы, білім және т.б салаларда кеңінен қолданыс тауып, халықтың әлеуметтік жағдайын жақсартуға септігін тигізуде. Цифрландыру процесін жүзеге асыру еліміздің бәсекеге қабілетті елдер қатарына енуіне мүмкіндік береді. Бұл процесті жүзеге асыру мақсатында бес бағыт бойынша жұмыс жасау қабылданды: экономикалық салаларды сандық түрге келтіру, сандық мемлекетке көшу, сандық жібек жолын іске асыру, адами капиталды дамыту, инновациялық экосистемелерді құру (Гурянова, 2017).

Бағыттардағы цифрлық деректердің үлкен көлемі және олардың құрылымданбауы ақпараттық кедергілер мен мәселелер тудырады. Сондықтан да осы мәселелерді шешуде жаңа шешімдерді, жаңа тәсілдерді, сонымен қатар ақпараттарды өңдеуші жаңа алгоритмдер қалыптастыру қажет.

Құжаттарды жіктеу кеңседе маңызды рөл атқаратын жұмыс, олар алдынала анықталған құжаттарды жіктеуге арналған және цифрлық мұрағатта сақтайды.

Кескіндерді жіктеу Deep Learning алгоритмдері арқылы жасалады және оның тиімділігі AlexNet-пен салыстырылады. Нәтижесінде NLP жіктелуіндегі сөз қоры мен LSTM алгоритмі әзірленеді.

Тақырыптың өзектілігі цифрлық мұрағаттарда көптеген құжаттарды сақтау ұйымдары үшін құжаттарды жіктеу маңызды процесс болып табылады. Зерттеу барысында Deep Learning алгоритмдерін қолдануды дамыту технологиясы жасалады. Құжаттар алдымен визуалды түрде, онан соң, бірінші кезеңде жіктелмейтін құжаттар табиғи тілде өңделінеді (NLP) арқылы жіктеледі. Оның ішінде топтар саны 20 шамасында, ал құжаттар саны 10 000-дай болады. Зерттеулер бір беттік құжаттан бастап жүргізіледі. Кескіндерді жіктеу Deep Learning алгоритмдері арқылы жасалады және оның тиімділігі AlexNet-пен салыстырылады. Нәтижесінде NLP жіктелуіндегі сөз қоры мен LSTM алгоритмі әзірленеді.

Классификация құжаттардың қазіргі таңда маңызды жұмыс атқарады. Құжаттарды алдын ала анықталғандарға жіктеуге бағытталған топтарды және оларды сәйкесінше цифрлық мұрағатта сақтау қажет. Жіктеуге бағытталған құжаттарды алдын-ала жіктелген топтарда сақталады. Бұл жұмыста енгізілген зерттеу 20 топқа бөлініп, соның ішінде 10 000-ға жуық құжат құрастырылады. Берілген құжаттар цифрлық мұрағатта сақталады және қажет болған жағдайда мұрағаттан алынып тасталады. Біздің жұмысымыз үшін әзірленген әдістер мен алгоритмдер құжаттарымызды кескін бойынша



жіктеленіледі. Терең оқыту, яғни терең құрылымдық оқыту – бұл жасанды нейрондық желілерге негізделген машиналық оқыту әдістерінің кең тобының бөлігі болып табылады. Оқытуды бақылауға, ішінара бақылауға немесе бақылауға болмайды (Bengio т.б., 2013).

Дереккөздерне сүйенсек, Фрэнк Розенблатт заманауи терең оқыту жүйелерінің барлық негізгі компоненттерін жасап, зерттегенін көрсетеді (Тапперт т.б., 2019). Интеллектуалды ақпараттық жүйелер тұрғысынан қарағанда бейне – бұл деректерді талдау жасаушы деректер жиынтығынан ерекшеліп алуға және қарастырылып отырған есептің шарттарымен сәйкес басқа объектілер мен топтауға мүмкіндік беретін нақты немесе абстрактілі объектілердегі (процестердегі, құбылыстардағы) деректер жиынтығы. Бастапқыда алгоритм мен әдіс таңдалынады аймағымызға сәйкес қолдану барысында оң нәтижеге қол жеткізіледі. Алдымен олар құжаттардың ерекшеліктерін анықтауға тырысты, содан кейін құжат топтарда зерттелді. Екіншіден, қолдануға болатын алгоритмдер мен әдістер құжатты сурет ретінде жіктеу зерттелді. Бұл зерттеудегі құжаттар, әдетте, мәтіннен жәнестелермен терілген бір беттен тұрады. Құжаттарымызды анықтау, талдау жасау, құжаттарды жіктеу үшін әртүрлі алгоритмдер мен әдістер қолданылды. Нәтижелеріміз AlexNet алгоритмі арқылы алынған.

Қол жеткізілген жетістік AlexNet алгоритмінің көмегімен шамамен 75% құрайды. Табиғи тілөндеу әдістері кейінірек топқа кірмеген адамдар үшін сыналды, AlexNet алгоритмін қолдана отырып анықтай алдық. Бұл тұрғыда алгоритм «Сөздік қорымыз» өте сәтті болды. Осы екі деңгейлі зерттеудің соңында жалпы өнімділік шамамен 95% көрсеткішке жеткізу болды. Басқалары жіктеуге қарапайым, бір беттік мәтіндер мәтіндерді жіктеуге негізделген. Терең конвульсиялық нейрондық желілер CNN кескінді классификациялауда кеңінен қолданылуда. AlexNet моделі кескіндерді классификациялауда терең оқытуда қолданыла бастады. Конвульсиялық нейрондық желілер CNN параметрлердің санын анықтайтын және конверсия мәселелерін тиімді шешеді. Терең оқытуға негізделіп отырған модельдер мәтіндерімізді зерттеу барысында көптеген машиналық оқыту әдістерден басым болуда.

Мәтіндерімізді, бейнемізді жіктеуге байланысты негізгі мәселе-оқытудың ең қолайлы түрін анықтау өте маңызды. Осы оқу бастамаларының көпшілігі өнімділікті арттыруға, қателерді азайтуға және қателіктердің алдын алуға көмектеседі. Алайда жұмысты жіктеудің негізгі түрлері мен әдістерін бөліп көрсету арқылы тану жұмысымыз жүргізеді (Mohammed т.б., 2021). Қазіргі таңда қауіпті шабуылдардың басым бөлігі терең білім негізінде яғни, кескін классификаторларын алдауға бағытталуда. Суреттерді классификациялау барысында шабуылдарды қауіпсіздікке сезімтал ақпараттық жүйеге қарсы қандай шабуылдар жасалуы мүмкін деген болжамдарды анықтайтын әртүрлі қауіп-қатер модельдеріне жатқызуға болады (Machado т.б.-2021:38).

Классификациялаудың тағы бір кең таралған түрі ол конвульсиялық нейрондық желі моделі болып табылады (Minaee т.б. 2019). Соңғы жылдары

CNN моделі компьютерлік көру және табиғи тілді өңдеу міндеттерінің кең ауқымды тиімдісі болды. Ұсынылған құрылымды бағалау барысында негізгі модельдерді оқыту үшін қолданылатын сәулет түрі LSTM және CNN архитектурасының әлсіз байланысы LSTM-CNN деп те аталады, ол LSTM және CNN екеуін де біріктіреді (Girshick т.б., 2014).

Материалдар мен әдістер. Технологияның үздіксіз даму барысында терең оқыту әдісі медициналық кескіндерді жіктеуде дәстүрлі машиналық оқыту мәселелерін шешудің жаңа түрін ұсынатын медициналық кескіндерді талдаудың тандаулы әдісіне айналған. Терең оқыту әдісі компьютерлік көру, содан кейін медициналық кескіндерді көруде кеңінен қолданылады. Терең оқытуда және оның құрылымында—оқытушысыз оқыту, бақыланатын оқыту және т.б. Жіктеудің ешқандай нұсқауынсыз CNN-дің дәстүрлі моделін үйрету қиын. Сондықтан, тақырып саласы білімге және тиісті сарапшылардың тәжірибесіне сүйенеді, CNN-ге негізделген қолданыстағы әдістердің көпшілігі MRI диагностикалық моделін құру үшін ақпараттық аймақты (мысалы, гиппокамп немесе учаске) эмпирикалық түрде анықтайды, бұл ми ауруларында диагностикалауда кезінде терең нейрондық желінің тиімділігін төмендетуі мүмкін. Дәстүрлі әдістермен салыстырғанда, терең оқыту модельдері жіктеу үшін суреттердің ерекшеліктерін автоматты түрде зерттеуге болады, диагностикалық өнімділік артады. Терең оқыту технологиясы жарнаманы дәл анықтау үшін қосымшаларды қолдануды қажет етеді.

Харли және т.б. (Harley т.б., 2015) берілген құжаттардың кескіні үшін терең оқыту мүмкіндіктерді қолдана отырып, құжаттардың кескіндерін жіктеу және шығару үшін CNN қолданылады. CNN функциялары сығылуға төзімді әрі құжаттардағы суреттерді оқиды. CNN жақсы тасымалданады әрі құжаттарды талдау тапсырмаларын орындау барысында орындауда белгілі бір мүмкіндікті оқытуға жеткілікті. CNN көп суретті өндеудің басқа әдістеріне қарағанда сәтті көрсеткіштерге ие, CNN байланысты емес суреттері қолданылған құжаттарға қатысты тапсырмаларды өте жақсы орындайды. Афзал және т.б. (Afzal т.б., 2015) CNN-ге негізделген құжаттарды жіктеуді зерттеуде дайындалған мәліметтердің өте көп мөлшері қажет деді.

Зерттеулерде тек жіктеу дәлдігіне қол жеткізу үшін бір сыныпқа 20 топтық үлгісі көрсетіледі. Олар үйренген AlexNet Imagenet суреттері олардың жұмысында қолданады. ImageNet-бұл мәліметтер жиынтығы 500-ден астам суреттер қамтылады, дайындалған әр түйінде және жалпы суреттер қарастырылады. AlexNet-бұл құрамында суреттер 1000 санат. ImageNet құрамында әр түрлі бұл зерттеу құжаттардың 10 түрін зерттеді. Құжаттардың бұл түрлері: дипломдық жұмыстар, бұйрықтар, транзакциялар, шығыс, кіріс журналдары, хаттамалар, ведомства, аттестаттар т.б. көрсетілген. CNN-нің салалардағы зерттеулер басқа зерттеулеріден өзгеше болып табылады. Deep Learning негізгі цифрлық құралдардың көмегімен құжаттарды ретпен басқаруды қамтамасыз етеді. CNN әдісі бейнені тану барысында құжат кескініндегі ең сәтті процесті жіктеу әдісі болып табылады.

Кадам бойынша орындалатын кезеңдер:

-Құжаттарды классификациялау жоспарланған, алдын-ала құрастырылған топтарды қарастырады. Жинақталған бейнелерді жіктейміз. Олар латын немесе кириллица тілінде болуы мүмкін.

- Терең оқыту әдістерін қолданатын қадам үшін AlexNet-ті таңдаймыз, себебі ол суретті тану немесе жіктеу үшін өте тиімді. Егер, біз мәтінді жіктей алмасақ немесе терең оқыту алгоритмдерін қолдана алмасақ, біз NLP-ді қолдана аламыз.

-10 000 құжатты қолмен жіктей отырып жинау қажет, бірақ алдымен топ сөзін таңдаудан бұрын топтың мәтіндік құжаттарын шешеміз. AlexNet RGB кескіндерінде оқытылады және шығарылады.

Үлкен және терең CNN классификациялауда Imagenet-те 10000-ға жуық жоғары ажыратылған суреттерді 20 түрлі топтарға бөлу үшін жаңа әдіс қолданылады. Жоғары оңтайландырылған графикалық суреттер іске асыруды жазуда іске асырылады (Krizhevsky т.б., 2017).

Терең оқыту CNN әдісін процестердің ең сәтті әдісі құжаттардың кескіндерін жіктеу барысында орындалады. Олар бірнеше факторларды қарастырды. Құжаттарды өңдеу кезінде CNN өнімділігі аса зор әсер етеді. Олароқу және оқу кезінде ауысымдық түрлендірулерді қолдануда үлкен кіріс кескіндерін пайдалануда үлкен өсуге әкелетін өнімділік болып табылған (Tensmeyer т.б., 2017). Классификация тапсырмасының қолданыстағы алгоритмдері стратегиялары құжаттар кескіні қатені екі есе азайтуға мүмкіндік береді (Afzal т.б., 2017). Санның әсерін талдауоқу құжатының суреттері және жіктеу мүмкіндіктеріне арналған басқа параметрлер: дипломдық жұмыстар, бұйрықтар, транзакциялар, шығыс, кіріс журналдары, хаттамалар, ведомства, аттестаттар т.б. бойынша жұмыс істеді. AlexNet-те жұмыс істеу барысында көптеген мәселелер туындауы мүмкін.

Ал, жаттығу деректерін Alex Net-ке Python бағдарламасында қолдануға болады. AlexNet-ті оқытқаннан кейін оны ұзақ жіктеу құжаттарын қолдану үшін пайдалануға болады. Alex Net қолданбасын қолдану үшін бір ғана үлгі алынады және бұл құжаттың тобы–міндетті түрде белгілі болады. Егер, AlexNet құжаттар тобын белгілесек болғаны, жіктей алмаған жағдай қарастырылады. Біз OCR Optical Character Recognition арқылы мәтінге түрлендіруге кірістік. Сөз таптарын тауып қолмен енгізуге болады. Талдау жүргізу барысында суреттер де, мәтін де құжаттар және нәтиже алуда екі талдаудың нәтижелерін біріктіре отырып құжаттар кескіндерінің классификациялау, бейнелеу әдісімен мәтінді тану нәтижелерін талдау барысында тиімдірек болады (Jain т.б., 2019)

Құжаттарды жіктеу мәселелеріне өндіріс ортасы мен құжаттарды жіктеудің қолданыстағы әдістерінде назар аудару қажет. Екі кезең ұсынылады: бірінші кезеңде, тереңдетілген желілік оқыту. Бұл мүмкіндіктерді алу тәсілі ретінде жұмыс істейді. Екінші кезеңде, тәуелсіз білім беру жүйесі, яғни CNN және LM нәтижелерін салыстыра отырып, қорытынды шығарамыз. Әлемге негізделген

құжаттарды құру кезінде оқудың жаңа перспективалары нақты уақыттағы ауқымды қосымшаларға сәйкес жіктелуде (Kolsch т.б., 2017). Жаңа жіктеу әдісін күрделі геометриялық құжаттарға ұсынамыз. Олардың көзқарасы негізделген нысандарын жылдам және сенімді анықтау алгоритмдеріне тығыз байланысты.

Ұсынылған әдістің артықшылықтары:

- жоғары жылдамдық;
- автоматты түрде құру мүмкіндігі;
- өңделмеген бейнелерді алдын ала өңдеудің кез-келген кезеңдерінде өңдеу.

Жұмысымызда ұсынылған алгоритмдер құжаттарды классификациялауға ғана емес, сонымен бірге құжаттарды анықтауға да мүмкіндік тудырады және суреттегі құжаттардың орналасуын және бағытын көрсетеді (Usilin т.б., 2010).

CNN мәселелері көптеген шығарылымдарға және ең тиімді параметрлерге ие болады. Олардың ұсынысы бойынша жақсартылған AlexNet негізінде суреттерді сұрыптау әдісі жанартылады. Бұл әдіс AlexNet қабаттарын толықтырады, толық қабатқа сілтемелер бойынша сұрыпталады.

Нәтижелер. Нәтижесінде ұсынылған әдісте параметрлер саны және арақатынас параметрлер пропорциясына толық байланыс деңгейі де жоғарылады сонымен қатар, AlexNet-пен салыстырғанда жіктеу дәлдігін арттырады (Li т.б., 2020).

Жалпы жұмыстарды бағаланған кезде, CNN құжатты терең жинақтауда, кескіндеуде, өңдеудегі ең табысты әдіс болып табылады. CNN-дегі өзгерістер және осы топтамадағы AlexNet-ті қарастырсақ.

Зерттеуіміздің бірінші кезеңінде, Alexnet көмегімен құжаттарды жіктеуге тырыстық және өнімділігін өлшедік. Өзгертулер енгізу арқылы табысты нәтижелерге қол жеткізу үшін AlexNet-ті толық мүмкіндігін қолдандық. Келесі қадамымыз құжаттарды жіктеуге тырыстық, мәтіндік классификатор арқылы суреттерді өңдеу арқылы жіктеуге болатын және жіктеуге болмайтын әдістерді қарастырдық. Құжаттардың суреттерін AlexNet көмегімен жіктеу, суреттерді жіктеу әдісімен құжаттарды жіктеу үшін 10000 құжаттан тұратын оқыту мәліметтер жиынтығы дайындалды (Сурет 1). Құжаттардың 20-ға жуық топтары ажыратылады.



Сурет 1 – Alexnet көмегімен құжаттарды жіктеу

Университеттің архив бөліміндегі құжаттар келесі түрлерде жіктеледі:

- 1) оқудан шығарылған немесе бітірушілердің жеке құжаттары;
- 2) университет қызметкерлерінің жұмыстан босатылғаннан кейін жеке құжаттары;
- 3) дипломдық жұмыстар (жобалар);
- 4) студенттер құрамына қабылдау бұйрықтары;
- 5) жеке құжаттарды қабылдау туралы актілер;
- 6) есепке алу құжаттары;
- 7) жалпы іс жүргізу: кіріс және шығыс құжаттары;
- 8) СМК жазбалары;
- 9) кадрлық есепке алуды жүргізу;
- 10) мемлекеттік органдардан студенттің университетте оқу фактісін растау туралы сұраныстар;
- 11) білім алушылар ауысқан жағдайда басқа білім беру ұйымдарының сұраныстары.

12) іс жүргізуде толтырылған басқа да құжаттар т.б. бойынша жұмыс істелінді. AlexNet-тің сәттілігін бағалау үшін 10000 құжат тексерілді

Сканерлеу өнімділігін өлшеу үшін әртүрлі өлшемдер сыналды. Мысалы, 75x75 пиксель және 200x200 пиксель, ақ-қара және 256 сұр деңгейдегі. Сканерлеу кезінде 75x75 пиксель ажыратымында кара-ақ форматта сканерлеу құжаттардың жеткілікті оқылмайтындығы анықталды.

Құжаттың класын анықтау үшін беттің жоғарғы жағын зерттеу жеткілікті болады. Осы себепті толық бет, жоғарғы жартысыбет және жоғарғы 1/3 беттер сканерленді және өнімділіктің әсерін өлшеуге әрекет жасалды.

**Талқылау.** Құжаттарды жіктеу оның кескіндерін жіктеу әдісіне негізделген әдіс осы зерттеу аясында жасалады. Құжаттардың суреттерін жіктеу AlexNet негіздері ретінде пайдаланылды, бірақ өзгертілген. Сондай-ақ, сканерлеу жеткілікті екендігі анықталған. Толық беттің орнына жоғарғы 1/3 бет өлшемі тура келеді. Бұл шешімсканерлеу процесін жеделдету тұрғысынан маңызды. LSTM әдістерін сурет бойынша жіктелуі мүмкін емес құжаттароларды өңдеу әдісін қолмен енгізу әдістермен жіктеуде қолдануға болады. Осы кезеңдердің соңында көптеген топтарға жіктелегенқұжаттардың 99% - ы сақталады, ал қалған 1%-ы қолмен енгізілуі мүмкін. Барлық сканерленген құжаттар қара және ақ түсті болады.

Алайда, мұның бәрін жіктеуге болады, ал сканерлейтін құжаттарда 256 сұр деңгейі бар. AlexNet архитектурасы өзгертілгенде жіктеуге келмейтін құжаттардағы кейбір мәтіндерді қолмен енгізуімізге тура келеді. Зерттеу жұмысында алынған нәтижелер, CNN-деқұжаттобы белгілі бірықтималдықпен анықталады. Басқа сөздермен айтқанда, ықтималдықтың белгілі бір мәнін қанағаттандыратын құжаттар тобы қажет. Жаңартылған AlexNet көмегімен құжат кескінінің классификациясы AlexNet архитектурасында сегіз зерттеуден қабаттардан тұрады, бес конволюционды және оған үш толық қосылады. Біз тек бастапқы AlexNet түпнұсқасының бірінші кіру деңгейін өзгертіп, ең жақсы көрсеткен нәтижелерді алуға мүмкіндік береді.

**Information about authors:**

**Baimakhanova Aigerim** – Doctoral student, Khoja Akhmet Yassawi International Kazakh-Turkish University *aigerimka01@mail.ru*. <https://orcid.org/0000-0002-5364-0146>;

**Seitmuratov Angisin** – Doktor of Physical and Matematical Sciences, Professor, The Korkyt Ata Kyzylorda State University. Kyzylorda. Kazakhstan. *angisin\_@mail.ru*. <https://orcid.org/0000-0002-9622-9584>.

**ӘДЕБИЕТТЕР:**

A.W. Harley, A. Ufkes, and K.G. Derpanis “Evaluation of Deep Convolutional Nets for Document Image Classification and Retrieval,” 13th IAPR International Conference on Document Analysis and Recognition, 2015.

A. Krizhevsky, I. Sutskever, G.E. Hinton, “ImageNet Classification with Deep Convolutional Neural Networks”, Communication of the ACM, Volume 60, Issue 6, June 2017 pp 84-90.

A. Kolsch, M.Z. Afzal, M. Ebbecke, M. Liwicki, “Real-Time Document Image Classification using Deep CNN and Extreme Learning Machines”, 14th IAPR International Conference on Document Analysis and Recognition, 2017.

Bengio Y.; Courville A.; Винсент П. (2013). «Репрезентативное обучение: обзор и новые перспективы». IEEE Transactions по анализу шаблонов и машинному анализу. 35(8): 1798–1828. *arXiv: 1206.5538* . DOI:10.1109 / *tpami*.2013.50. PMID 23787338. S2CID 393948.

Гурянова (Кибитова) В.Н. Ансамбль алгоритмов для определения ишемической болезни сердца //Сборник тезисовXXVI международный научной конференции Ломоносов.– М:Издательский отдел факультета вычислительной математики и кибернетики МГУ имени М.В. Ломоносова,2017.–С.15-17.

Girshick R., Donahue J., Darrell T. & Malik J. (2014). Rich feature hierarchies for accurate object detection and semantic segmentation. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 580-587).

Mohammed A. & Kora R. (2021). An effective ensemble deep learning framework for text classification. Journal of King Saud University-Computer and Information Sciences.

Machado G.R., Silva E. & Goldschmidt R.R. (2021). Adversarial Machine Learning in Image Classification: A Survey Toward the Defender’s Perspective. ACM Computing Surveys (CSUR), 55(1), 1-38.

Minaee S., Azimi E. & Abdolrashidi A. (2019). Deep-sentiment: Sentiment analysis using ensemble of cnn and bi-lstm models. arXiv preprint arXiv:1904.04206.

M.Z. Afzal, S. Capobianco, M.I. Malik, S. Marinai, T.M. Breuel, A. Dengel, and M. Liwicki, “DeepDocClassifier: Document Classification with Deep Convolutional Neural Network”, 13<sup>th</sup> International Conference on Document Analysis and Recognition (ICDAR), 2015.

M.Z. Afzal, A. Kölsch, S. Ahmed, M. Liwicki, “Cutting the Error by Half: Investigation of Very Deep CNN and Advanced Training Strategies for Document Image Classification”, 14th IAPR International Conference on Document Analysis and Recognition, 2017.

R. Jain and C. Wigington, “Multimodal Document Image Classification”, 15<sup>th</sup> IAPR International Conference on Document Analysis and Recognition, 2019.

C. Tensmeyer and T. Martinez, “Analysis of Convolutional Neural Networks for Document Image Classification”, 14th IAPR International Conference on Document Analysis and Recognition, 2017.

S. Usilin, D. Nikolaev, V. Postnikov and G.Schaefer, “Visual Appearance Based Document Image Classification”, Proceedings of IEEE 17th International Conference on Image Processing. 2010.

S. Li, L. Wang, J. Li and Y. Yao, “Image Classification Algorithm Based on Improved AlexNet”, MBIDAS 2020 Journal of Physics: Conference Series, 2020.

Тапперт, Чарльз С. (2019). «Кто отец глубокого обучения?» . 2019 Международная кон-

ференция по вычислительным наукам и вычислительного интеллекта (CSCI). IEEE. С. 343–348. DOI: 10.1109/CSCI49370.2019.00067. ISBN978-1-7281-5584-5. S2CID 216043128. Проверено 31 мая 2021 года.

«Цифрлық Қазақстан» мемлекеттік бағдарламасы // (алынған күні 13.05.2019). [https://prime-minister.kz/kz/page/view/tsifirlik\\_kazakstan\\_memlekettik\\_bagdarlama](https://prime-minister.kz/kz/page/view/tsifirlik_kazakstan_memlekettik_bagdarlama)

#### REFERENCES:

A.W. Harley, A. Ufkes, and K.G. Derpanis “Evaluation of Deep Convolutional Nets for Document Image Classification and Retrieval,” 13th IAPR International Conference on Document Analysis and Recognition, 2015.

A. Krizhevsky, I. Sutskever, G.E. Hinton, “ImageNet Classification with Deep Convolutional Neural Networks”, Communication of the ACM, Volume 60, Issue 6, June 2017 pp 84-90.

A. Kolsch, M.Z. Afzal, M. Ebbecke, M. Liwicki, “Real-Time Document Image Classification using Deep CNN and Extreme Learning Machines”, 14th IAPR International Conference on Document Analysis and Recognition, 2017.

Bengio Y.; Courville A.; Винсент П. (2013). «Репрезентативное обучение: обзор и новые перспективы». IEEE Transactions по анализу шаблонов и машинному.

Guryanova (Kibitova) V.N. An ensemble of algorithms for the definition of ischemic heart disease//Sbornik tezisov XXVI mezhdunarodnyy nauchnoy konferencii Lomonosov.–M: Izdatelskiy otdel fakulteta vychislitelnoy matematiki i cybernetiki MGU imeni M.V. Lomonosova, 2017.–P. 15-17.

Girshick R., Donahue J., Darrell T. & Malik J. (2014). Rich feature hierarchies for accurate object detection and semantic segmentation. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 580-587).

Mohammed A. & Kora R. (2021). An effective ensemble deep learning framework for text classification. Journal of King Saud University-Computer and Information Sciences.

Machado G.R., Silva E. & Goldschmidt R.R. (2021). Adversarial Machine Learning in Image Classification: A Survey Toward the Defender’s Perspective. ACM Computing Surveys (CSUR), 55(1), 1-38.

Minaee S., Azimi E. & Abdolrashidi A. (2019). Deep-sentiment: Sentiment analysis using ensemble of cnn and bi-lstm models. arXiv preprint arXiv:1904.04206.

M.Z. Afzal, S. Capobianco, M.I. Malik, S. Marinai, T.M. Breuel, A. Dengel, and M. Liwicki, “DeepDocClassifier: Document Classification with Deep Convolutional Neural Network”, 13<sup>th</sup> International Conference on Document Analysis and Recognition (ICDAR), 2015.

M.Z. Afzal, A. Kölsch, S. Ahmed, M. Liwicki, “Cutting the Error by Half: Investigation of Very Deep CNN and Advanced Training Strategies for Document Image Classification”, 14th IAPR International Conference on Document Analysis and Recognition, 2017.

R. Jain and C. Wigington, “Multimodal Document Image Classification”, 15<sup>th</sup> IAPR International Conference on Document Analysis and Recognition, 2019.

C. Tensmeyer and T. Martinez, “Analysis of Convolutional Neural Networks for Document Image Classification”, 14th IAPR International Conference on Document Analysis and Recognition, 2017.

S. Usilin, D. Nikolaev, V. Postnikov and G. Schaefer, “Visual Appearance Based Document Image Classification”, Proceedings of IEEE 17th International Conference on Image Processing. 2010.

S. Li, L. Wang, J. Li and Y. Yao, “Image Classification Algorithm Based on Improved AlexNet”, MBDA 2020 Journal of Physics: Conference Series, 2020.

Tappert, Charles S. (2019). “Who is the father of deep learning?”. 2019 International Conference on Computational Sciences and Computational Intelligence (CSCI). IEEE. pp. 343-348. DOI : 10.1109 / CSCI49370.2019.00067 . ISBN 978-1-7281-5584-5. S2CID 216043128 . Verified on May 31 , 2021 .

“Digital Kazakhstan” state program / [https://primeminister.kz/kz/page/view/tsifirlik\\_kazakstan\\_memlekettik\\_bagdarlamasi](https://primeminister.kz/kz/page/view/tsifirlik_kazakstan_memlekettik_bagdarlamasi) (13.05.2019).

## МАЗМҰНЫ

<b>А.С. Баймаханова, А.Ж. Сейтмуратов</b> DEEP LEARNING АЛГОРИТМІН ҚОЛДАНУ НЕГІЗІНДЕ ЦИФРЛЫҚ ҚҰЖАТТАРДЫ ЖІКТЕУ.....	5
<b>М.А. Болатбек, Ш.Ж. Мусиралиева, К. Багитова, А.Т. Нюсупов, Е. Абайұлы</b> ВЕБ-РЕСУРСТАРДАҒЫ ФИШИНГТІК ХАБАРЛАМАЛАР ЖӘНЕ ОЛАРДЫ МАШИНАЛЫҚ ОҚЫТУ ӘДІСТЕРІ АРҚЫЛЫ АНЫҚТАУ.....	16
<b>М.А. Кантуреева, А.Ш. Хасенов, Д.А. Тусупов, А.Б. Закирова, А.З. Алимагамбетова</b> ЭВАКУАЦИЯ ДИНАМИКАСЫНА АРНАЛҒАН FLOOR FIELD МОДЕЛІ...30	30
<b>А.Д. Кубегенова, К.Т. Искаков, Е.С. Кубегенов, О.И. Криворотько</b> ДЕРЕКТЕРДІ ИНТЕЛЕКТУАЛДЫ ТАЛДАУ АРҚЫЛЫ ЭПИДЕМИОЛОГИЯЛЫҚ ЖАҒДАЙДЫ БАҚЫЛАУ ЖӘНЕ МОДЕЛЬДЕУ.....	43
<b>Г. Қалман, М.А. Самбетбаева, Д.А. Ақтаева, А.С. Илюбаев</b> МАШИНАЛЫҚ ОҚЫТУ ӘДІСТЕРІНЕ НЕГІЗДЕЛГЕН АНАФОРАНЫ ШЕШУ МОДЕЛІ.....	56
<b>С.Т. Мамбетов, Е.Е. Бегимбаева, С.К. Джолдасбаев, Б.О. Куламбаев, Г.Н. Казбекова</b> АҚПАРАТТЫҚ ЖҮЙЕНІҢ ҚАУІПТЕРІ МЕН ОСАЛ ТҰСТАРЫНЫҢ МОНИТОРИНГІ ТУРАЛЫ.....	68
<b>У.Т. Махажанова, Б. Тасуов, А.А. Муханова, А. Мухиядин, Р.К. Жеткиншеков</b> БҰЛДЫР ЖИЫНДАР ТЕОРИЯСЫ НЕГІЗІНДЕ БИЗНЕСТІҢ НЕСИЕ ҚАБІЛЕТІЛІГІН БАҒАЛАУ АЛГОРИТМІ.....	81
<b>Р.Н. Молдашева, А.А. Исмаилова, А.К. Жамангара, А.М. Задағали, Г.Б. Турмуханова</b> СУ ЭКО ЖҮЙЕЛЕРІН ЗЕРТТЕУДЕ АТЖ ӨЗІРЛЕУГЕ ҚОЙЫЛАТЫН ТАЛАПТАР.....	93
<b>А.А. Муханова, У.Т. Махажанова, Н.Д. Мархабатов, Б. Тасуов, Ж.Б. Ламашева</b> ЭКОНОМИКАЛЫҚ ЖҮЙЕЛЕРДІ ТАЛДАУДА БҰЛДЫР ЛОГИКАНЫ ҚОЛДАНУ.....	106



<b>Н.А. Сейлова, А.Б. Батыргалиев, Ж.А. Джангозин, Д.А. Байбатчаева, Н. Нұрғабылов</b> ШУ КЕДЕЛДЕРІН БҮРКЕУДІҢ САПАСЫН БАҒАЛАУ ӘДІСТЕМЕСІ.....	120
<b>А.Ш. Хасенов, М.А. Кантурсева, Д.А. Тусупов, А.С. Омарбекова, Г.Б. Абдикеримова</b> АГЕНТТІК МОДЕЛЬДЕУ ЖҮЙЕСІНДЕ ЭВАКУАЦИЯ МОДЕЛІН ЖҮЗЕГЕ АСЫРУ ТӘСІЛІ.....	134
<b>А. Шаушенова, А. Нурпейсова, Д. Досалянов, Г. Мауина</b> ПРОКТОРИНГ ЖҮЙЕСІНДЕ ЖАСАНДЫ НЕЙРОНДЫҚ ЖЕЛІЛЕРГЕ НЕГІЗДЕЛГЕН СӨЙЛЕУДІ ТАҢУ МӘСЕЛЕЛЕРІ.....	146
<b>А.Ә. Шекербек, Г.Б. Абдикеримова, Ә.М. Сабыр, Ж.С. Әбілқайыр</b> КЕУДЕ КЛЕТКАСЫНЫҢ ПАТОЛОГИЯСЫН АНЫҚТАУ ҮШІН ӘДІС ПЕН АЛГОРИТМДІ ҚОЛДАНУ.....	159

## СОДЕРЖАНИЕ

<b>А.С. Баймаханова, А.Ж. Сейтмуратов</b> КЛАССИФИКАЦИЯ ЦИФРОВЫХ ДОКУМЕНТОВ НА ОСНОВЕ ИСПОЛЬЗОВАНИЯ АЛГОРИТМА DEEP LEARNING.....	5
<b>М.А. Болатбек, Ш.Ж. Мусиралиева, К. Багитова, А.Т. Нюсупов, Е. Абайулы</b> ФИШИНГОВЫЕ СООБЩЕНИЯ НА ВЕБ-РЕСУРСАХ И ИХ ОПРЕДЕЛЕНИЕ МЕТОДАМИ МАШИННОГО ОБУЧЕНИЯ.....	16
<b>М.А. Кантуреева, А.Ш. Хасенов, Д.А. Тусупов, А.Б. Закирова, А.З. Алимагамбетова</b> FLOOR FIELD МОДЕЛЬ ДЛЯ ДИНАМИКИ ЭВАКУАЦИИ.....	30
<b>А.Д. Кубегенова, К.Т. Искаков, Е.С. Кубегенов, О.И. Криворотько</b> МОНИТОРИНГ И МОДЕЛИРОВАНИЕ ЭПИДЕМИОЛОГИЧЕСКОЙ СИТУАЦИИ С ПОМОЩЬЮ ИНТЕЛЛЕКТУАЛЬНОГО АНАЛИЗА ДАННЫХ.....	43
<b>Г. Қалман, М.А. Самбетбаева, Д.А. Актаева, А.С. Илюбаев</b> МОДЕЛЬ РАЗРЕШЕНИЯ АНАФОРЫ НА ОСНОВЕ МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ.....	56
<b>С.Т. Мамбетов, Е.Е. Бегимбаева, С.К. Джолдасбаев, Б.О. Куламбаев, Г.Н. Казбекова</b> О МОНИТОРИНГЕ УГРОЗ И УЯЗВИМОСТЕЙ ИНФОРМАЦИОННОЙ СИСТЕМЫ.....	68
<b>У.Т. Махажанова, Б. Тасуов, А.А. Муханова, А. Мухиядин, Р.К. Жеткиншеков</b> АЛГОРИТМ ОЦЕНКИ КРЕДИТОСПОСОБНОСТИ БИЗНЕСА НА ОСНОВЕ ТЕОРИИ НЕЧЕТКИХ МНОЖЕСТВ.....	81
<b>Р.Н. Молдашева, А.А. Исмаилова, А.К. Жамангара, А.М. Задағали, Г.Б. Турмуханова</b> ТРЕБОВАНИЯ К РАЗРАБОТКЕ ИАС-ИССЛЕДОВАНИЙ ВОДНЫХ ЭКОСИСТЕМ.....	93
<b>А.А. Муханова, У.Т. Махажанова, Н.Д. Мархабатов, Б. Тасуов, Ж.Б. Ламашева</b> ПРИМЕНЕНИЕ НЕЧЕТКОЙ ЛОГИКИ ПРИ АНАЛИЗЕ ЭКОНОМИЧЕСКИХ СИСТЕМ.....	106

<b>Н.А. Сейлова, А.Б. Батыргалиев, Ж.А. Джангозин, Д.А. Байбатчаева, Н. Нұрғабылов</b> МЕТОДИКА ОЦЕНКИ КАЧЕСТВА МАСКИРУЮЩИХ ШУМОВЫХ ПОМЕХ.....	120
<b>А.Ш. Хасенов, М.А. Кантуреева, Д.А. Тусупов, А.С. Омарбекова, Г.Б. Абдикеримова</b> ПОДХОД К РЕАЛИЗАЦИИ МОДЕЛИ ЭВАКУАЦИИ В СИСТЕМЕ АГЕНТНОГО МОДЕЛИРОВАНИЯ.....	134
<b>А.Г. Шаушенова, А.А. Нурпейсова, Д.Б. Досалянов, Г.М. Мауина</b> ПРОБЛЕМЫ РАСПОЗНАВАНИЯ РЕЧИ НА ОСНОВЕ ИСКУССТВЕННЫХ НЕЙРОННЫХ СЕТЕЙ В СИСТЕМЕ ПРОКТОРИНГА.....	146
<b>А.А. Шекербек, Г.Б. Абдикеримова, А.М. Сабыр, Ж.С. Абулхаир</b> ПРИМЕНЕНИЕ МЕТОДА И АЛГОРИТМА ДЛЯ ВЫЯВЛЕНИЯ ПАТОЛОГИИ ГРУДНОЙ КЛЕТКИ.....	159

## CONTENTS

<b>A. Baimakhanova, A. Seitmuratov</b> CLASSIFICATION OF DIGITAL DOCUMENTS USING DEEP LEARNING ALGORITHM.....	5
<b>M. Bolatbek, Sh. Musiralieva, K Bagitova, A. Нюсупов, E. Abaiuly</b> PHISHING MESSAGES ON WEB RESOURCES AND THEIR DETECTION BY MACHINE LEARNING METHODS.....	16
<b>M. Kantureyeva, A. Khassenov, D. Tussupov, A. Zakirova, A. Alimagambetova</b> FLOOR FIELD MODEL FOR EVACUATION DYNAMICS.....	30
<b>A.D. Kubegenova, K.T. Iskakov, E.S. Kubegenov, O.I. Krivorotko</b> MONITORING AND MODELING OF THE EPIDEMIOLOGICAL SITUATION USING DATA MINING.....	43
<b>G. Kalman, M.A. Sambetbayeva, A.C. Ilyubayev, D.A. Aktaeva</b> ANAPHORA RESOLUTION MODEL BASED ON MACHINE LEARNING METHODS.....	56
<b>S.T. Mambetov, Ye.Ye. Begimbayeva, S. Joldasbayev, B.O. Kulambayev, G.N. Kazbekova</b> ABOUT MONITORING THREATS AND VULNERABILITIES OF THE INFORMATION SYSTEM.....	68
<b>U. Makhazhanova, B. Tassuov, A. Mukhanova, A. Mukhiyadin, R. Zetkinshekov</b> AN ALGORITHM FOR ASSESSING THE CREDITWORTHINESS OF A BUSINESS BASED ON THE THEORY OF FUZZY SETS.....	81
<b>R.M. Moldasheva, A.A. Ismailova, A.K. Zhamangara, A.M. Zadagali, G.B. Turmukhanova</b> REQUIREMENTS TO DEVELOPMENT OF IAS FOR RESEARCH OF AQUEOUS ECOSYSTEMS.....	93
<b>A. Mukhanova, U. Makhazhanova, N. Markhabatov, B. Tassuov, Zh. Lamasheva</b> APPLICATION OF FUZZY LOGIC IN THE ANALYSIS OF ECONOMIC SYSTEMS N.....	106

<b>N.A. Seilova, A. Batyrgaliyev, Zh. Dzhangozin, D. Baibatchayeva, N. Nurgabylov</b> METHOD FOR ASSESSING THE QUALITY OF MASKING NOISE INTERFERENCES.....	120
<b>A. Khassenov, M. Kantureyeva, D. Tussupov, A. Omarbekova, G. Abdikerimova</b> APPROACH TO THE IMPLEMENTATION OF EVACUATION MODEL IN THE AGENT-BASED MODELING SYSTEM.....	134
<b>A.G. Shaushenova, A.A. Nurpeisova, D.B. Dosalyanov, G.M. Mauina</b> PROBLEMS OF SPEECH RECOGNITION BASED ON ARTIFICIAL NEURAL NETWORKS IN THE PROCTORING SYSTEM.....	146
<b>A. Shekerbek, G. Abdikerimova, A. Sabyr, Zh. Abilkaiyr</b> APPLICATION OF THE METHOD AND ALGORITHM FOR THE DETECTION OF CHEST PATHOLOGY.....	159

**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)**

Директор отдела издания научных журналов НАН РК *А. Ботанқызы*

Заместитель директор отдела издания научных журналов НАН РК *Р. Жәліқызы*

Редакторы: *М.С. Ахметова, Д.С. Аленов*

Верстка на компьютере *Г.Д. Жадыранова*

Подписано в печать 15.09.2022.

Формат 60x88/8. Бумага офсетная. Печать – ризограф.

10,5 п.л. Тираж 300. Заказ 4.