

ISSN 2518-1726 (Online),

ISSN 1991-346X (Print)

**ACADEMIC SCIENTIFIC
JOURNAL OF COMPUTER SCIENCE**

**№2
2026**

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



ACADEMIC SCIENTIFIC JOURNAL OF COMPUTER SCIENCE

3 (359)

JULY – SEPTEMBER 2026

**PUBLISHED SINCE JANUARY 1963
PUBLISHED 4 TIMES A YEAR**

ALMATY, NAS RK

Chief Editor:

MUTANOV Galimkair Mutanovich, doctor of technical sciences, professor, academician of NAS RK, (Almaty, Kazakhstan), <https://www.scopus.com/authid/detail.uri?authorId=6506682964>, <https://www.webofscience.com/wos/author/record/1423665>

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), <https://www.scopus.com/authid/detail.uri?authorId=56153126500>, <https://www.webofscience.com/wos/author/record/2428551>

MAMYRBAEV 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), <https://www.scopus.com/authid/detail.uri?authorId=55967630400>, <https://www.webofscience.com/wos/author/record/1774027>

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), <https://www.scopus.com/authid/detail.uri?authorId=6506823633>, <https://www.webofscience.com/wos/author/record/1923423>

WOICIK Waldemar, Doctor of Technical Sciences (Phys.-Math.), Professor of the Lublin University of Technology (Lublin, Poland), <https://www.scopus.com/authid/detail.uri?authorId=7005121594>, <https://www.webofscience.com/wos/author/record/678586>

SMOLARJ Andrej, Associate Professor Faculty of Electronics, Lublin polytechnic university (Lublin, Poland), <https://www.scopus.com/authid/detail.uri?authorId=56249263000>, <https://www.webofscience.com/wos/author/record/1268523>

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), <https://www.scopus.com/authid/detail.uri?authorId=8701101900>, <https://www.webofscience.com/wos/author/record/1436451>

KHAIROVA Nina, Doctor of Technical Sciences, Professor, Chief Researcher of the Institute of Information and Computational Technologies CS MES RK (Almaty, Kazakhstan), <https://www.scopus.com/authid/detail.uri?authorId=37461441200>, <https://www.webofscience.com/wos/author/record/1768515>

OTMAN Mohamed, PhD, Professor of Computer Science Department of Communication Technology and Networks, Putra University Malaysia (Selangor, Malaysia), <https://www.scopus.com/authid/detail.uri?authorId=56036884700>, <https://www.webofscience.com/wos/author/record/747649>

NYSANBAYEVA Saule Yerkebulanovna, Doctor of Technical Sciences, Associate Professor, Senior Researcher of the Institute of Information and Computing Technologies CS MES RK (Almaty, Kazakhstan), <https://www.scopus.com/authid/detail.uri?authorId=55453992600>, <https://www.webofscience.com/wos/author/record/3802041>

USATOVA Olga Alexandrovna, PhD, Associate Professor, Chief Scientific Secretary of the Institute of Information and Computing Technologies of the National Academy of Sciences of the Republic of Kazakhstan (Almaty, Kazakhstan), <https://www.scopus.com/authid/detail.uri?authorId=57204581062>, <https://www.webofscience.com/wos/author/record/JCO-3058-2023>

KAPALOVA Nursulu Aldazharovna, Candidate of Technical Sciences, Head of the Laboratory cybersecurity, Institute of Information and Computing Technologies CS MES RK (Almaty, Kazakhstan), <https://www.scopus.com/authid/detail.uri?authorId=57191242124>,

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), <https://www.scopus.com/authid/detail.uri?authorId=7202799321>, <https://www.webofscience.com/wos/author/record/38481396>

MIKHALEVICH Alexander Alexandrovich, Doctor of Technical Sciences, Professor, Academician of the National Academy of Sciences of Belarus (Minsk, Belarus), <https://www.scopus.com/authid/detail.uri?authorId=7004159952>, <https://www.webofscience.com/wos/author/record/46249977>

TIGHINEANU Ion Mihailovich, Doctor of Physical and Mathematical Sciences, Academician, President of the Academy of Sciences of Moldova, Technical University of Moldova (Chisinau, Moldova), <https://www.scopus.com/authid/detail.uri?authorId=7006315935>, <https://www.webofscience.com/wos/author/record/524462>

Academic Scientific Journal of Computer Science

ISSN 2518-1726 (Online),

ISSN 1991-346X (Print)

Owner: «Central Asian Academic Research Center» LLP (Almaty).

Certificate № KZ77VPY00121154 on the re-registration of the periodical printed and online publication of the information agency, issued on 05.06.2025 by the Republican State Institution «Information Committee» of the Ministry of Culture and Information of the Republic of Kazakhstan

Subject area: *information and communication technologies*.

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

Periodicity: 4 times a year.

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

© «Central Asian Academic Research Center» LLP, 2026

БАС РЕДАКТОР:

МУТАНОВ Ғалымқайыр Мұтанұлы, техника ғылымдарының докторы, профессор, ҚР ҰҒА академигі, (Алматы, Қазақстан), <https://www.scopus.com/authid/detail.uri?authorId=6506682964>, <https://www.webofscience.com/wos/author/record/1423665>

РЕДАКЦИЯ АЛҚАСЫ:

КАЛИМОЛДАЕВ Мақсат Нұрәділұлы, (бас редактордың орынбасары), физика-математика ғылымдарының докторы, профессор, ҚР ҰҒА академигі, ҚР ҒЖБМ ҒК «Ақпараттық және есептеу технологиялары институты» бас директорының кеңесшісі, зертхана меңгерушісі (Алматы, Қазақстан), <https://www.scopus.com/authid/detail.uri?authorId=56153126500>, <https://www.webofscience.com/wos/author/record/2428551>

МАМЫРБАЕВ Өркен Жұмажанұлы (ғалым хатшы), Ақпараттық жүйелер саласындағы техника ғылымдарының (PhD) докторы, ҚР ҒЖБМ ҒК «Ақпараттық және есептеу технологиялары институты» директорының ғылым жөніндегі орынбасары (Алматы, Қазақстан), <https://www.scopus.com/authid/detail.uri?authorId=55967630400>, <https://www.webofscience.com/wos/author/record/1774027>

БАЙГУНЧЕКОВ Жұмаіл Жаңабайұлы, техника ғылымдарының докторы, профессор, ҚР ҰҒА академигі, Кибернетика және ақпараттық технологиялар институты, Қолданбалы механика және инженерлік графика кафедрасы, Сәтбаев университеті (Алматы, Қазақстан), <https://www.scopus.com/authid/detail.uri?authorId=6506823633>, <https://www.webofscience.com/wos/author/record/1923423>

ВОЙЧИК Вальдемар, техника ғылымдарының докторы (физ-мат), Люблин технологиялық университетінің профессоры (Люблин, Польша), <https://www.scopus.com/authid/detail.uri?authorId=7005121594>, <https://www.webofscience.com/wos/author/record/678586>

СМОЛАРЖ Анджей, Люблин политехникалық университетінің электроника факультетінің доценті (Люблин, Польша), <https://www.scopus.com/authid/detail.uri?authorId=56249263000>, <https://www.webofscience.com/wos/author/record/1268523>

КЕЙЛАН Әлімхан, техника ғылымдарының докторы, профессор (ғылым докторы (Жапония)), ҚР ҒЖБМ ҒК «Ақпараттық және есептеу технологиялары институтының» бас ғылыми қызметкері (Алматы, Қазақстан), <https://www.scopus.com/authid/detail.uri?authorId=8701101900>, <https://www.webofscience.com/wos/author/record/1436451>

ХАЙРОВА Нина, техника ғылымдарының докторы, профессор, ҚР ҒЖБМ ҒК «Ақпараттық және есептеу технологиялары институтының» бас ғылыми қызметкері (Алматы, Қазақстан), <https://www.scopus.com/authid/detail.uri?authorId=37461441200>, <https://www.webofscience.com/wos/author/record/1768515>

ОТМАН Мохаммед, PhD, Информатика, Коммуникациялық технологиялар және желілер кафедрасының профессоры, Путра университеті Малайзия (Селангор, Малайзия), <https://www.scopus.com/authid/detail.uri?authorId=56036884700>, <https://www.webofscience.com/wos/author/record/747649>

НЫСАНБАЕВА Сауле Еркебұланқызы, техника ғылымдарының докторы, доцент, ҚР ҒЖБМ ҒК «Ақпараттық және есептеу технологиялары институтының» аға ғылыми қызметкері (Алматы, Қазақстан), <https://www.scopus.com/authid/detail.uri?authorId=55453992600>, <https://www.webofscience.com/wos/author/record/3802041>

УСАТОВА Ольга Александровна, PhD, қауымдастырылған профессор, ҚР ҒЖБМ "Ақпараттық және есептеу технологиялары институтының" бас ғалым хатшысы (Алматы, Қазақстан), <https://www.scopus.com/authid/detail.uri?authorId=57204581062>, <https://www.webofscience.com/wos/author/record/JCO-3058-2023>

КАПАЛОВА Нұрсұлу Алдажарқызы, техника ғылымдарының кандидаты, ҚР ҒЖБМ ҒК «Ақпараттық және есептеу технологиялары институты», Киберқауіпсіздік зертханасының меңгерушісі (Алматы, Қазақстан), <https://www.scopus.com/authid/detail.uri?authorId=57191242124>,

КОВАЛЕВ Александр Михайлович, физика-математика ғылымдарының докторы, Украина Ұлттық Ғылым академиясының академигі, Қолданбалы математика және механика институты (Донецк, Украина), <https://www.scopus.com/authid/detail.uri?authorId=7202799321>, <https://www.webofscience.com/wos/author/record/38481396>

МИХАЛЕВИЧ Александр Александрович, техника ғылымдарының докторы, профессор, Беларусь Ұлттық Ғылым академиясының академигі (Минск, Беларусь), <https://www.scopus.com/authid/detail.uri?authorId=7004159952>, <https://www.webofscience.com/wos/author/record/46249977>

ТИГИНЯНУ Ион Михайлович, физика-математика ғылымдарының докторы, академик, Молдова Ғылым Академиясының президенті, Молдова техникалық университеті (Кишинев, Молдова), <https://www.scopus.com/authid/detail.uri?authorId=7006315935>, <https://www.webofscience.com/wos/author/record/524462>

Academic Scientific Journal of Computer Science

ISSN 2518-1726 (Online),

ISSN 1991-346X (Print)

Меншіктеуші: «Орталық Азия академиялық ғылыми орталығы» ЖШС (Алматы).

Ақпарат агенттігінің мерзімді баспасөз басылымын, ақпарат агенттігін және желілік басылымды қайта есепке қою туралы ҚР Мәдениет және Ақпарат министрлігі «Ақпарат комитеті» Республикалық мемлекеттік мекемесі **05.06.2025 ж.** берген **№ KZ77VPY00121154** Куәлік.

Тақырыптық бағыты: *ақпараттық-коммуникациялық технологиялар*

Қазіргі уақытта: *«ақпараттық-коммуникациялық технологиялар» бағыты бойынша ҚР БҒМ БҒСБК ұсынған журналдар тізіміне енді.*

Мерзімділігі: *жылына 4 рет.*

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

© «Орталық Азия академиялық ғылыми орталығы» ЖШС, 2026

Главный редактор:

МУТАНОВ Галимканир Мутанович, доктор технических наук, профессор, академик НАН РК, (Алматы, Казахстан), <https://www.scopus.com/authorid/detail.uri?authorId=6506682964>, <https://www.webofscience.com/wos/author/record/1423665>

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), <https://www.scopus.com/authorid/detail.uri?authorId=56153126500>, <https://www.webofscience.com/wos/author/record/2428551>

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), <https://www.scopus.com/authorid/detail.uri?authorId=55967630400>, <https://www.webofscience.com/wos/author/record/1774027>

БАЙГУНЧЕКОВ Жумадил Жанабаевич, доктор технических наук, профессор, академик НАН РК, Институт кибернетики и информационных технологий, кафедра прикладной механики и инженерной графики, Университет Сатпаева (Алматы, Казахстан), <https://www.scopus.com/authorid/detail.uri?authorId=6506823633>, <https://www.webofscience.com/wos/author/record/1923423>

ВОЙЧИК Вальдемар, доктор технических наук (физ.-мат.), профессор Люблинского технологического университета (Люблин, Польша), <https://www.scopus.com/authorid/detail.uri?authorId=7005121594>, <https://www.webofscience.com/wos/author/record/678586>

СМОЛАРЖ Анджей, доцент факультета электроники Люблинского политехнического университета (Люблин, Польша), <https://www.scopus.com/authorid/detail.uri?authorId=56249263000>, <https://www.webofscience.com/wos/author/record/1268523>

КЕЙЛАН Алимхан, доктор технических наук, профессор (Doctor of science (Japan)), главный научный сотрудник РГП «Института информационных и вычислительных технологий» КН МНВО РК (Алматы, Казахстан), <https://www.scopus.com/authorid/detail.uri?authorId=8701101900>, <https://www.webofscience.com/wos/author/record/1436451>

ХАЙРОВА Нина, доктор технических наук, профессор, главный научный сотрудник РГП «Института информационных и вычислительных технологий» КН МНВО РК (Алматы, Казахстан), <https://www.scopus.com/authorid/detail.uri?authorId=37461441200>, <https://www.webofscience.com/wos/author/record/1768515>

ОТМАН Мохамед, доктор философии, профессор компьютерных наук, Департамент коммуникационных технологий и сетей, Университет Путра Малайзия (Селангор, Малайзия), <https://www.scopus.com/authorid/detail.uri?authorId=56036884700>, <https://www.webofscience.com/wos/author/record/747649>

НЫСАНБАЕВА Сауле Еркебулановна, доктор технических наук, доцент, старший научный сотрудник РГП «Института информационных и вычислительных технологий» КН МНВО РК (Алматы, Казахстан), <https://www.scopus.com/authorid/detail.uri?authorId=55453992600>, <https://www.webofscience.com/wos/author/record/3802041>

УСАТОВА Ольга Александровна, PhD, ассоциированный профессор, Главный ученый секретарь «Института информационных и вычислительных технологий» КН МНВО РК (Алматы, Казахстан), <https://www.scopus.com/authorid/detail.uri?authorId=57204581062>, <https://www.webofscience.com/wos/author/record/JCO-3058-2023>

КАПАЛОВА Нурсулу Алдажаровна, кандидат технических наук, заведующий лабораторией кибербезопасности РГП «Института информационных и вычислительных технологий» КН МНВО РК (Алматы, Казахстан), <https://www.scopus.com/authorid/detail.uri?authorId=57191242124>

КОВАЛЕВ Александр Михайлович, доктор физико-математических наук, академик НАН Украины, Институт прикладной математики и механики (Донецк, Украина), <https://www.scopus.com/authorid/detail.uri?authorId=7202799321>, <https://www.webofscience.com/wos/author/record/38481396>

МИХАЛЕВИЧ Александр Александрович, доктор технических наук, профессор, академик НАН Беларуси (Минск, Беларусь), <https://www.scopus.com/authorid/detail.uri?authorId=7004159952>, <https://www.webofscience.com/wos/author/record/46249977>

ТИГИНЯНУ Ион Михайлович, доктор физико-математических наук, академик, президент Академии наук Молдовы, Технический университет Молдовы (Кишинев, Молдова), <https://www.scopus.com/authorid/detail.uri?authorId=7006315935>, <https://www.webofscience.com/wos/author/record/524462>

Academic Scientific Journal of Computer Science

ISSN 2518-1726 (Online),

ISSN 1991-346X (Print)

Собственник: *ТОО «Центрально-азиатский академический научный центр» (г. Алматы).*

Свидетельство о постановке на переучет периодического печатного издания, информационного агентства и сетевого издания № **KZ77VPY00121154**. Дата выдачи **05.06.2025**

Тематическая направленность: *информационно-коммуникационные технологии.*

В настоящее время: *вошел в список журналов, рекомендованных КОКШВО МНВО РК по направлению «информационно-коммуникационные технологии».*

Периодичность: *4 раза в год.*

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

© ТОО «Центрально-азиатский академический научный центр», 2026

CONTENTS

Akhmetova A.Zh., Kantureyeva M.A., Serimbetov B.A., Zhasuzakova M.Zh., Mustafayeva N.T. Identification of communities and active users in social networks.....	11
Akzhigitov D., Mukhtarkhanov M., Zhetpisbayeva A., Kulakayeva A., Zhetpisbayev K. Analysis of the mechanical structure of a 12U CubeSat nanosatellite and assessment of its mechanical stability.....	29
Alibekov A., Nurimanov A., Mukhametkali A., Kassym D., Varol H.A. Qolda-Math: a small vision-language model for mathematical reasoning in Kazakh.....	50
Alisheva S.S., Ospanova A.O., Shaimerdenova G.S., Shaimerdenova L., Zhidebayeva A.N. Detecting Kazakh language offensive comments in online content using deep learning algorithms.....	68
Askar G., Mutaikhan A., Aitkozhayeva M., Nauryzbaikyzy G., Makhataeva Z. Retrieval beats fine-tuning: adapting a Kazakh small language model for historical question answering.....	86
Beissenova G.I., Shynzhigit B.B., Beisebayeva A.Zh., Kaldarova B.S., Lesbayev A.U. Deep neural network-based speaker authentication for secure access control systems.....	106
Bekenuly B., Baimakhanova A., Baiturganova V., Bolatbekova D. Comparison of CNN architectures for automated ACL injury classification from knee MRI.....	123
Bekbolat A., Razak A., Alimbayeva Z., Ozhikenova A. Multi-task deep learning for knee MRI pathology classification.....	139
Issakhov A., Akylbekkyzy A., Abylkassymova A. Predicting the spread of airborne viral infections in enclosed spaces.....	158
Kisseleva O., Abdramanov K., Naumenko V., Mukanova Z., Maintser D. Evaluation of serverless cold start optimization techniques.....	177
Koshkarov Ye.Zh., Amirgaliyev B.Ye., Mulesa O. Physics-aware masked pretraining for cross-basin well log foundation models.....	197
Kurmetkan T., Turysbek Zh., Mukhitova A., Nauryzbaeva A. Research on methods for reducing knowledge conflicts in large language models in Kazakh question-and-answer systems.....	214
Makatov Ye.K., Abdul Razaque, Ismukanova A.N., Salykova A.K., Glock E.S. Machine learning-based model for phishing URL detection in online social networks.....	233
Mimenbayeva A.B., Naizagarayeva A.A., Ospanova T.T., Koikelova D.K., Kudabayeva R.N. Spatial and climate data analysis for machine learning-based crop yield prediction in Northern Kazakhstan.....	255
Mukhamediev R.I., Abdoldina F.N., Gertsen Y.A. Economic opportunities and constraints of artificial intelligence adoption in Kazakhstan's resource-oriented economy.....	274

Mutanov G.M., Karyukin V.I., Ussatova O.A., Begimbayeva Y.Y., Baispay G.B. The assessment of blockchain readiness for academic document verification.....	289
Mussiraliyeva Sh., Bolatbek M., Baisylbayeva K. Classification of cyber-extremist ideologies in the Kazakh language considering multi-annotator disagreement.....	306
Omarov B.S., Sultan D.R., Rakhymberdiev D.S., Abibullayeva A.A., Toktarova A.B. Deformable attention-guided Transformer for brain tumor segmentation.....	326
Oralbekova D., Bekarystankyzy A., Yerimbetova A., Berzhanova U. Development of a streaming Kazakh speech recognition system using end-to-end models.....	346
Serek A., Shoiynbek A., Zheludkov M., Kuanyshbay D., Meraliyev B. Comparative analysis of machine learning methods for detecting fraud in text messages.....	365
Smakhanova A.K., Uskenbayeva R.K., Kassymova A.B., Kalpeyeva Zh.B., Konyrbaev N.B. Experimental research using models and methods of the university digital ecosystem.....	384
Sultanali A.E., Kazbekova G.N. Machine learning-based detection of AI-generated content in students' written works.....	401
Tashenova Zh., Yergazin A., Nurlybaeva E., Abdugulova Zh., Amanzholova Sh. Experimental evaluation of a hybrid IDPS with adaptive learning for low-latency 5G networks using edge computing.....	418
Telman D., Yerimbetova A., Daiyrbayeva E. Leakage-resistant fake news detection for Kazakh: temporal evaluation of lexical, Transformer and hybrid models.....	441
Zhanegizov A., Mamyrbayev O., Abdigappar M. Application of YOLO12 for road defect detection in GPR images.....	457

МАЗМҰНЫ

Ахметова А.Ж., Кантуреева М.А., Серимбетов Б.А., Жасұзақова М.Ж., Мұстафаева Н.Т. Әлеуметтік желілердегі қауымдастықтар мен белсенді қолданушыларды анықтау.....	11
Ақжігітов Д.Ж., Мұхтарханов М.О., Жетписбаева А.Т., Кулакаева А.Е., Жетписбаев К.У. 12U CubeSat үлгісіндегі наноспутниктің механикалық құрылымын талдау және оның механикалық орнықтылығын бағалау.....	29
Алибеков А., Нуриманов А., Мұхаметқали А., Қасым Д., Варол Х.А. Qolda-Math: қазақ тіліндегі математикалық пайымдауға арналған кіші көру-тілдік модель.....	50
Алишева С.С., Оспанова А.О., Шаймерденова Г.С., Шаймерденова Л., Жидебаева А.Н. Онлайн-контенттегі қазақ тіліндегі бейәдеп пікірлерді терең оқыту алгоритмдері арқылы анықтау.....	68
Асқар Ғ., Мұтайхан А., Айтқожаева М., Наурызбайқызы Г., Махатаева Ж. Retrieval fine-tuning-тен басым: тарихи сұрақтарға жауап беру үшін қазақша шағын тілдік модельді бейімдеу.....	86
Бейсенова Г.И., Шынжігіт Б.Б., Бейсебаева А.Ж., Қалдарова Б.С., Лесбаев А.У. Терең нейрондық желіге негізделген қауіпсіз басқару жүйелерінде сөйлеушіні аутентификациялау.....	106
Бекенұлы Б., Баймаханова А., Байтурганова В., Болатбекова Д. Алдыңғы крест тәрізді байлам жарақатын анықтауға арналған CNN архитектураларын салыстыру.....	123
Бекболат А., Разак А., Алимбаева Ж., Ожикенова А. Tize MPT патологиясын жіктеуге арналған көп тапсырмалы терең оқыту.....	139
Исахов А.А., Ақылбекқызы А., Абылкасымова А. Шектеулі кеңістіктерде әуе тамшыларымен вирустық инфекцияның таралуын сандық модельдеу.....	158
Киселева О., Абдраманов Қ., Науменко В., Мұқанова Ж., Майнцер Д. Serverless жүйелеріндегі суық іске қосуды оңтайландыру әдістерін эксперименттік бағалау.....	177
Кошкарров Е.Ж., Амиргалиев Б.Е., Мүлеса О. Бассейнаралық ұңғымалық диаграммалардың іргелі модельдерін физикалық тұрғыдан негізделген масқаланған алдын ала оқыту.....	197
Құрметқан Т., Тұрысбек Ж., Мұхитова А., Наурызбаева А. Қазақ тіліндегі сұрақ-жауап жүйелерінде үлкен тілдік модельдердегі білім қайшылықтарын азайту әдістерін зерттеу.....	214
Мақатов Е.К., Abdul Razaque, Исмуканова А.Н., Салыкова А.К., Глюк Е.С. Онлайн әлеуметтік желілердегі фишингтік URL-дарды анықтауға арналған машиналық оқыту негізіндегі модель.....	233

Мименбаева А.Б., Найзағараева А.А., Оспанова Т.Т., Қойкелова Д.К., Қудабаева Р.Н. Солтүстік Қазақстандағы дақылдар өнімділігін машиналық оқыту арқылы болжауда кеңістіктік және климаттық деректерді талдау.....	255
Мухамедиев Р.И., Абдолдина Ф.Н., Герцен Е.А. Қазақстанның ресурстық бағытталған экономикасына жасанды интеллектті енгізудің экономикалық мүмкіндіктері мен шектеулері.....	274
Мутанов Г.М., Карюкин В.И., Усатова О.А., Бегимбаева Е.Е., Байспай Г.Б. Білім беру құжаттарының түпнұсқалығын тексеруге блокчейн технологиясының дайындығын бағалау.....	289
Мусиралиева Ш., Болатбек М., Байсылбаева К. Көпаннынотаторлы келіспеушілікті ескере отырып қазақ тіліндегі киберэкстремистік идеологияларды жіктеу.....	306
Омаров Б.С., Сұлтан Д.Р., Рахымбердиев Д.С., Абибуллаева А.А., Тоқтарова А.Б. Ми ісіктерін сегменттеуде деформацияланатын зейін механизміне негізделген мультимодальды трансформер.....	326
Оралбекова Д.О., Бекарыстанқызы А., Еримбетова А.С., Бержанова У.Г. Интегралды модельдерді қолдану арқылы қазақ тіліндегі ағындық сөйлеуді тану жүйесін әзірлеу.....	346
Серек А., Шойынбек А., Желудков М., Қуанышбай Д., Мералиев Б. Мәтіндік хабарламалардағы алаяқтықты анықтауға арналған машиналық оқыту әдістерінің салыстырмалы талдауы.....	365
Смаханова А.К., Ускенбаева Р.К., Касымова А.Б., Кальпеева Ж.Б., Қоңырбаев Н.Б. Университеттің цифрлық экожүйесінің модельдері мен әдістерін қолдана отырып жүргізілген эксперименттік зерттеу.....	384
Сұлтанәлі А.Е., Казбекова Г.Н. Білім алушылардың жазбаша жұмыстарындағы AI-контентті машиналық оқыту арқылы анықтау.....	401
Ташенова Ж.М., Ергазин Ә., Нұрлыбаева Ә.Н., Абдуғулова Ж.К., Аманжолова Ш.А. Edge Computing көмегімен төмен кідірісті 5G желілеріне арналған бейімделгіш оқытуы бар гибриді IDPS-ті эксперименттік бағалау.....	418
Тельман Д., Еримбетова А., Дайырбаева Ә. Қазақ тіліндегі жалған жаңалықтарды анықтау: лексикалық, трансформерлік және гибридті модельдерді уақыт бойынша бағалау.....	441
Жанегизов А.С., Мамырбаев Ә.Ж., Абдіғаппар М.Д. YOLO12 моделін георадар кескіндеріне қолдану.....	457

СОДЕРЖАНИЕ

Ахметова А.Ж., Кантуреева М.А., Серимбетов Б.А., Жасузакова М.Ж., Мустафаева Н.Т. Выявление сообществ и активных пользователей в социальных сетях.....	11
Акжигитов Д.Ж., Мухтарханов М.О., Жетписбаева А.Т., Кулакаева А.Е., Жетписбаев К.У. Анализ механической конструкции наноспутника модели 12U CubeSat и оценка его механической устойчивости.....	29
Алибеков А., Нуриманов А., Мухаметкали А., Касым Д., Варол Х.А. Qolda-Math: маленькая мультимодальная модель зрения и языка для математических рассуждений на казахском языке.....	50
Алишева С.С., Оспанова А.О., Шаймерденова Г.С., Шаймерденова Л., Жидебаева А.Н. Выявление оскорбительных комментариев на казахском языке в онлайн-контенте с помощью алгоритмов глубокого обучения.....	68
Аскар Г., Мутайхан А., Айткожаева М., Наурызбайкызы Г., Махатаева Ж. Retrieval эффективнее fine-tuning: адаптация малой языковой модели казахского языка для ответов на исторические вопросы.....	86
Бейсенова Г.И., Шынжигит Б.Б., Бейсебаева А.Ж., Калдарова Б.С., Лесбаев А.У. Аутентификация личности по голосу на основе глубокой нейронной сети для систем безопасного управления доступом.....	106
Бекенулы Б., Баймаханова А., Байтурганова В., Болатбекова Д. Сравнение архитектур сверточных нейронных сетей для автоматической классификации разрыва ПКС по МРТ коленного сустава.....	123
Бекболат А., Разак А., Алимбаева Ж., Ожикенова А. Многозадачное глубокое обучение для классификации патологий колена по МРТ.....	139
Исахов А.А., Акылбеккызы А., Абылкасымова А. Прогнозирование распространения вирусной инфекции воздушно-капельным путем в замкнутых помещениях.....	158
Киселева О., Абдраманов К., Науменко В., Муканова Ж., Майнцер Д. Экспериментальная оценка методов оптимизации холодного запуска в serverless-системах.....	177
Кошкарар Е.Ж., Амиргалиев Б.Е., Мулеса О. Физически-осознанное маскированное предобучение для межбассейновых фундаментальных моделей каротажных диаграмм.....	197
Курметкан Т., Турысбек Ж., Мухитова А., Наурызбаева А. Исследование методов снижения конфликтов знаний в больших языковых моделях казахских вопросно-ответных систем.....	214
Макатов Е.К., Abdul Razaque, Исмуканова А.Н., Салыкова А.К., Глок Е.С. Модель обнаружения фишинговых URL в онлайн-социальных сетях на основе машинного обучения.....	233

Мименбаева А.Б., Найзагараева А.А., Оспанова Т.Т., Койкелова Д.К., Кудабаяева Р.Н. Анализ климатических данных для прогнозирования урожайности пшеницы методами машинного обучения в Северном Казахстане.....	255
Мухамедиев Р.И., Абдолдина Ф.Н., Герцен Е.А. Экономические возможности и ограничения внедрения искусственного интеллекта в ресурсно-ориентированной экономике Казахстана.....	274
Мутанов Г.М., Карюкин В.И., Усатова О.А., Бегимбаева Е.Е., Байспай Г.Б. Оценка готовности технологии блокчейн к проверке подлинности документов об образовании.....	289
Мусиралиева Ш., Болатбек М., Байсылбаева К. Классификация киберэкстремистских идеологий на казахском языке с учетом разногласий нескольких аннотаторов.....	306
Омаров Б.С., Султан Д.Р., Рахымбердиев Д.С., Абибуллаева А.А., Токтарова А.Б. Мультимодальный трансформер на основе механизма деформируемого внимания для сегментации опухолей головного мозга.....	326
Оралбекова Д.О., Бекарыстанкызы А., Еримбетова А.С., Бержанова У.Г. Распознавание казахской речи с применением интегральных моделей.....	346
Серек А., Шойынбек А., Желудков М., Куанышбай Д., Мералиев Б. Сравнительный анализ методов машинного обучения для выявления мошенничества в текстовых сообщениях.....	365
Смаханова А.К., Ускенбаева Р.К., Касымова А.Б., Кальпеева Ж.Б., Конырбаев Н.Б. Экспериментальное исследование с использованием моделей и методов цифровой экосистемы университета.....	384
Султанали А.Е., Казбекова Г.Н. Обнаружение AI-контента в письменных работах обучающихся методами машинного обучения.....	401
Ташенова Ж.М., Ергазин А., Нурлыбаева Э.Н., Абдугулова Ж.К., Аманжолова Ш.А. Экспериментальная оценка гибридной системы обнаружения и предотвращения вторжений (IDPS) с адаптивным обучением для сетей 5G с низкой задержкой с использованием граничных вычислений.....	418
Тельман Д., Еримбетова А., Дайырбаева Э. Обнаружение фейковых новостей на казахском языке: временная оценка лексических, трансформерных и гибридных моделей.....	441
Жанегизов А.С., Мамырбаев О.Ж., Абдигаппар М.Д. Применение модели YOLOv12 для анализа данных георадара.....	457

ACADEMIC SCIENTIFIC JOURNAL OF COMPUTER SCIENCE
ISSN 1991-346X
Volume 3.
Number 359 (2026). 356–383

<https://doi.org/10.32014/2026.2518-1726.468>

IRSTI 28.23.25

UDC 004.855.5

© Serek A.¹, Shoiynbek A.², Zheludkov M.^{2*}, Kuanyshbay D.²,
Meraliyev B.³, 2026.

¹ Astana IT University, Astana, Kazakhstan;

² Narxoz University, Almaty, Kazakhstan;

³ SDU University, Kaskelen, Kazakhstan.

E-mail: maxim.zheludkov@narxoz.kz

COMPARATIVE ANALYSIS OF MACHINE LEARNING METHODS FOR DETECTING FRAUD IN TEXT MESSAGES

Azamat Serek — PhD, Assistant-Professor, Astana IT University, Astana, Kazakhstan,
E-mail: azamat.serek@astanait.edu.kz. ORCID ID: <https://orcid.org/0000-0001-7096-6765>;

Aisultan Shoiynbek — PhD, Professor, Narxoz University, Almaty, Kazakhstan,
E-mail: aisultan.shoiynbek@gmail.com. ORCID ID: <https://orcid.org/0000-0002-9328-8300>;

Maxim Zheludkov — Master's student, Narxoz University, Almaty, Kazakhstan,
E-mail: maxim.zheludkov@narxoz.kz. ORCID ID: <https://orcid.org/0009-0002-2819-5201>;

Darkhan Kuanyshbay — PhD, Assistant-Professor, Narxoz University, Almaty, Kazakhstan,
E-mail: darkhan.kuanyshbay@narxoz.kz. ORCID ID: <https://orcid.org/0000-0001-5952-8609>;

Bakhtiyor Meraliyev — Master, Assistant-Professor, SDU University, Kaskelen, Kazakhstan,
E-mail: bakhtiyor.meraliyev@sdu.edu.kz. ORCID ID: <https://orcid.org/0000-0002-5640-4966>.

Abstract. The rapid migration of financial fraud to encrypted messenger platforms such as WhatsApp and Telegram has created an urgent need for lightweight, endpoint-based detection tools capable of operating on small, multilingual datasets characteristic of real-world "cold start" conditions. Unlike traditional email spam, messenger-based fraud is short-lived, conversational, and heavily obfuscated through mixed scripts and artificial urgency, rendering static rule-based filters ineffective. Existing machine learning solutions predominantly address English-language SMS spam and rely on large labeled corpora, leaving a significant gap for post-Soviet and Central Asian deployment contexts where Russian, Kazakh, and English frequently co-occur within a single message. This study aims to identify the optimal classical machine learning architecture for binary fraud classification of such multilingual text messages under severe data scarcity constraints. A balanced pilot dataset of 189 messages (95 legitimate / 94 fraudulent) was manually collected from Telegram, WhatsApp, Reddit, Pikabu, and verified news sources. Text was vectorized using TF-IDF (unigrams and bigrams, max 100 features, min_df=2)

without morphological preprocessing to preserve cross-language token integrity. Five classifiers were benchmarked: Logistic Regression, Linear SVM, Multinomial Naive Bayes, Random Forest, and XGBoost. Regularization strategies including L2 penalty, additive smoothing, and structural tree constraints, combined with class-weight balancing, were applied to mitigate overfitting on the small sample. Model evaluation used 5-fold stratified cross-validation and an independent test set, with Recall (Fraud) as the primary safety metric alongside Accuracy, F1-score, and ROC-AUC. Logistic Regression and Linear SVM achieved the highest overall performance (Accuracy = 0.97, Recall = 1.00, F1 = 0.97, ROC-AUC = 0.957–0.962), successfully identifying all fraudulent messages with only one false alarm each. XGBoost also reached perfect recall (1.00) at a cost of lower precision (Precision = 0.88), while Random Forest and Naive Bayes were less stable, achieving accuracies of 0.86 and 0.79 respectively. The findings demonstrate that for small, sparse, and multilingual fraud datasets, aggressively regularized linear models with minimal feature spaces outperform complex ensemble architectures, offering a viable, computationally lightweight foundation for real-time endpoint fraud filtering on encrypted platforms.

Keywords: Fraud Detection, Machine Learning, Cold Start Problem, Small-sample Learning, Multilingual Text Classification

Information on funding. *This research has been funded by the Science Committee of the Ministry of Science and Higher Education of the Republic of Kazakhstan (Grant No. AP27510301 “Development of technology for recognizing fraudulent actions during a telephone conversation and/or text message exchange in messengers based on artificial intelligence algorithms”).*

For citations: Serek A., Shoiynbek A., Zheludkov M., Kuanyshbay D., Meraliyev B. Comparative analysis of machine learning methods for detecting fraud in text messages. Academic Scientific Journal of Computer Science, 2026. — No.3. — P. 356–383. DOI: <https://doi.org/10.32014/2026.2518-1726.468>

© Серек А.¹, Шойынбек А.², Желудков М.^{2*}, Қуанышбай Д.²,
Мералиев Б.³, 2026.

¹ Astana IT University, Астана, Қазақстан;

² Narxoz University, Алматы, Қазақстан;

³ SDU University, Қаскелең, Қазақстан.

E-mail: maxim.zheludkov@narxoz.kz

МӘТІНДІК ХАБАРЛАМАЛАРДАҒЫ АЛАЯҚТЫҚТЫ АНЫҚТАУҒА АРНАЛҒАН МАШИНАЛЫҚ ОҚЫТУ ӘДІСТЕРІНІҢ САЛЫСТЫРМАЛЫ ТАЛДАУЫ

Серек Азамат — PhD, ассистент-профессор, Astana IT University, Астана, Қазақстан,
E-mail: azamat.serek@astanait.edu.kz, <https://orcid.org/0000-0001-7096-6765>;

Шойынбек Айсұлтан — PhD, профессор, Narxoz University, Алматы, Қазақстан,
E-mail: aisultan.shoynbek@gmail.com, <https://orcid.org/0000-0002-9328-8300>;

Желудков Максим — магистрант, Narxoz University, Алматы, Қазақстан,
E-mail: maxim.zheludkov@narxoz.kz, <https://orcid.org/0009-0002-2819-5201>;

Қуанышбай Дархан — PhD, ассистент-профессор, Narxoz University, Алматы, Қазақстан,
E-mail: darkhan.kuanyshbay@narxoz.kz, <https://orcid.org/0000-0001-5952-8609>;

Мералиев Бахтиёр — магистр, ассистент-профессор, SDU University, Қаскелең, Қазақстан,
E-mail: bakhtiyor.meraliyev@sdu.edu.kz, <https://orcid.org/0000-0002-5640-4966>.

Аннотация. Қаржылық алаяқтықтың WhatsApp және Telegram сияқты шифрланған мессенджерлерге жылдам ауысуы нақты жағдайларға тән шағын әрі көптілді деректер жиынтықтарымен жұмыс істей алатын жеңіл анықтау құралдарына деген қажеттілікті арттырды. Дәстүрлі электрондық пошта спамынан айырмашылығы, мессенджерлердегі алаяқтық қысқа мерзімді және диалогтік сипатқа ие әрі аралас жазуды, бірнеше тілді және жасанды шұғылдық сезімін белсенді пайдаланады, бұл статикалық ережелерге негізделген сүзгілердің тиімділігін төмендетеді. Қолданыстағы машиналық оқыту шешімдері көбіне ағылшын тіліндегі SMS-спамға бағытталған және үлкен таңбаланған корпустарға сүйенеді. Бұл бір хабарлама ішінде орыс, қазақ және ағылшын тілдері қатар қолданылатын посткеңестік және Орталық Азия кеңістігі үшін маңызды зерттеу олқылығын қалыптастырады. Зерттеудің мақсаты - деректердің айқын тапшылығы жағдайында көптілді мәтіндік хабарламалардағы алаяқтықты бинарлық жіктеуге арналған классикалық машиналық оқыту әдістерінің ішінен ең тиімді архитектураны анықтау. Telegram, WhatsApp, Reddit, Pikabu және тексерілген жаңалықтар көздерінен қолмен жиналған 189 хабарламадан тұратын теңгерімді пилоттық деректер жиынтығы құрылды, оның 95-і заңды және 94-і алаяқтық хабарламалар болды. Мәтіндер тілдер арасындағы токендердің тұтастығын сақтау үшін морфологиялық алдын ала өңдеусіз униграммалар мен биграммаларға негізделген TF-IDF әдісімен векторланды; белгілердің максималды саны 100, $\min_df = 2$ болып белгіленді. Logistic Regression, Linear SVM, Multinomial Naive Bayes, Random Forest және XGBoost модельдері салыстырылды. Шағын іріктемеде артық үйренуді азайту үшін L2-регуляризация, аддитивті тегістеу, ағаш модельдерінің құрылымдық шектеулері және сынып салмақтарын теңестіру қолданылды. Бағалау бес бөлікті стратификацияланған кросс-валидация және тәуелсіз тестілік жиынтық негізінде жүргізілді. Негізгі қауіпсіздік метрикасы ретінде алаяқтық класы үшін Recall көрсеткіші, қосымша Accuracy, F1-score және ROC-AUC метрикалары қолданылды. Logistic Regression және Linear SVM ең жоғары жалпы нәтижелерді көрсетті: Accuracy = 0,97, Recall = 1,00, F1-score = 0,97, ROC-AUC = 0,957–0,962. Екі модель де барлық алаяқтық хабарламаларды дұрыс анықтап, тек бір жалған оң нәтиже көрсетті. XGBoost та Recall = 1,00 нәтижесіне қол жеткізді, алайда Precision = 0,88 болды. Random Forest және Naive Bayes модельдері тиісінше 0,86 және 0,79 Accuracy көрсеткіштерімен төменірек тұрақтылық

көрсетті. Алынған нәтижелер шағын, сирек және көптілді алаяқтық деректер жиынтықтарында ықшам белгілер кеңістігі бар қатаң регуляризацияланған сызықтық модельдер күрделі ансамбльдік архитектуралардан тиімдірек болуы мүмкін екенін көрсетеді. Мұндай модельдер шифрланған мессенджерлік платформаларда алаяқтықты нақты уақыт режимінде құрылғы деңгейінде сүзуге арналған есептеу тұрғысынан жеңіл негіз бола алады.

Түйін сөздер: алаяқтықты анықтау; машиналық оқыту; «суық старт» мәселесі; шағын іріктемелермен оқыту; көптілді мәтіндерді жіктеу

© Серек А.¹, Шойынбек А.², Желудков М.^{2*}, Куанышбай Д.²,
Мералиев Б.³, 2026.

¹ Astana IT University, Астана, Қазақстан;

² Narxoz University, Алматы, Қазақстан;

³ SDU University, Каскелен, Қазақстан.

E-mail: maxim.zheludkov@narxoz.kz

СРАВНИТЕЛЬНЫЙ АНАЛИЗ МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ ДЛЯ ВЫЯВЛЕНИЯ МОШЕННИЧЕСТВА В ТЕКСТОВЫХ СООБЩЕНИЯХ

Серек Азамат — PhD, ассистент-профессор, Astana IT University, Астана, Қазақстан,
E-mail: azamat.serek@astanait.edu.kz, <https://orcid.org/0000-0001-7096-6765>;

Шойынбек Айсултан — PhD, профессор, Narxoz University, Алматы, Қазақстан,
E-mail: aisultan.shoynbek@gmail.com, <https://orcid.org/0000-0002-9328-8300>;

Желудков Максим — магистрант, Narxoz University, Алматы, Қазақстан,
E-mail: maxim.zheludkov@narxoz.kz, <https://orcid.org/0009-0002-2819-5201>;

Куанышбай Дархан — PhD, ассистент-профессор, Narxoz University, Алматы, Қазақстан,
E-mail: darkhan.kuanysbay@narxoz.kz, <https://orcid.org/0000-0001-5952-8609>;

Мералиев Бахтиёр — магистр, ассистент-профессор, SDU University, Каскелен, Қазақстан,
E-mail: bakhtiyor.meraliyev@sdu.edu.kz, <https://orcid.org/0000-0002-5640-4966>.

Аннотация. Стремительное перемещение финансового мошенничества на зашифрованные мессенджер-платформы, такие как WhatsApp и Telegram, обуславливает необходимость разработки легковесных средств обнаружения, способных работать с небольшими многоязычными наборами данных, характерными для реальных условий «холодного старта». В отличие от традиционного почтового спама, мошенничество в мессенджерах носит краткосрочный и диалоговый характер, активно использует смешение письменностей, языков и искусственно создаваемое ощущение срочности, что снижает эффективность статических фильтров на основе правил. Существующие решения машинного обучения преимущественно ориентированы на англоязычный SMS-спам и используют крупные размеченные корпуса, вследствие чего сохраняется существенный пробел для постсоветского и центральноазиатского пространства, где русский, казахский и английский языки часто встречаются в одном сообщении. Цель

исследования - определить наиболее эффективную архитектуру классического машинного обучения для бинарной классификации многоязычных мошеннических текстовых сообщений в условиях острого дефицита данных. Был сформирован сбалансированный пилотный набор из 189 сообщений, включающий 95 легитимных и 94 мошеннических сообщения, вручную собранных из Telegram, WhatsApp, Reddit, Pikabu и проверенных новостных источников. Тексты векторизовались методом TF-IDF на основе униграмм и биграмм при максимальном количестве 100 признаков и $\text{min_df} = 2$ без морфологической предобработки, что позволило сохранить межъязыковую целостность токенов. Сравнивались пять классификаторов: Logistic Regression, Linear SVM, Multinomial Naive Bayes, Random Forest и XGBoost. Для снижения риска переобучения применялись L2-регуляризация, аддитивное сглаживание, структурные ограничения деревьев и балансировка весов классов. Оценка проводилась с использованием пятикратной стратифицированной кросс-валидации и независимой тестовой выборки. Основной метрикой безопасности являлся Recall для класса мошенничества; дополнительно оценивались Accuracy, F1-score и ROC-AUC. Logistic Regression и Linear SVM продемонстрировали наилучшие результаты: Accuracy = 0,97, Recall = 1,00, F1-score = 0,97, ROC-AUC = 0,957–0,962. Обе модели выявили все мошеннические сообщения, допустив лишь по одному ложноположительному срабатыванию. XGBoost также достиг Recall = 1,00, однако его Precision составил 0,88. Random Forest и Naive Bayes показали менее стабильные результаты с Accuracy 0,86 и 0,79 соответственно. Результаты показывают, что на небольших, разреженных и многоязычных наборах данных о мошенничестве строго регуляризованные линейные модели с компактным пространством признаков могут превосходить более сложные ансамблевые архитектуры. Такой подход формирует вычислительно эффективную основу для фильтрации мошеннических сообщений в реальном времени непосредственно на конечных устройствах.

Ключевые слова: выявление мошенничества; машинное обучение; проблема холодного старта; обучение на малых выборках; многоязычная классификация текстов

Introduction. The current cybersecurity environment shows attackers now prefer social engineering tactics instead of trying to breach infrastructure systems (Wang et al., 2021). The main communication method which phishers and fraudsters use today consists of text channels which include SMS and WhatsApp and Telegram (Lim and Phuangsuwan, 2025). The filtering solutions which exist for email spam do not work against mobile text fraud because this type of fraud exists as a constantly changing and short-lived threat (Salman et al., 2024). The methods used by fraudsters to commit their crimes continue to change because they employ techniques which include using confusing language and combining different writing systems and creating a sense of emergency to outsmart standard

security systems based on predefined rules. The situation demands immediate development of adaptable lightweight Machine Learning (ML) classifiers which can identify semantic meaning instead of recognizing specific words (Tyagi and Szénási, 2024).

This study is a pilot project on binary classification of fraudulent Russian-language texts using classical machine learning methods. To achieve this goal, a database of unique examples was compiled, a comparative analysis of five architectures was conducted, and hyperparameter tuning was thoroughly explored in order to mitigate overfitting on a small dataset.

The main research contributions include the introduction of a specialized dataset reflecting "cold-start" fraud scenarios, an empirical evaluation of regularization strategies for sparse multilingual data, and the identification of lightweight linear models that achieve perfect recall, providing a robust methodological framework for effective learning from limited samples (Charilaou, and Battat, 2022).

Literature Review. The automatic detection of fraudulent text messages (smishing) has evolved into a significant subfield of cybersecurity, driven by the increasing reliance on mobile messaging platforms for financial authentication. Early research in this domain predominantly focused on adapting email spam filtering techniques to Short Message Services (SMS). Feature extraction methods that are commonly used are the Bag-of-Words (BoW) and Term Frequency-Inverse Document Frequency (TF-IDF) that convert the text into a numerical vector. Early studies demonstrated that probabilistic classifiers, particularly Naive Bayes and Support Vector Machines (SVM), could achieve high accuracy on standard datasets like the UCI SMS Spam Collection due to the distinct lexical features of promotional spam (Almeida et al., 2011). They rely heavily on the static keyword set presented in the texts and are therefore not adequate for dealing with informal and brief text messages, which are common in mobile communication. Moreover, these early datasets were collected under controlled conditions and do not reflect the adversarial evolution of modern fraud, where attackers deliberately introduce typographical noise, abbreviations, and mixed-language constructs to evade keyword-based detection.

A critical evolution in this domain is the rapid migration of fraud from traditional cellular SMS to Over-The-Top (OTT) instant messengers such as WhatsApp and Telegram. Unlike SMS, which are limited by the number of characters and the cost of cellular communication, messengers provide virtually unlimited text length and rich formatting, which allows you to create more complex social engineering plots. This shift presents a unique challenge, as the end-to-end encryption employed by these platforms renders traditional network-level filtering ineffective, necessitating lightweight endpoint detection solutions (Lim and Phuangsuwan, 2025). Furthermore, messenger-based attacks are more dynamic and short-lived than email spam, often employing "conversational" fraud that bypasses static rule-based filters designed for bulk SMS (Salman et al., 2024).

Convolutional Neural Networks (CNNs) and Long Short-Term Memory

(LSTM) networks are often employed to model the sequential data in the messages. Also, a hybrid approach using combined CNN-LSTM models has been shown to significantly outperform traditional machine learning classifiers using handcrafted features, and are able to learn high-level features from the character-level embeddings (Ghourabi, 2022). Similarly, Transformer-based models, such as BERT and RoBERTa, have set new benchmarks in text classification by leveraging attention mechanisms to understand the intent behind a message (Mishra and Soni, 2022). These deep architectures excel at detecting subtle obfuscation techniques used by attackers, such as homoglyph attacks and context-based social engineering. However, deploying such models in production requires substantial computational resources: BERT-base contains 110 million parameters and demands GPU inference for real-time throughput, making it impractical for on-device or edge deployment scenarios. Furthermore, fine-tuning transformer models on datasets of fewer than 500 samples frequently lead to unstable training dynamics and poor generalization, limiting their applicability in cold-start fraud detection contexts (Charilaou and Battat, 2022).

Despite these technological advancements, a significant gap remains in addressing the multilingual and code-switching nature of modern fraud. The vast majority of existing research relies on monolingual English datasets, such as the UCI corpus or the TiW-SMS dataset (Jain et al., 2021). In real-world scenarios, particularly in post-Soviet and Central Asian regions, fraudulent messages frequently mix languages (e.g., Russian, English, and Kazakh) or use Latin scripts for Cyrillic terms to bypass filters. It has been noted that standard NLP preprocessing pipelines, such as lemmatization and stemming, often degrade performance on such mixed-script data. This degradation occurs because morphological analyzers are typically trained on monolingual corpora and fail to correctly segment tokens that span two writing systems or contain code-switched affixes. As a result, language-agnostic approaches such as character-level n-grams or raw TF-IDF over mixed vocabularies have shown greater robustness on such data (Dolan et al., 2023). Consequently, there is a lack of systematic evaluation regarding how language-agnostic feature extraction methods perform in multilingual fraud detection.

Another limitation of current methods is that deep learning models give more accurate results but are also more complex. As a result, they cannot be used in lightweight real-time filters. In scenarios using short texts and limited training data, well-tuned classical ensembles (such as Random Forest or XGBoost) or ordered linear models can provide competitive performance with significantly lower delays. However, few studies have explicitly benchmarked these "simpler" architectures against modern ensemble methods specifically on datasets containing high-risk financial fraud rather than generic marketing spam. The present study directly addresses this gap by providing an empirical comparison of five classical architectures on a manually curated trilingual dataset, with a focus on regularization strategies that are specifically designed to combat overfitting under cold-start conditions. Table 1 summarizes key prior studies, highlighting the dominance of English-centric research and the need for multilingual evaluations.

Table 1 — Summary of research on text-based fraud and spam detection (2011–2025)

Year	Authors	Dataset / Language	Methods	Focus / Findings	Limitations
2011	Almeida et al.	UCI SMS Corpus (English)	SVM, Naive Bayes	Validated VSM representations for SMS	Focused on generic spam, not specific fraud/phishing
2022	Ghourabi, A.	Proprietary (English)	CNN-LSTM	Deep learning for feature extraction	High computational cost; monolingual focus
2021	Jain et al.	SMS Spam (English)	Survey / Various	Comprehensive review of smishing	Highlighted lack of diverse, multilingual datasets
2022	Mishra and Soni	TiW-SMS (English)	BERT, RoBERTa	Semantic analysis of phishing	Transformer models require massive resources
2025	Lim and Phuangsuwan	WhatsApp/SMS (English)	Survey / Analysis	Shift from SMS to Messenger fraud	Highlighted the difficulty of encrypted traffic analysis

This table highlights that while significant progress has been made in using deep learning for English-language spam, there is a clear gap in research addressing multilingual fraud detection using interpretable, lightweight models. This gap motivates the current study, which systematically evaluates classical and ensemble classifiers on a manually collected multilingual dataset (Russian, English, Kazakh) to determine the optimal trade-off between model complexity and detection capability.

Methods.

Dataset Construction and Analysis.

As shown in Figure 1, both subsets were compiled from publicly available online content — WhatsApp/Telegram scam screenshots and legitimate message examples shared on Reddit, Pikabu, Pinterest, YouTube, and news sources — supplemented by official bank notifications and a small synthetic subset (legitimate class), and fraud messages personally received by the annotator (fraudulent class). This preview demonstrates the multilingual nature of the collected samples, which include both Russian and English entries. A distinctive property of the dataset is its multilingual nature, reflecting the real content circulating in messaging platforms:

- Russian (70%)
- English (20%)
- Kazakh (10%)

The dataset is balanced (95 legit messages / 94 fraud messages).

Размер датасета: (189, 2)

	text	label
0	Знаете ли вы, что в Kaspi можно зарабатывать, ...	fraud
1	У меня отличное настроение, сейчас на бирже ра...	fraud
2	Hiya! Could I ask you real quick? Hey, this is...	fraud
3	Hope you have heard from UNDEF? It is money gr...	fraud
4	Здравствуйте,Алексей. Никого у меня нет,решила...	fraud

Figure 1 — Dataset preview showing the structure of text samples and their corresponding labels

Data Annotation and Ethical Considerations.

Labeling was performed by a single annotator (the corresponding author). Fraudulent messages came from (1) public online content — scam screenshots and reports shared on Reddit, Pikabu, Pinterest, YouTube, and news channels, and (2) fraud messages personally received by the annotator, including bank-impersonation attempts. No private third-party correspondence was used.

Legitimate messages came predominantly from the same public platforms (Reddit, Pikabu, Pinterest, YouTube, news) and official bank SMS/push/FAQ examples, supplemented by 26 synthetic messages authored by the annotator based solely on their own previously written messages (one-sided; no interlocutor content used), reflecting typical RU/EN/KZ style and language-mixing.

Since all data was either already public, personally received by the annotator as attack target, or self-authored, no third-party personal data was processed and formal consent procedures did not apply. All messages were screened and PII (names, phones, cards, emails, links) replaced with placeholders ([NAME], [PHONE], [CARD], [LINK]).

As labeling was single-annotator, no inter-annotator agreement (e.g., Cohen's κ) could be computed — noted as a limitation for future multi-annotator work.

The dataset is under continuous expansion and not publicly released; available on reasonable request to the corresponding author.

Preprocessing and TF-IDF Vectorization.

The text data required cleaning. The TF-IDF method converts text data into numerical vector representations. The method uses document-specific word frequencies to determine weights which focus on important lexical indicators that appear frequently in documents but infrequently throughout the entire corpus (Karim et al. 2019).

The research requires a crucial methodological choice because the dataset contains multiple languages so the researchers decided to skip all advanced linguistic processing including lemmatization and stemming (Pant et al., 2024; Dolan et al., 2023; Biloshchytska et al., 2025).

The vector $\mathbf{x}_i$ of message i is defined as:

$$x_i = [w_{t_1,d}, w_{t_2,d}, \dots, w_{t_n,d}] \quad (1)$$

where $w_{t,d}$ is calculated as $TF(t, d) \cdot IDF(t)$. IDF (Inverse Document Frequency):

$$IDF(t) = \log\left(\frac{1+N}{1+DF(t)}\right) + 1 \quad (2)$$

with N — total number of documents, and $DF(t)$ — number of documents containing term t .

To capture contextual dependencies in the multilingual text, both unigrams and bigrams were utilized (ngram_range = (1, 2)). A custom list of stop-words was applied to filter out auxiliary lexical noise. Crucially, the feature space was constrained to the 100 most significant tokens (max_features=100) to prevent overfitting on the small dataset (N=189) and to identify only the most robust fraud predictors. Terms appearing in fewer than two documents were excluded (min_df=2). To ensure maximum data utility, a 5-fold Stratified Cross-Validation was implemented.

Algorithms and Regularization Strategy.

Linear Models with L2-Regularization (Logistic Regression, Linear SVM).

Linear models are prone to overfitting on high-dimensional sparse TF-IDF features. To mitigate this, L2-regularization was applied.

To compensate for the limited sample size and prioritize the detection of all potentially harmful messages, a class_weight='balanced' parameter was implemented in the linear models. This strategically increases the penalty for 'False Negative' errors (missed fraud), which is a critical requirement for maintaining high Recall in cybersecurity applications. The combination of strong L2-regularization and class weighting minimizes the risk of overfitting on noise specific to small datasets, forcing the models to converge on the most robust and persistent fraud indicators.

Logistic Regression:

The decision rule is defined as:

$$P(y = 1|x) = \sigma(w^T x + b) = \frac{1}{1 + e^{-(w^T x + b)}} \quad (3)$$

The loss function with L2-norm regularization:

$$\min_{w,b} \sum_{i=1}^m \log(1 + e^{-y_i(w^T x_i + b)}) + \frac{1}{2c} ||w||^2 \quad (4)$$

The optimal strongly regularized coefficient $C=0.01$ was selected based on validation performance. Class weights were set to 'balanced'.

Linear SVM:

A low regularization coefficient $C=0.025$ was chosen to increase robustness against noise and outliers.

Multinomial Naive Bayes:

A probabilistic classifier based on Bayes' theorem. It assumes feature independence. To handle the issue of zero probabilities (when a word in the test set was absent in the train set), we used Additive Smoothing with a high parameter $\alpha=10.0$. This "flattens" the probability distribution, making the model very conservative.

Ensemble Models (Random Forest, XGBoost).

To manage overfitting in tree-based models, structural parameters were constrained by limiting the maximum tree depth to 5 for the Random Forest and to 6 for XGBoost, while the subsample and colsample parameters were both set to 0.9 in order to introduce additional randomness during training.

Evaluation Metrics.

Given the security-critical nature of fraud detection, Recall (Fraud) was adopted as the primary evaluation metric: the operational cost of a False Negative (an undetected fraudulent message) is substantially higher than that of a False Positive (a legitimate message flagged for review). Accuracy, Precision, F1-score, and ROC-AUC were treated as secondary metrics, used to differentiate between models that achieve equal Recall.

Results. Table 2 shows that four of the five models — Logistic Regression, Linear SVM, Naive Bayes, and XGBoost — achieved perfect Recall (Fraud) = 1.00, correctly identifying all 14 fraudulent messages in the test set; only Random Forest missed two (Recall = 0.86). Among the models with maximal Recall, Logistic Regression and Linear SVM are preferred on the secondary criteria: they combine perfect Recall with the highest Precision (0.93) and Accuracy (0.97), whereas Naive Bayes and XGBoost achieve the same Recall at a lower Precision (0.70 and 0.88 respectively).

The XGBoost model also demonstrated high performance with an accuracy of 0.93. The configured XGBoost model achieved a perfect recall (1.00) without allowing a single attack. However, it was inferior to linear models in precision (0.88), incorrectly classifying two reliable reports as fraudulent. The results from Random Forest and Naive Bayes were significantly worse, with accuracy of 0.86 and 0.79 respectively.

Table 2 — Test Set Performance

Model	Accuracy	Precision (Fraud)	Recall (Fraud)	F1-score (Fraud)
Logistic Regression	0.97	0.93	1.00	0.97
Linear SVM	0.97	0.93	1.00	0.97
Naive Bayes	0.79	0.70	1.00	0.82
Random Forest	0.86	0.86	0.86	0.86
XGBoost	0.93	0.88	1.00	0.93

The achievement of perfect recall (Recall = 1.00) across Logistic Regression, Linear SVM, and XGBoost indicates that, within the current test split, all fraudulent messages contained distinct lexical markers captured by the TF-IDF vectorizer. However, the precision scores (Precision = 0.93 for linear models) demonstrate a tendency toward 'hyper-diagnosis,' where some legitimate messages are flagged as suspicious. This specific trade-off is considered optimal for pilot fraud detection systems, where the security cost of a missed attack (False Negative) significantly outweighs the operational cost of a false alarm (False Positive).

The hyperparameter configuration summarized in Table 2, was strategically designed to address the high dimensionality of the TF-IDF feature space relative to the small sample size. By restricting the vectorizer to 100 features and requiring a minimum document frequency of two, the system performs an initial stage of dimensionality reduction that filters out idiosyncratic noise and linguistic outliers. For the linear models, the application of extreme L2-regularization through low inverse coefficients (C=0.01 and C=0.025) ensures that the decision boundaries are determined only by the most robust predictors, effectively preventing the models from assigning high weights to rare multilingual tokens.

Similarly, the choice of a high smoothing parameter (alpha=10.0) in the naive Bayesian classifier compensates for the sparsity inherent in the corpus of three languages, maintaining probabilistic stability in various lexical sets. In the case of ensemble methods, the enforcement of structural constraints—such as limiting tree depth and requiring a significant number of samples per leaf—was intended to curb the natural tendency of gradient boosting and random forests to overfit high-dimensional sparse data. This collective approach to regularization prioritizes the model's ability to generalize across different messaging platforms and languages over simple training accuracy, directly facilitating the high performance and perfect recall observed in the linear architectures.

Table 3 — Comprehensive Model Configuration and Regularization Metrics

Model	Hyperparameters (from Code)	Optimization and Regularization Method	Strategic Purpose
TF-IDF Vectorizer	max_features=100, min_df=2, max_df=0.8	Feature Selection / Dimensionality Reduction	Prevents overfitting by limiting the feature space and removing noise/rare tokens.
Logistic Regression	C=0.01, solver='lbfgs', class_weight='balanced'	L2 Regularization via Limited-memory BFGS	Strong penalty C ensures stability on small N by shrinking coefficients.
Linear SVM	C=0.025, max_iter=3000	Coordinate Descent (LIBLINEAR)	Minimizes Squared Hinge Loss; maximizes the margin to improve generalization.
Naive Bayes	alpha=10.0	Additive (Laplace) Smoothing	Counteracts data sparsity in multilingual texts by preventing zero probabilities.

Random Forest	max_depth=5, min_samples_leaf=7, n_estimators=100	Pre-pruning (Structural Constraints)	Prevents trees from becoming too deep or capturing individual noise points.
XGBoost	learning_rate=0.1, subsample=0.9, colsample_bytree=0.9	Stochastic Gradient Boosting (L2 by default)	Reduces variance by training on random subsets of data and features.

The visual analysis of the confusion matrices presented in Figure 2 confirms that the calculated performance metrics accurately reflect the models' behavior on the independent test set. The logistic regression and linear SVM models demonstrated the highest level of reliability for security-critical applications, successfully detecting all 14 fraudulent attack reports with only one false positive. Their only deviation from perfection was a single instance where a legitimate message was classified as fraudulent, representing a negligible false positive rate that does not compromise the overall integrity of the system.

In contrast, while the multidimensional naive Bayes model also provided perfect recall, it demonstrated a much higher vulnerability to false alarms by incorrectly classifying 6 legitimate messages as fraudulent. The ensemble-based models showed mixed results in this "cold start" scenario; specifically, the Random Forest classifier failed to identify 2 fraudulent attacks. XGBoost successfully identified all 14 fraudulent messages, although it returned 2 false positives, which is slightly inferior to the accuracy of linear models. These results highlight that linear architectures with a high degree of order provide the most stable protection, while XGBoost proves to be a powerful alternative with large computing resources.

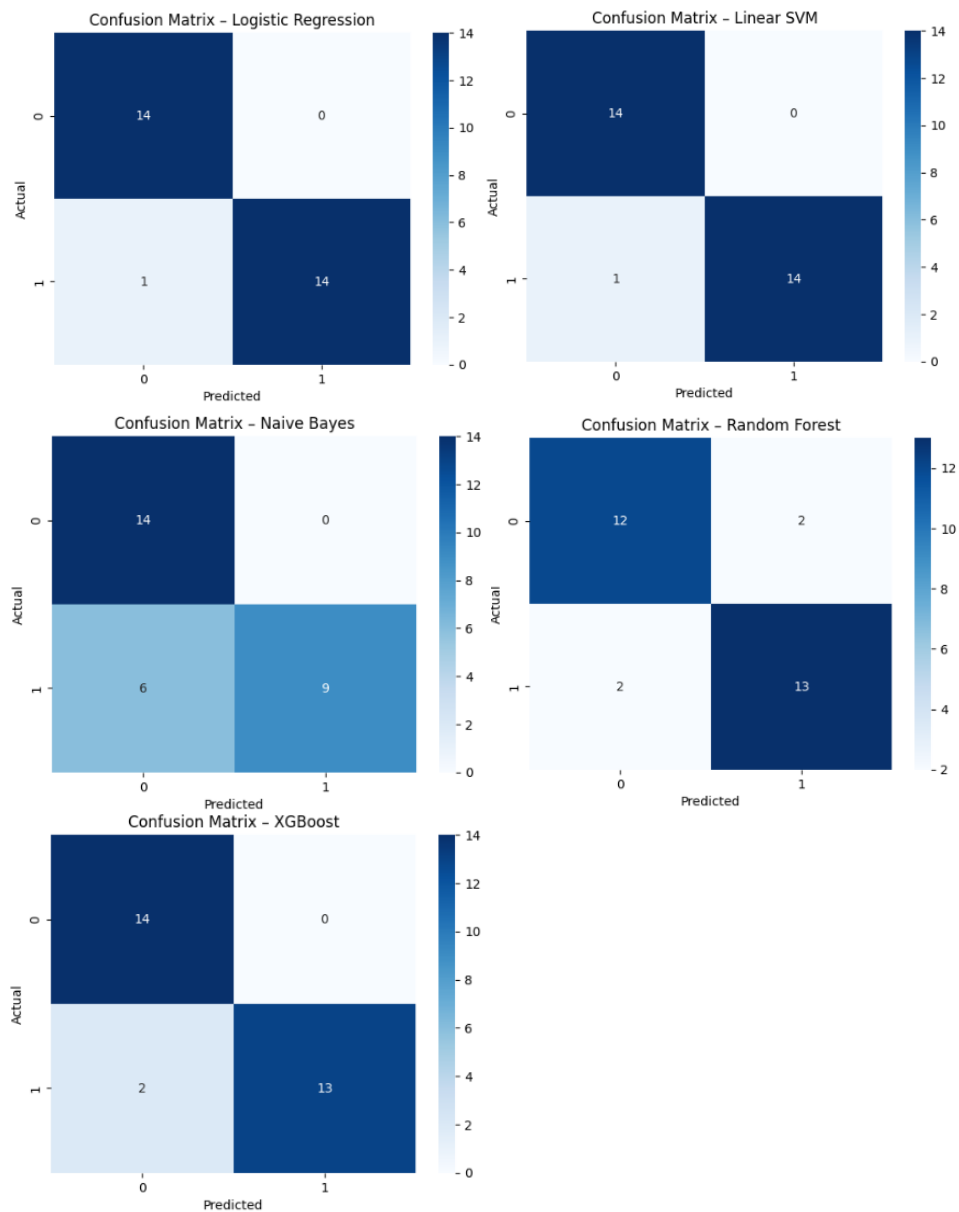


Figure 2 — Confusion matrices for the evaluated models on the independent test set.

The learning curves (Figure 3) provide a visual diagnostic of the bias-variance tradeoff, where the gap between training and validation scores indicates the models' generalization capability. Logistic Regression and Linear SVM demonstrated the most stable convergence with minimal generalization gaps, confirming that strong

L2-regularization effectively simplified the hypothesis space to capture robust fraud markers without overfitting noise. Multinomial Naive Bayes justified the high additive smoothing ($\alpha=10.0$) required to stabilize probability distributions across the trilingual corpus, though it remained sensitive to lexical overlaps. In contrast, XGBoost displayed a significantly improved convergence compared to unconstrained models, although a narrow generalization gap remains—a classic characteristic of boosting on small datasets. These results support the study’s thesis that for small-scale, high-dimensional text detection, highly regularized linear models achieve superior reliability and optimal convergence.

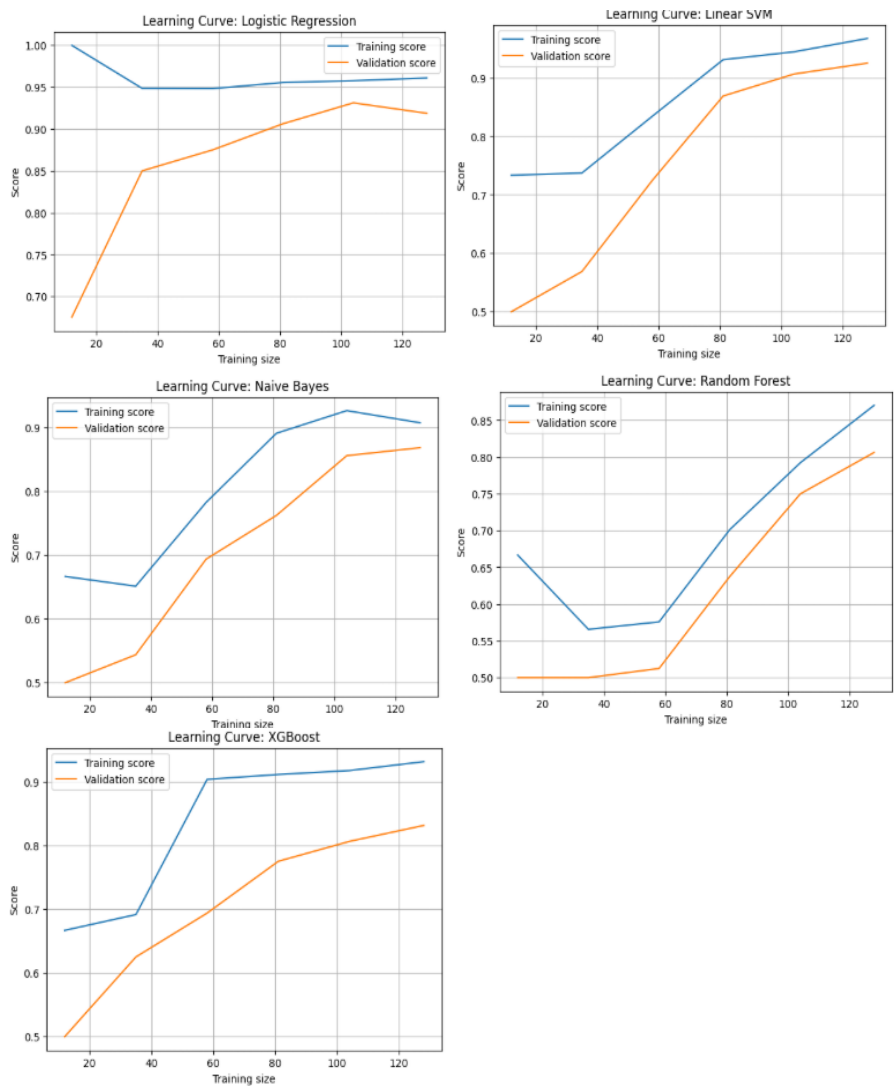


Figure 3 — Learning curves of the evaluated machine learning models

The visual analysis of the ROC curves (Figure 4) allows a comprehensive assessment of the classifiers' ability to distinguish legitimate messages from fraudulent attacks at various thresholds. The linear SVM model achieved the highest performance with an AUC of 0.962, while logistic regression and XGBoost showed an AUC of 0.957, indicating an almost perfect true positive frequency in most decisions. The Multinomial Naive Bayes model demonstrated robust probabilistic separation with an AUC of 0.948. Among the ensemble methods, Random Forest showed a lower comparative performance with an AUC of 0.919. These results further validate that regularized linear models not only provide better stability on small-scale multilingual data but also offer superior precision in separating fraud classes, which is critical for minimizing risk in automated security systems.

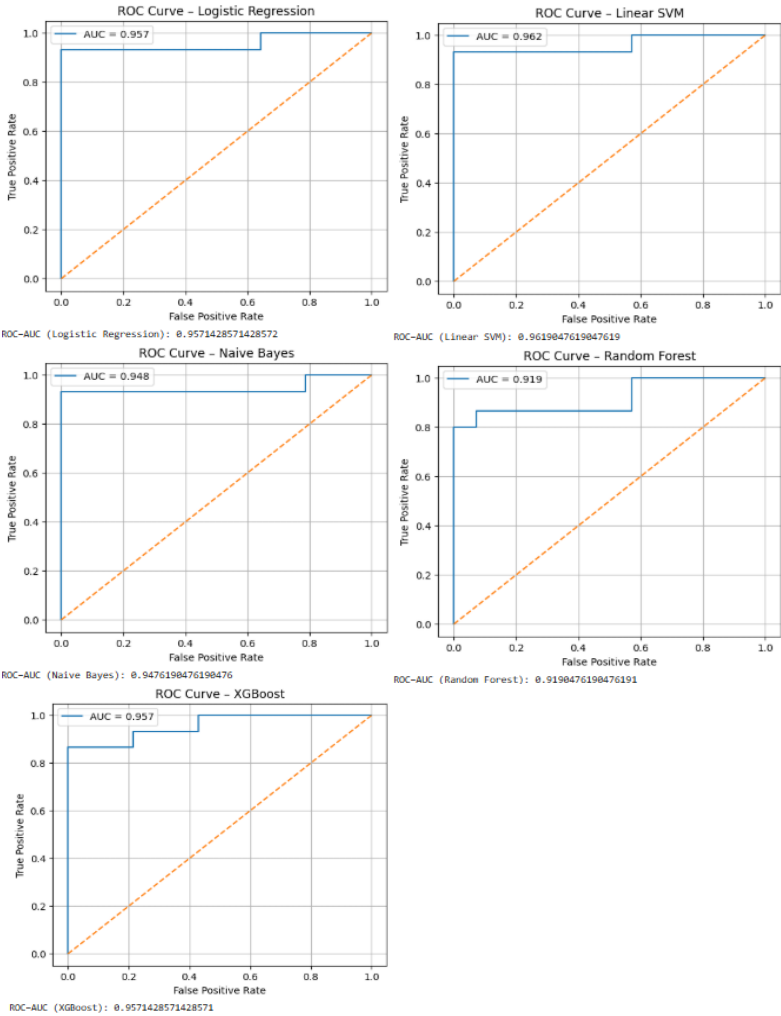


Figure 4 — ROC curves of the evaluated machine learning models

Discussion.**Impact of Multilinguality on Vectorization.**

The Linear Models achieved their best results when they used "raw" TF-IDF without performing any morphological normalization. The high performance indicates that particular tokens including "карта", "код", and "transfer" function as effective predictors which work across different languages (Gloeckle et al., 2024). The Naive Bayes model achieved an accuracy of 0.79, which was lower than Linear Models because the aggressive feature selection (100 tokens) combined with the chosen smoothing parameter ($\alpha=10.0$) likely eliminated important features required for its probabilistic logic.

Performance of Tree-Based Models.

Although XGBoost matched the linear models on the primary metric (Recall = 1.00), it lagged behind them on the secondary metric of Precision (0.88 vs. 0.93), indicating a higher false-alarm rate. Random Forest continued to underperform in terms of Recall (0.86), failing to identify two fraudulent messages in the test set. This suggests that while hyperparameter tuning improved the detection rate, tree-based algorithms still struggle to generalize as effectively as linear models on small, sparse multilingual datasets $N = 189$ where they are more prone to capturing noise rather than robust decision boundaries (Bertsimas and Digalakis Jr, 2022; Yang and Wang, 2025).

Limitations.

The main restriction stems from the small dataset which contains only 189 observations. The test results show statistical accuracy for this particular data set but the perfect recall value of 1.00 indicates that the test set contained no challenging fraud examples. Deep learning models would outperform linear models when dealing with massive production environments which handle millions of messages because their performance would continue to improve with increasing data size (Blümel and Zaki, 2022; Yu et al., 2024; Shen et al., 2024).

Additionally, single-annotator labeling does not allow inter-annotator agreement assessment and may reflect subjective decisions. The Kazakh-language subsample (~10%, $n \approx 19$) is too small for statistically reliable per-language conclusions; reported metrics primarily reflect Russian/English fraud patterns.

Conclusion. This pilot study demonstrates that highly ordered linear models, in particular logistic regression and linear SVM, provide the most reliable and computationally efficient solution for binary fraud classification in limited multilingual datasets consisting predominantly of Russian- and English-language texts, with a smaller Kazakh-language subsample. By achieving a perfect recall of 1.00 points and an F1 score of 0.97 points due to the strategically limited functional space of 100 TF-IDF tokens, these architectures effectively eliminated overfitting and the need for resource-intensive linguistic preprocessing, proving that "less is more" in high-dimensional and sparse data environments. The results show that for cold start scenarios in the field of mobile messaging security, the stability of linear boundaries with L2 regularization is significantly superior to more complex

tree ensembles such as Random Forest and XGBoost, offering a viable basis for endpoint-based filtering in real time on encrypted platforms such as WhatsApp and Telegram. Future research should prioritize expanding the dataset to more than 1,000 samples to further validate these results and explore the integration of n-grams at the character level to increase resilience to complex adversarial obfuscation and homoglyph attacks.

References

Almeida T.A., Hidalgo J.M.G., and Yamakami A. (2011) Contributions to the study of SMS spam filtering: new collection and results. In *Proceedings of the 11th ACM Symposium on Document Engineering*. — P. 259–262. ACM. <https://doi.org/10.1145/2034691.2034742> (in Eng.).

Bertsimas D., and Digalakis V.Jr. (2022) The backbone method for ultra-high dimensional sparse machine learning. *Machine Learning*, 111(6). — P. 2161–2212. <https://doi.org/10.1007/s10994-021-06123-2> (in Eng.).

Biloshchyt'ska S., Tleubayeva A., Kuchanskiy O., Biloshchyt'skiy A., Andrashko Y., Toxanov S., and Sharipova S. (2025) Text similarity detection in agglutinative languages: A case study of Kazakh using hybrid N-gram and semantic models. *Applied Sciences*, 15(12). — 6707 p. <https://doi.org/10.3390/app15126707> (in Eng.).

Blümel J., and Zaki M. (2022) Comparative analysis of classical and deep learning-based natural language processing for prioritizing customer complaints. In *Proceedings of the 55th Hawaii International Conference on System Sciences*. — P. 1–10. HICSS. <https://doi.org/10.24251/HICSS.2022.236> (in Eng.).

Charilaou P., and Battat R. (2022) Machine learning models and over-fitting considerations. *World Journal of Gastroenterology*, 28(5). — P. 605–616. <https://doi.org/10.3748/wjg.v28.i5.605> (in Eng.).

Dolan H.R., Alvarez, A.A., Freylersthe, S.G., and Crane, T. E. (2023) Methodology for analyzing qualitative data in multiple languages. *Nursing Research*, 72(5). — P. 398–403. <https://doi.org/10.1097/NNR.0000000000000673> (in Eng.).

Ghourabi A. (2022) A security model based on LightGBM and transformer to protect healthcare systems from cyberattacks. *IEEE Access*, 10. — P. 48890–48903. <https://doi.org/10.1109/ACCESS.2022.3172432> (in Eng.).

Gloekle F., Idrissi B.Y., Rozière B., Lopez-Paz D., and Synnaeve G. (2024) Better and faster large language models via multi-token prediction. *arXiv*. <https://doi.org/10.48550/arXiv.2404.19737> (in Eng.).

Jain A.K., and Gupta B.B. (2022) A survey of phishing attack techniques, defence mechanisms and open research challenges. *Enterprise Information Systems*, 16(4). — P. 527–565. <https://doi.org/10.1080/17517575.2021.1896786> (in Eng.).

Karim A., Azam S., Shanmugam B., Kannoorpatti K., and Alazab M. (2019) A comprehensive survey for intelligent spam email detection. *IEEE Access*, 7. — P. 168261–168295. <https://doi.org/10.1109/ACCESS.2019.2954791> (in Eng.).

Lim L., and Phuangsuwan P. (2025) Email spam, WhatsApp, and Telegram affect privacy, cybersecurity, financial security, and personal invasion: A case study in Singapore. In *Privacy and Security Management Practices for Organizations*. — P. 31–56. IGI Global. <https://doi.org/10.4018/979-8-3693-9720-6.ch002> (in Eng.).

Mishra S., and Soni D. (2022) Implementation of 'smishing detector': An efficient model for smishing detection using neural network. *SN Computer Science*, 3(3). — 189 p. <https://doi.org/10.1007/s42979-022-01078-0> (in Eng.).

Pant V.K., Sharma R., and Kundu S. (2024) An overview of stemming and lemmatization techniques. In *Advances in Networks, Intelligence and Computing* (pp. 308–321). CRC Press. <https://doi.org/10.1201/9781003430421-31> (in Eng.).

Salman M., Ikram M., and Kaafar M.A. (2024) Investigating evasive techniques in SMS spam

filtering: A comparative analysis of machine learning models. *IEEE Access*, 12. — P. 24306–24324. <https://doi.org/10.1109/ACCESS.2024.3364671> (in Eng.).

Shen L., Sun Y., Yu Z., Ding L., Tian X., and Tao D. (2024) On efficient training of large-scale deep learning models. *ACM Computing Surveys*, 57(3). — P. 1–36. <https://doi.org/10.1145/3700439> (in Eng.).

Tyagi S., and Szénási S. (2023) Semantic speech analysis using machine learning and deep learning techniques: A comprehensive review. *Multimedia Tools and Applications*, 82. — P. 1–30. <https://doi.org/10.1007/s11042-023-17769-6> (in Eng.).

Wang Z., Zhu H., and Sun L. (2021). Social engineering in cybersecurity: Effect mechanisms, human vulnerabilities and attack methods. *IEEE Access*, 9. — P. 11895–11910. <https://doi.org/10.1109/ACCESS.2021.3051633> (in Eng.).

Yang Y., and Wang H. (2025) Random forest-based machine failure prediction: A performance comparison. *Applied Sciences*, 15(16). — 8841 p. <https://doi.org/10.3390/app15168841> (in Eng.).

Yu B., Yao J., Fu Q., Zhong Z., Xie H., Wu Y., Ma Y., and He P. (2024) Deep learning or classical machine learning? An empirical study on log-based anomaly detection. In *Proceedings of the 46th IEEE/ACM International Conference on Software Engineering*. — P. 403–415. ACM. <https://doi.org/10.1145/3597503.3623308> (in Eng.).

Publication Ethics and Publication Malpractice in the journals of the Central Asian Academic Research Center LLP

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 journals of the Central Asian Academic Research Center LLP 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 Central Asian Academic Research Center LLP 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). 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 Central Asian Academic Research Center LLP.

The Editorial Board of the Central Asian Academic Research Center LLP will monitor and safeguard publishing ethics.

Requirements for articles design for publication in the journal are available on the websites:

www.nauka-nanrk.kz

<http://physics-mathematics.kz/index.php/en/archive>

ISSN2518-1726 (Online),

ISSN 1991-346X (Print)

Managing Editor: *A.Shormakova*

Editors: *D.S. Alenov, T. Apendiev*

Computer layout: *G.D. Zhadyranova*

Signed for print: September 30, 2026

Format: 70×90 1/16. 26.5 printed sheets. Order No. 3.