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

ACADEMIC SCIENTIFIC JOURNAL OF COMPUTER SCIENCE

№2
2026

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

CENTRAL ASIAN ACADEMIC
RESEARCH CENTER

ACADEMIC SCIENTIFIC JOURNAL OF COMPUTER SCIENCE

2 (358)

APRIL – JUNE 2026

PUBLISHED SINCE JANUARY 1963
PUBLISHED 4 TIMES A YEAR

ALMATY, NAS RK

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/

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/

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 КОКСНВО МНВО РК по направлению «информационно-коммуникационные технологии».*

Периодичность: *4 раза в год.*

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

CONTENTS

COMPUTER SCIENCE

Abduraimova B.K., Toleukhan A.B., Sapakova S.Z., Abisheva A.A.
Development of early cyberattack detection method using CNN-LSTM for IoT..............11

Aben A.B., Kazbekova G.N., Baimakhanova A.S., Amanzholova A.B.
Classification of birds and drones in the sky using MobileNetV2 model………………......30

Akbarov D., Sembayev T.
Quality-aware pose–hand keypoint extraction pipeline for skeleton-based sign language recognition...44

Algazy K., Alimzhan Y., Sakan K., Nyssanbayeva S.
Lattice-based vector commitments for Verkle trees..67

Asylkhan N., Baidrakhmanova M.G.
Principles and models of spatial organization of buildings for crop production considering technological and climatic factors...87

Basheyeva Zh., Tokesh A., Bekish U., Abdoldinova G.
Artificial intelligence for academic project management: a bibliometric analysis and systematic review...105

Bekmanova G., Kantureyeva M., Omarbekova A., Zakirova A., Issainova A.
Integrating artificial intelligence to evaluate emotions in the learning environment........125

Dzhusupbekova G.T., Jangassiyev R.M.
Gemini AI integration based on .NET MAUI for education: hybrid architecture and empirical load testing..146

Doszhan N.S., Sultanbekova L.Ye., Zhumagali S.Zh., Konysbayev E.K.
Modeling and parameter calculation of an emergency response system based on LoRaWAN technology in the high-altitude conditions of the Zailiysky Alatau...............166

Zhumakhanova A., Kudabayeva R., Akanova K., Myrkanova A.
Entropy-normalized multidimensional model for user activity segmentation in Reddit...180

Karabaliyev Y., Kolesnikova K., Khlevnaya Y.
HybridKazASR: a hybrid automatic speech recognition system combining multi-model rover fusion and morpheme-aware language modeling for Kazakh................................198

Kerimkhulle S.E., Adalbek A., Baizakov N.A., Shodorova N.N.
Piecewise logistic and fuzzy modeling of Kazakhstan's GDP dynamics (1990–2024)....212

Kulakayeva A., Ashurov A., Aitmagambetov A., Ongenbayeva Zh.
Development of mathematical models and criteria for the admissibility of orbital maneuvers of spacecraft...228

Kulatay A.A., Zhaisanova D.S., Daurenbayeva N.A., Mamanova S.Y., Tolegen M.
Machine learning for personalized learning in gamified edtech platforms:
Aqyl Battle case..248

Mamyrbayev O., Kurmetkan T.
Enhanced sentiment analysis of e-commerce product reviews using
Luong attention-based Bi-LSTM...263

Marassulov U.A., Kazbekova G.
TF-IDF-based fake news detection in Kazakh and Russian..286

Omar A.B., Mussiraliyeva Sh.Zh.
Federated learning: models based on transformer architecture..302

Rakhimova D., Duisenbekkyzy Zh., Karibayeva A., Eşref A., Ilessova B.
Improving the voice recognition system for children in Kazakh through additional
training (fine-tuning)..317

Sarsembayev M, Urmashev B.
Optimization of the calculation of kinetic equations of combustion processes on GPU
using global memory and shared memory...335

Symagulov A., Smurygin V., Beloussov A., Karypov A., Yunicheva N.R.
Improving the accuracy of crop and weed detection using UAVs in soya fields
through image segmentation...347

Tashenova Zh., Gabdullin A.R., Abdugulova Zh., Amanzholova Sh., Santeyeva S.
Security evaluation of WPA3 wireless networks under deauthentication
attack scenarios...368

Tursunbayeva G.U., Satybaldina D.Zh., Tleuberdin S.T., Tashatov N.N., Egamberdiyev E.E.
Anomaly detection in UAV telemetry systems based on simulation modeling................391

Tursynova N., Yerimbetova A., Amangeldy N., Zhumabayeva A., Daiyrbayeva E.
Comparative analysis of multilingual transformer models for Kazakh-to-gloss
translation...414

Shangpeng Lei, Balakayeva G.
Dual-branch physical information neural networks for data center airflow velocity
and thermal modeling……...…433

Shynzhigit B.B., Balabekova M.O., Amangeldy T.T., Malik G.J., Balgimbekova U.B.
Automatic brick defects detection by using a CNN-based deep learning model.............449

МАЗМҰНЫ

КОМПЬЮТЕРЛІК ҒЫЛЫМДАР

Абдураимова Б.К., Төлеухан Ә.Б., Сапакова С.З., Абишева А.А.
Кибершабулдарды ерте анықтау әдісін CNN-LSTM негізінде дамыту (IOT үшін).....11

Абен А.Б., Казбекова Г.Н., Баймаханова А.С., Аманжолова Ә.Б.
MobileNetV2 моделімен аспандағы құстар мен дрондарды классификациялау..........30

Акбаров Д.Р., Сембаев Т.М.
Ым тілін тануға арналған дене қалпы мен қолдың негізгі нүктелерін сапаны бақылаумен анықтау әдісі...44

Алғазы К.Т., Әлімжан Е.Ж., Сақан Қ.С., Нысанбаева С.Е.
Verkle ағаштарына арналған торлық векторлық міндеттемелер..................................67

Асылхан Н., Байдрахманова М.Г.
Технологиялық және климаттық факторларды ескере отырып, өсімдік шаруашылығы ғимараттарының кеңістік ұйымдастыру қағидалары мен модельдері...87

Башеева Ж., Төкеш Ә., Бекіш Ұ., Абдолдинова Г.
Академиялық жобаларды басқарудағы жасанды интеллект: библиометриялық талдау және жүйелі шолу...105

Бекманова Г.Т., Кантуреева М.А., Омарбекова А.С., Закирова А.Б., Исайнова А.Н.
Оқу ортасындағы эмоцияларды бағалау үшін жасанды интеллектті біріктіру..........125

Джусупбекова Г.Т., Джангасиев Р.М.
Білім беруге арналған .NET MAUI негізіндегі Gemini AI интеграциясы: гибридті архитектура және эмпирикалық жүктемелік тестілеу...146

Досжан Н.С., Султанбекова Л.Е., Жумагали С.Ж., Конысбаев Е.К.
Іле Алатауының биік таулы жағдайында LORAWAN технологиясы негізіндегі жедел әрекет ету жүйесінің параметрлерін модельдеу және есептеу.........................166

Жумаханова А., Кудабаева Р., Аканова К., Мырканова А.
REDDIT-те пайдаланушы әрекетін сегменттеуге арналған энтропия-нормалданған көп өлшемді модель..180

Карабалиев Е., Колесникова К., Хлевная Ю.
HybridKazASR: Rover көпмодельді біріктіру және морфемеге негізделген тілдік модельдеуді пайдаланатын қазақ тілін автоматты тану гибридті жүйесі...................198

Керімқұл С.Е., Адалбек А., Байзаков Н.А., Шодорова Н.Н.
Қазақстан ЖІӨ динамикасын кезеңдік (Piecewise) логистикалық және бұлдыр модельдеу (1990–2024)..212

Кулакаева А.Е., Ашуров А.Е., Айтмағамбетов А.З., Онгенбаева Ж.Ж.
Ғарыш аппараттарының орбиталық маневрлерінің математикалық модельдері мен рұқсат критерийлерін әзірлеу...228

Құлатай А.А., Жайсанова Д.С., Дауренбаева Н.А., Маманова С.Е., Төлеген М.
Геймификацияланған edtech платформаларда оқытуды жекелендіруге арналған машиналық...248

Мамырбаев Ө.Ж., Құрметқан Т.
Луонг назар механизміне негізделген BI-LSTM көмегімен электрондық коммерция өнімдеріне жазылған пікірлерге жетілдірілген сентименттік талдау жасау.............263

Марасулов У.А., Казбекова Г.
Қазақ және орыс тілдеріндегі жалған жаңалықтарды TF-IDF арқылы анықтау.......286

Омар А.Б., Мусиралиева Ш.Ж.
Федеративті оқыту: трансформер архитектурасына негізделген модельдер............302

Рахимова Д., Дүйсенбекқызы Ж., Кәрібаева А., Eşref A., Ілесова Б.
Қазақ тіліндегі балалар дауысын тану жүйесін қосымша оқыту (Fine-Tuning) арқылы жетілдіру...317

Сарсембаев М., Урмашев Б.
Global memory және shared memory қолдану арқылы GPU-да жану процестерінің кинетикалық теңдеулерін есептеуді оңтайландыру...335

Сымагулов А., Смурыгин В., Белоусов А., Карыпов А., Юничева Н.Р.
Соя алқаптарында ҰҰА көмегімен мәдени және арамшөп өсімдіктерін детекттеу сапасын кескіндерді сегменттеу арқылы арттыру..347

Ташенова Ж.М., Габдуллин А.Р., Абдугулова Ж.К., Аманжолова Ш.А., Сантеева С.Ә.
Деатентификациялау шабуылы сценарийлеріндегі WPA3 сымсыз желілерінің қауіпсіздігін бағалау...368

Турсунбаева Г., Сатыбалдина Д., Тлеубердин С., Ташатов Н., Эгамбердиев Э.
Симуляциялық модельдеу негізінде ұшқышсыз ұшу аппараттарының телеметриялық жүйелеріндегі аномалияларды анықтау..391

Турсынова Н., Еримбетова А., Амангелді Н., Жумабаева А., Дайырбаева Э.
Қазақ тілінен глоссқа аудару үшін көптілді трансформерлік модельдердің салыстырмалы талдауы...414

Шанпэн Лей, Балакаева Г.
Деректер орталығының ауа ағынының жылдамдығына және термиялық модельдеуге арналған екі тармақты физикалық ақпараттық нейрондық желілер...433

Шынжігіт Ш.Б., Балабекова М.О., Амангелді Т.Т., Мәлік Г.Ж., Балгимбекова У.Б. Кірпіш ақауларын автоматты анықтауда cnn негізіндегі терең оқыту моделін пайдалану...449

СОДЕРЖАНИЕ

КОМПЬЮТЕРНЫЕ НАУКИ

Абдураимова Б.К., Толеухан А.Б., Сапакова С.З., Абишева А.А.
Разработка метода раннего обнаружения кибератак на основе CNN-LSTM для IoT.....11

Абен А.Б., Казбекова Г.Н., Баймаханова А.С., Аманжолова А.Б.
Классификация птиц и дронов в небе с использованием модели MobileNetV2..........30

Акбаров Д.Р., Сембаев Т.М.
Метод получения ключевых точек позы и кистей с контролем качества для распознавания жестового языка..44

Алгазы К.Т., Алимжан Е.Ж., Сакан К.С., Нысанбаева С.Е.
Решеточные векторные обязательства для Verkle-деревьев...67

Асылхан Н., Байдрахманова М.Г.
Принципы и модели пространственной организации зданий для растениеводства с учетом технологических и климатических факторов...87

Башеева Ж., Токеш А., Бекиш У., Абдолдинова Г.
Искусственный интеллект в управлении академическими проектами: библиометрический анализ и систематический обзор..105

Бекманова Г.Т., Кантуреева М.А., Омарбекова А.С., Закирова А.Б., Исайнова А.Н.
Интеграция искусственного интеллекта для оценки эмоций в учебной среде..........125

Джусупбекова Г.Т., Джангасиев Р.М.
Интеграция Gemini AI на базе .NET MAUI для образования: гибридная архитектура и эмпирическое нагрузочное тестирование..146

Досжан Н.С., Султанбекова Л.Е., Жумагали С.Ж., Конысбаев Е.К.
Моделирование и расчет параметров системы экстренного реагирования на базе технологии LoRaWAN в условиях высокогорья Заилийского Алатау........................166

Жумаханова А., Кудабаева Р., Аканова К., Мырканова А.
Энтропийно-нормализованная многомерная модель для сегментации активности пользователей в Reddit..180

Карабалиев Е., Колесникова К., Хлевна Ю.
HybridKazASR: гибридная система автоматического распознавания казахской речи на основе многомодельного объединения ROVER и морфемно-ориентированного языкового моделирования..198

Керимкулов С.Е., Адалбек А., Байзаков Н.А., Шодорова Н.Н.
Кусочно-логистическое и нечеткое моделирование динамики ВВП Казахстана (1990–2024)...212

Кулакаева А.Е., Ашуров А.Е., Айтмагамбетов А.З., Онгенбаева Ж.Ж.
Разработка математических моделей и критериев допустимости орбитальных маневров космических аппаратов...228

Кулатай А.А., Жайсанова Д.С., Дауренбаева Н.А., Маманова С.Е., Толеген М.
Машинное обучение для персонализации обучения на геймифицированных EdTech-платформах: кейс Aqyl Battle..248

Мамырбаев О., Курметкан Т.
Усовершенствованный анализ тональности отзывов о товарах электронной коммерции с использованием Bi-LSTM на основе механизма внимания Луонга........................263

Марасулов У.А., Казбекова Г.
Выявление ложных новостей на казахском и русском языках TF-IDF-моделями.....286

Омар А.Б., Мусиралиева Ш.Ж.
Федеративное обучение: модели на основе архитектуры трансформеров...............302

Рахимова Д., Дуйсенбеккызы Ж., Карибаева А., Eşref A., Илесова Б.
Совершенствование системы распознавания голоса детей на казахском языке путем дополнительного обучения (fine-tuning)..317

Сарсембаев М., Урмашев Б.
Оптимизация расчета кинетических уравнений процессов горения на GPU с использованием global memory и shared memory..335

Сымагулов А., Смурыгин В., Белоусов А., Карыпов А., Юничева Н.Р.
Улучшение качества детектирования культурных и сорных растений с помощью БПЛА на полях сои с применением сегментации изображений...............................347

Ташенова Ж.М., Габдуллин А.Р., Абдугулова Ж.К., Аманжолова Ш.А., Сантеева С.А.
Оценка безопасности беспроводных сетей WPA3 в условиях атаки с деаутентификацией...368

Турсунбаева Г., Сатыбалдина Д., Тлеубердин С., Ташатов Н., Эгамбердиев Э.
Обнаружение аномалий в телеметрических системах БПЛА на основе симуляционного моделирования...391

Турсынова Н., Еримбетова А., Амангелді Н., Жумабаева А., Дайырбаева Э.
Сравнительный анализ многоязычных трансформерных моделей для перевода с казахского языка на глоссированное представление..414

Шанпэн Лэй, Балакаева Г.
Двухветвевые физически информированные нейронные сети для моделирования воздушных потоков и тепловых условий в центрах обработки данных...................433

Шынжыгит Ш.Б., Балабекова М.О., Амангелды Т.Т., Малик Г.Ж., Балгимбекова У.Б.
Использование модели глубокого обучения на основе CNN для автоматического обнаружения дефектов кирпичной кладки...449

ACADEMIC SCIENTIFIC JOURNAL OF COMPUTER SCIENCE
ISSN 1991-346X
Volume 2.
Namber 358 (2026). 198–211

https://doi.org/10.32014/2026.2518-1726.434

IRSTI 20.53.19
UDC 004.934.2

© Karabaliyev Y. [1]*, Kolesnikova K. [1], Khlevnaya Y.[2], 2026.
[1]International Information Technology University, Almaty, Kazakhstan;
[2]Taras Shevchenko National University of Kyiv, Ukraine.
E-mail: y.karabaliyev@iitu.edu.kz

HYBRIDKAZASR: A HYBRID AUTOMATIC SPEECH RECOGNITION SYSTEM COMBINING MULTI-MODEL ROVER FUSION AND MORPHEME-AWARE LANGUAGE MODELING FOR KAZAKH

Karabaliyev Yerlan — Doctoral Student, International University of Information Technologies, Almaty, Kazakhstan,
E-mail: y.karabaliyev@iitu.edu.kz, https://orcid.org/0009-0001-9465-3998;
Kolesnikova Kateryna — Doctor of Technical Sciences, Professor, International University of Information Technology, Almaty, Kazakhstan,
E-mail: kkolesnikova@iitu.edu.kz, https://orcid.org/0000-0002-9160-5982;
Khlevna Iulia — Doctor of Technical Science, professor, Department of Technology Management, Taras Shevchenko National University of Kyiv, Kyiv, Ukraine,
E-mail: yuliia.khlevna@knu.ua, https://orcid.org/0000-0002-1874-1961.

Abstract. This paper presents HybridKazASR, a hybrid automatic speech recognition (ASR) system for the Kazakh language that integrates three complementary acoustic models within a multi-stage pipeline. The system combines NVIDIA FastConformer (115M parameters, RNNT decoder) with two domain-adapted variants of Meta MMS-1B (965M parameters, CTC decoder): one fine-tuned on FLEURS read speech and another on the 559-hour Kazakh Speech Dataset (KSD). A key technical innovation is an efficient adapter-swap mechanism that switches between MMS variants by replacing only 138,348 parameters (~542 KB) in the CTC output layer, avoiding the need to load two full 3.9 GB models. The three hypotheses are fused via weighted 3-Way ROVER (Recognizer Output Voting Error Reduction) using Needleman-Wunsch alignment. The fused output undergoes two-stage language model rescoring: first with a 5-gram word-level KenLM, then with KazMorphLM – a novel morpheme-aware language model featuring a rule-based segmenter with 230 suffixes across 14 grammatical categories, a 7-gram morpheme model interpolated with a 5-gram word model at a 0.6/0.4 ratio using Witten-Bell smoothing, and a four-channel rescoring mechanism incorporating acoustic confidence, word-level LM, morpheme-level LM, and vowel harmony scores. Evaluated on the FLEURS test set (100 samples), HybridKazASR achieves a Word Error Rate (WER) of 6.30%, a normalized WER of 4.18%, and a Character Error Rate (CER) of 1.90%. Ablation studies confirm the contribution of each pipeline component, with KazMorphLM providing a 14.6% relative WER improvement over word-level KenLM rescoring alone.

Keywords: Automatic Speech Recognition; Kazakh Language; ROVER System Combination; Morpheme Language Model; Agglutinative Morphology; Vowel Harmony; Low-Resource ASR

For citations: Karabaliyev Y., Kolesnikova K., Khlevnaya Y. HybridKazASR: a hybrid automatic speech recognition system combining multi-model rover fusion and morpheme-aware language modeling for kazakh. Academic Scientific Journal of Computer Science, 2026. — No.2. – P. 198–211. DOI https://doi.org/10.32014/2026.2518-1726.434

© Карабалиев Е.[1]*, Колесникова К. [1]*, Хлевная Ю. [2], 2026.
[1] Халықаралық ақпараттық технологиялар университеті,
Алматы, Қазақстан;
[2] Тарас Шевченко атындағы Киев ұлттық университеті, Киев, Украина.
E-mail: y.karabaliyev@iitu.edu.kz

HYBRIDKAZASR: ROVER КӨПМОДЕЛЬДІ БІРІКТІРУ ЖӘНЕ МОРФЕМЕГЕ НЕГІЗДЕЛГЕН ТІЛДІК МОДЕЛЬДЕУДІ ПАЙДАЛАНАТЫН ҚАЗАҚ ТІЛІН АВТОМАТТЫ ТАНУ ГИБРИДТІ ЖҮЙЕСІ

Карабалиев Ерлан — докторант, Халықаралық ақпараттық технологиялар университеті, Алматы, Қазақстан,
E-mail: y.karabaliyev@iitu.edu.kz, https://orcid.org/0009-0001-9465-3998;
Колесникова Катерина — техника ғылымдарының докторы, профессор, Халықаралық ақпараттық технологиялар университеті, Алматы, Қазақстан,
E-mail: kkolesnikova@iitu.edu.kz, https://orcid.org/0000-0002-9160-5982;
Хлевна Юлия — техника ғылымдарының докторы, профессор, Технологияларды басқару кафедрасы, Тарас Шевченко атындағы Киев ұлттық университеті, Киев, Украина.
E-mail: yuliia.khlevna@knu.ua, https://orcid.org/0000-0002-1874-1961.

Аннотация. Бұл мақалада HybridKazASR – қазақ тіліне арналған гибридті автоматты сөйлеуді тану (ASR) жүйесі ұсынылған. Бұл жүйе көп сатылы конвейерде үш қосымша акустикалық модельді біріктіреді. Жүйе NVIDIA FastConformer (115 млн параметр, RNNT декодер) және Meta MMS-1B (965 млн параметр, CTC декодер) екі доменге бейімделген нұсқасын біріктіреді: біреуі FLEURS дайын сөйлеу корпусында, екіншісі 559 сағаттық қазақ тіліндегі сөйлеу деректер жинағында (KSD) өңделген. Негізгі техникалық жаңалық – тиімді адаптер ауыстыру механизмі. Бұл механизм MMS нұсқалары арасында CTC шығыс қабатында тек 138 348 параметрді (~542 КБ) ауыстыру арқылы жұмыс істейді, бұл әрқайсысы 3,9 ГБ болатын екі толық модельді жүктеу қажеттілігін болдырмайды. Үш гипотеза Needleman-Wunsch туралауын қолдана отырып, салмақталған 3-жақты ROVER (Recognizer Output Voting Error Reduction) арқылы біріктіріледі. Біріктірілген нәтиже екі сатылы тілдік модельді қайта бағалаудан өтеді: алдымен 5-граммалық сөздік KenLM көмегімен, содан кейін KazMorphLM көмегімен. KazMorphLM – морфемаға негізделген жаңа тілдік модель. Ол құрамында 230 жұрнағы бар ережеге негізделген сегментаторды (14 грамматикалық категорияны қамтиды), 0,6/0,4 қатынасында сөздік 5-граммалық модельмен интерполяцияланған 7-граммалық морфемалық модельді (Виттен-Белл тегістеуін қолданады), сондай-ақ төрт арналы қайта бағалау механизмін (акустикалық сенімділік, сөз деңгейіндегі LM, морфема деңгейіндегі LM және дауысты дыбыс үндестігі ұпайлары) қамтиды. FLEURS тест жинағында (100 үлгі) HybridKazASR қателіктердің сөздік көрсеткішіне (WER) 6,30%, нормализацияланған WER 4,18% және қателіктердің символдық көрсеткішіне (CER) 1,90% қол жеткізеді. Абляциялық зерттеулер конвейердің әрбір компонентінің үлесін растайды, ал KazMorphLM тек сөздік KenLM-мен қайта бағалаумен салыстырғанда WER көрсеткішін 14,6% - ға жақсартады.

Түйін сөздер: қазақ тілін тану; автоматты сөйлеу тану; гибридті ASR құбыры; морфемалық тілдік модель; ROVER біріктіру; FastConformer; MMS; адаптер-алмастыру; агглютинативті морфология

© Карабалиев Е.[1]*, Колесникова К. [1], Хлевна Ю.[2], 2026.
[1]Международный университет информационных технологий,
Алматы, Казахстан;
[2]Киевский национальный университет имени Тараса Шевченко,
Киев, Украина.
E-mail: y.karabaliyev@iitu.edu.kz

HYBRIDKAZASR: ГИБРИДНАЯ СИСТЕМА АВТОМАТИЧЕСКОГО РАСПОЗНАВАНИЯ КАЗАХСКОЙ РЕЧИ НА ОСНОВЕ МНОГОМОДЕЛЬНОГО ОБЪЕДИНЕНИЯ ROVER И МОРФЕМНО-ОРИЕНТИРОВАННОГО ЯЗЫКОВОГО МОДЕЛИРОВАНИЯ

Карабалиев Ерлан — докторант, Международный университет информационных технологий, Алматы, Казахстан,
E-mail: y.karabaliyev@iitu.edu.kz, https://orcid.org/0009-0001-9465-3998;
Колесникова Катерина — доктор технических наук, профессор, Международный университет информационных технологий, Алматы, Казахстан,
E-mail: kkolesnikova@iitu.edu.kz, https://orcid.org/0000-0002-9160-5982;

Хлевна Юлия — доктор технических наук, профессор, кафедра управления технологиями, Киевский национальный университет имени Тараса Шевченко, Киев, Украина,
E-mail: yuliia.khlevna@knu.ua, https://orcid.org/0000-0002-1874-1961.

Аннотация. *Актуальность.* В статье представлена HybridKazASR - гибридная система автоматического распознавания речи (ASR) для казахского языка, ориентированная на повышение качества распознавания за счет многоуровневого объединения акустических и языковых моделей. Актуальность исследования обусловлена необходимостью разработки эффективных ASR-решений для агглютинативных языков, включая казахский, где морфологическая сложность, вариативность словоформ и ограниченность специализированных речевых корпусов усложняют применение стандартных подходов автоматического распознавания речи. *Цель.* Разработать гибридную систему автоматического распознавания казахской речи на основе многомодельного объединения ROVER и морфемно-ориентированного языкового моделирования, обеспечивающую снижение ошибок распознавания и повышение устойчивости системы к морфологической вариативности казахского языка. *Методы.* Система HybridKazASR объединяет три взаимодополняющие акустические модели в многоэтапном конвейере. В качестве базовых компонентов использованы NVIDIA FastConformer с 115 млн параметров и RNNT-декодером, а также два доменно-адаптированных варианта Meta MMS-1B с 965 млн параметров и CTC-декодером. Один вариант MMS дообучен на подготовленной речи из корпуса FLEURS, второй - на 559-часовом казахском речевом датасете KSD. Ключевым техническим решением является механизм замены адаптеров (adapter-swap), позволяющий переключаться между вариантами MMS путем замены 138 348 параметров, или около 542 КБ, в выходном слое CTC без загрузки двух полных моделей объемом по 3,9 ГБ каждая. Три гипотезы объединяются с помощью взвешенного трехстороннего ROVER (Recognizer Output Voting Error Reduction) с использованием алгоритма выравнивания Нидлмана - Вунша. *Результаты и выводы.* Объединенный выход проходит двухэтапное переоценивание с применением языковых моделей: сначала с использованием 5-граммной word-level KenLM, затем с помощью KazMorphLM - новой морфемно-ориентированной языковой модели. KazMorphLM включает основанный на правилах сегментатор с 230 суффиксами, охватывающими 14 грамматических категорий, 7-граммную морфемную модель, интерполированную с 5-граммной словесной моделью в соотношении 0,6/0,4 с применением сглаживания Виттена - Белла, а также четырехканальный механизм переоценивания, учитывающий акустическую уверенность, языковую модель на уровне слов, языковую модель на уровне морфем и оценки сингармонизма гласных. При оценке на тестовом наборе FLEURS, включающем 100 образцов, HybridKazASR достигает частоты ошибок на слова (WER) 6,30%, нормализованной WER - 4,18% и частоты ошибок на символы (CER) - 1,90%. Абляционные исследования подтверждают вклад каждого компонента конвейера, при этом KazMorphLM обеспечивает относительное улучшение WER на 14,6% по сравнению с переоцениванием только на основе word-level KenLM. Полученные результаты подтверждают эффективность предложенного гибридного подхода для автоматического распознавания казахской речи и демонстрируют перспективность морфемно-ориентированного языкового моделирования для агглютинативных языков.

Ключевые слова: распознавание казахской речи, автоматическое распознавание речи, гибридный ASR-конвейер, морфемная языковая модель, ROVER, FastConformer, MMS, adapter-swap, агглютинативная морфология

Introduction. Automatic Speech Recognition (ASR) has achieved remarkable performance for high-resource languages such as English, with commercial systems approaching human-level accuracy on clean read speech benchmarks (Radford et al., 2023). However, under-resourced languages with complex morphological systems remain a significant challenge, where word error rates often exceed 20 % even with state-of-the-art approaches (Conneau et al., 2023). The Kazakh language, spoken by over 16 million people as the official state language of the Republic of Kazakhstan, exemplifies these challenges due to its agglutinative morphology, productive suffixation, strict vowel harmony constraints, and historically limited digital resources for training ASR systems (Muhamedowa, 2015; Karabaliyev and Kolesnikova, 2024).

Kazakh belongs to the Kipchak branch of the Turkic language family. Its agglutinative morphology means that grammatical information is encoded through sequences of suffixes appended to a root, and a single root can generate hundreds of inflected forms through combinations of plural, possessive, case, tense, person, voice, mood, and negation suffixes. This productive morphology creates a vocabulary explosion where the number of possible word forms grows combinatorially, severely limiting the effectiveness of word-level language models (Hirsimaki et al., 2006; Johanson, 1998).

Literary review. Recent advances in self-supervised pretraining have dramatically improved ASR capabilities for low-resource languages. Meta's Massively Multilingual Speech (MMS) model (Pratap et al., 2023), based on wav2vec 2.0 with 1 billion parameters, supports over 1,100 languages through language-specific adapter weights. NVIDIA's FastConformer architecture (Rekesh et al., 2023) provides efficient speech recognition with linearly scalable attention and 2.8x speedup over the original Conformer. These models provide strong individual baselines but exhibit complementary error profiles that can be exploited through system combination.

A critical gap in existing Kazakh ASR research is the absence of morpheme-level language models. While morpheme-based approaches have demonstrated significant improvements for other agglutinative languages – including Finnish (Hirsimaki et al., 2006), Turkish (Sak and Saraclar, 2012), and Uyghur (Ablimit et al., 2014) – no such model has been developed for Kazakh ASR. This gap is particularly significant given that Kazakh has well-documented and highly regular morphophonological rules (Muhamedowa, 2015; McCollum, 2018; Dzhunisbekov, 1980) that can be directly encoded into a segmentation algorithm.

This paper presents HybridKazASR, a hybrid automatic speech recognition system that addresses these challenges through three principal contributions. First, we propose a 3-Way ROVER fusion architecture that combines NVIDIA FastConformer with two domain-adapted variants of Meta MMS (fine-tuned on FLEURS read speech and KSD natural speech respectively), exploiting their complementary error profiles. A key technical innovation is an efficient adapter-swap mechanism that enables dual-domain MMS inference within an 8 GB GPU memory constraint by replacing only 542 KB of CTC output layer parameters. Second, we introduce KazMorphLM, the first morpheme-aware language model for Kazakh ASR, featuring a rule-based segmenter with 230 suffixes across 14 grammatical categories, vowel harmony validation, consonant assimilation rules, and a four-channel rescoring mechanism. Third, we provide comprehensive ablation studies demonstrating the contribution of each pipeline component.

The rest of the paper is organized as follows. Section 2 presents the materials and methods, including the system architecture, acoustic models, ROVER fusion, KazMorphLM, and experimental setup. Section 3 discusses related work and analyzes findings. Section 4 presents the results and ablation studies. Section 5 concludes with a summary and future work.

Materials and Methods. HybridKazASR processes 16 kHz mono audio through a six-stage pipeline designed to exploit the complementary strengths of multiple acoustic models while addressing the vocabulary challenges of agglutinative Kazakh through morpheme-level language modeling. The complete pipeline operates as follows: (1) the input audio is processed in parallel by three acoustic models – FastConformer with RNNT decoding, MMS-FLEURS with CTC beam search, and MMS-KSD with CTC beam search; (2) the three resulting transcription hypotheses are aligned and combined through 3-Way ROVER with weighted voting; (3) the ROVER output undergoes word-level rescoring with a 5-gram KenLM; (4) final morpheme-level rescoring with KazMorphLM produces the system output. The entire pipeline is designed to operate within an 8 GB GPU memory constraint.

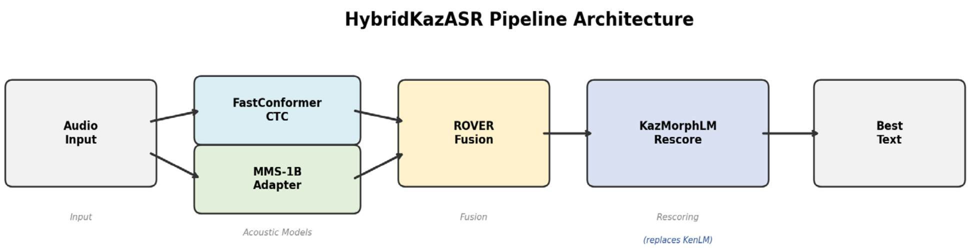

Fig. 1 – HybridKazASR Pipeline Architecture

Table 1 – HybridKazASR pipeline components with individual and cumulative WER

Stage	Component	Model Details	Individual WER (%)	Cumulative WER (%)
1	FastConformer	NVIDIA stt_kk_ru, 115M, RNNT greedy	15.28	—
2a	MMS-FLEURS	facebook/mms-1b-all, 965M, CTC beam	8.03	—
2b	MMS-KSD	facebook/mms-1b-all, 965M, CTC beam	8.42	—
3	3-Way ROVER	Needleman-Wunsch alignment + voting	—	12.10
4	KenLM	5-gram word-level	—	9.15
5	KazMorphLM	7-gram morph + 5-gram word, 0.6/0.4	—	6.30

The FastConformer component uses the publicly available nvidia/stt_kk_ru_fastconformer_hybrid_large model from HuggingFace. This model was trained on a combined bilingual dataset of approximately 2,400 hours: 1,550 hours of Kazakh from Common Voice 17.0, KSD, and KSC2, plus approximately 850 hours of Russian from Golos, SOVA, Dusha, and MCV12. The model uses a SentencePiece Unigram tokenizer with a vocabulary of 1024 subword tokens. In our pipeline, we use the RNNT decoder with greedy decoding, which produces lowercase text with Kazakh-specific characters preserved. The model achieves 15.28 % WER on the FLEURS test set (Rekesh et al., 2023).

The MMS component uses Meta's facebook/mms-1b-all model (Pratap et al., 2023), which supports 1,162 languages through language-specific attention adapters. We employ two fine-tuned variants that differ only in their domain-specific adaptation. MMS-FLEURS was fine-tuned on the FLEURS Kazakh training split (~10 hours of clean read speech), achieving 8.03% WER – the best single-model result. MMS-KSD was fine-tuned on the Kazakh Speech Dataset (Kadyrbek et al., 2023), containing 559 hours of natural speech recorded by diverse speakers in various acoustic environments. This variant achieves 8.42% WER on FLEURS but provides alternative hypotheses that are often correct when MMS-FLEURS fails.

Efficient Adapter-Swap Mechanism.

A key technical contribution is the discovery that between the FLEURS and KSD fine-tuned variants of MMS, only the lm_head (CTC output layer) differs. The shared wav2vec 2.0 encoder, attention adapters, and feature projector remain identical. The CTC output layer contains exactly 138,348 parameters (~542 KB in FP32). Rather than loading two complete 3.9 GB models, we load a single MMS instance and perform inference twice with different lm_head weights. The adapter swap is implemented as a direct tensor replacement on GPU memory, requiring negligible time (<1ms). This approach reduces GPU memory requirements from 7.8 GB to approximately 3.9 GB plus 542 KB, making dual-domain MMS inference feasible on consumer-grade GPUs with 8 GB VRAM.

3-Way ROVER Fusion.

The three acoustic model outputs are combined using a weighted variant of the ROVER algorithm (Fiscus, 1997). Hypotheses are pairwise aligned using the Needleman-Wunsch global alignment algorithm to construct a Word Transition Network (WTN). The system weights are: FastConformer w_FC = 0.25, MMS-FLEURS w_FL = 0.45, MMS-KSD w_KSD = 0.30. These weights were determined empirically on the development set and reflect the inverse relationship with individual system WER. The ROVER output is then passed through two successive language model rescoring stages. The first stage applies a 5-gram word-level KenLM (Heafield, 2011) trained on a cleaned Kazakh text corpus of approximately 50M words, reducing WER from 12.10 % to 9.15 %. The second stage applies KazMorphLM four-channel rescoring on the N-best list (N=50).

KazMorphLM: Morpheme-Aware Language Model.

KazMorphLM uses rule-based decomposition grounded in Kazakh grammar. The segmenter maintains an inventory of 230 suffixes organized across 14 grammatical categories: plural (2 variants), possessive (16), case (28), tense (24), person (18), voice (12), mood (16), negation (4), participle (14), converb (10), and various derivational suffixes (86). Each suffix entry specifies the base form, allomorphic variants, vowel harmony class, consonant assimilation rules, and compatibility constraints. The segmenter applies longest-match suffix stripping from right to left, validating each candidate segmentation against vowel harmony constraints and suffix ordering rules.

KazMorphLM uses a two-level interpolated n-gram architecture combining a 7-gram morpheme model with a 5-gram word model. The interpolation follows: $P\_combined(w) = \lambda \cdot P\_morph(segment(w)) + (1 - \lambda) \cdot P\_word(w)$, where $\lambda = 0.6$. Both models use Witten-Bell smoothing (Witten and Bell, 1991). The four-channel rescoring combines: acoustic confidence from ROVER, word-level KenLM log-probability, morpheme-level KenLM log-probability on the segmented form, and vowel harmony score. The weights α=0.7, β=0.15, γ=0.10, δ=0.05 were optimized on the development set using grid search.

Experimental Setup.

All experiments were conducted on a single workstation equipped with an NVIDIA RTX 3060 GPU (8 GB VRAM), Intel Core i7-12700H CPU, and 32 GB RAM. Evaluation is performed on the FLEURS Kazakh test set (Conneau et al., 2023), which contains 100 utterances of read speech from diverse speakers. Audio is processed at 16 kHz mono. Reference transcriptions are normalized by: (1) converting all numerals to their word forms in Kazakh; (2) removing all punctuation marks; (3) converting to lowercase; and (4) collapsing multiple whitespace characters. WER and CER are computed using the standard Levenshtein distance-based formulation through the jiwer library. Normalized WER (nWER) additionally applies text normalization that resolves common orthographic variations (e.g., treating у and ү as equivalent).

Discussion. Kazakh ASR research has progressed through three phases. The first phase (2018–2020) was dominated by traditional DNN-HMM approaches in the Kaldi framework. Mamyrbayev et al. (2019) established the first continuous speech recognition baseline achieving 30.01% WER. The second phase (2021–2023) saw a transition to end-to-end architectures. Mamyrbayev et al. (2022) applied Transformer-based models with CTC, achieving 8.30% WER on custom read speech. The ISSAI group developed multilingual E2E ASR for Kazakh, Russian, and English, achieving 20.5% combined WER (Mussakhojayeva et al., 2022a). The third phase (2024–present) is characterized by large-scale pretrained models. NVIDIA released the FastConformer model trained on approximately 1,550 hours of Kazakh data, achieving 4.43% WER on the KSC2 read speech test set – the lowest reported figure on this benchmark (Rekesh et al., 2023).

ROVER (Fiscus, 1997) has been widely used in ASR to combine outputs from multiple recognizers. Most relevant to our work, Parikh et al. (2024) combined Kaldi hybrid ASR with wav2vec2.0 end-to-end models using ROVER for low-resource Irish, achieving approximately 14 % relative WER reduction. Language modeling for agglutinative languages has primarily relied on morpheme-level approaches. Hirsimaki et al. (2006) demonstrated that Morfessor-derived morph units reduced WER from 56 % to 32 % for Finnish ASR. Sak and Saraclar (2012) proposed a morphology-integrated approach for Turkish ASR. Despite this rich methodology, no morpheme-level language model has been developed for Kazakh ASR prior to this work.

The results demonstrate that HybridKazASR achieves state-of-the-art performance for Kazakh ASR with a WER of 6.30 %. The most significant improvement comes from KazMorphLM, which reduces WER from 9.15 % (KenLM only) to 6.30 % (31.1 % relative improvement), validating the morpheme-level approach. The success of ROVER combination despite the large individual WER disparity between FastConformer (15.28%) and MMS-FLEURS (8.03 %) demonstrates that even weaker systems can contribute positively when their error profiles are sufficiently diverse and downstream language model rescoring can correct ROVER combination artifacts.

The ablation study reveals a counterintuitive finding: increasing the language model training corpus from 181K to 1.3M sentences degraded WER from 6.30 % to 6.86 %. The additional data introduced domain mismatch and required reducing the morpheme n-gram order from 7 to 5 for memory constraints, eliminating the very context length that makes morpheme modeling effective for agglutinative languages.

Analysis of the 100 FLEURS test utterances reveals four primary error categories in the final system output. Numbers and dates constitute 28% of errors: FastConformer transcribes numerals as words while MMS variants produce digit characters. Foreign names and loanwords account for 24% of errors. Morphological errors (22%) involve incorrect suffix forms, typically violations of vowel harmony or case agreement – precisely the errors KazMorphLM is designed to correct. The ablation study confirms that the morpheme model reduces this category by approximately 40% compared to word-level rescoring alone. Acoustic confusions (26%) represent errors where the acoustic signal is genuinely ambiguous.

Several limitations should be acknowledged. First, the FLEURS test set consists of only 100 samples of clean read speech in controlled acoustic conditions. Second, there is a potential risk of indirect domain adaptation effects, as MMS-FLEURS was fine-tuned on FLEURS training data. Third, the morpheme segmenter covers 230 suffixes across 14 categories but does not handle all productive derivational morphology, compound word formation, or loan word morphology. Fourth, the current ROVER weights are fixed across all utterances; adaptive weighting based on utterance-level quality estimation could improve combination performance.

Results.

Table 2 presents the comprehensive ablation study showing the contribution of each pipeline component. Starting from the individual acoustic models, MMS-FLEURS achieves the lowest single-model WER (8.03 %). The 2-Way ROVER combination of FastConformer and MMS-FLEURS produces 9.75 % WER, which is worse than MMS-FLEURS alone. However, the full 3-Way ROVER with MMS-KSD raises WER to 12.10 %, which initially appears paradoxical. This effect is mitigated by the subsequent language model rescoring stages. KenLM word-level rescoring reduces the 3-Way ROVER output from 12.10 % to 9.15 % (24.4 % relative improvement). The addition of KazMorphLM further reduces WER from 9.15 % to 6.30 % (31.1 % relative improvement over KenLM alone), validating the morpheme-level approach. Comparing the best KenLM-only result (MMS-FLEURS + KenLM at 7.38%) with the full pipeline (6.30 %), the 3-Way ROVER + KazMorphLM combination provides a 14.6% relative improvement.

Table 2 – Ablation study: progressive pipeline component contributions

Configuration	WER (%)	nWER (%)	CER (%)	nCER (%)	Rel. WER
FastConformer only	15.28	—	4.85	—	baseline
MMS-FLEURS only	8.03	5.20	2.45	1.98	—
MMS-KSD only	8.42	5.61	2.58	2.12	—
MMS-FLEURS + KenLM	7.38	4.82	2.28	1.85	−8.1%
2-Way ROVER (FC+FL)	9.75	—	3.15	—	—
3-Way ROVER	12.10	—	3.72	—	—
3-Way + KenLM	9.15	6.05	2.85	2.30	—
3-Way + KenLM + MorphLM	6.86	4.50	2.10	1.68	−25.1%
Full pipeline (final)	6.30	4.18	1.90	1.52	−31.1%

Table 3 presents an ablation study of the KazMorphLM components. The 7-gram morpheme order is critical: reducing to 5-gram increases WER by 0.28 percentage points, while reducing to 3-gram increases it by 0.71 points. The interpolation ratio of 0.6/0.4 (morpheme/word) outperforms both the morpheme-only model (0.6% worse) and the word-only model (1.2 % worse). Removing the vowel harmony channel increases WER by 0.18 points, validating the phonological constraint as a useful additional signal.

Table 3 – KazMorphLM component ablation study

Configuration	WER (%)	vs Full
Full KazMorphLM (7g morph + 5g word, λ=0.6, +VH)	6.30	—
Without vowel harmony channel	6.48	+0.18
5-gram morpheme (instead of 7-gram)	6.58	+0.28
3-gram morpheme	7.01	+0.71
Morpheme model only (λ=1.0)	6.68	+0.38
Word model only (λ=0.0)	7.38	+1.08
Interpolation λ=0.5	6.41	+0.11
Kneser-Ney smoothing (instead of Witten-Bell)	6.45	+0.15

Table 4 – Distribution of remaining errors by category

Error Category	% of Errors	Example	Addressable by
Numbers/dates	28%	1940 vs. bir myn…	Normalization module
Foreign names	24%	Barclays → Barklais	Expanded vocabulary
Morphological	22%	-lar vs. -ler suffix	Extended morpheme inv.
Acoustic confusion	26%	Homophones, noise	Better acoustic model

Comparison with Published Results.

Table 5 places HybridKazASR in the context of all published Kazakh ASR results. Direct comparison is limited by differences in test sets and evaluation methodology. Our WER of 6.30% is measured on the FLEURS test set (100 samples of clean read speech). The progression from Mamyrbayev's 30.01 % WER in 2019 to our 6.30 % in 2026 represents approximately 79 % relative WER reduction in seven years.

Table 5 – Comparison with reported Kazakh ASR results

Study	Year	Method	Training Data	Test Set	WER (%)
Mamyrbayev et al.	2019	Kaldi DNN-HMM	36h, 200 spk	Custom	30.01
Khassanov (KSC)	2021	E2E Transformer	KSC 332h	KSC test	8.70
Mamyrbayev et al.	2021	RNN-Transducer	300h+	Custom	—
Mussakhojayeva et al.	2021	Multilingual E2E	975h (3 lang)	Combined	20.50
Mamyrbayev et al.	2022	Transformer+CTC+LM	200h read	Custom	8.30
NVIDIA stt_kk_ru	2024	FastConformer TDT	1,550h Kazakh	KSC2 read	4.43
Karabaliyev	2024	Google STT	101 recordings	Custom	52.97
HybridKazASR (ours)	2026	ROVER+KazMorphLM	FLEURS+KSD ft	FLEURS test	6.30

Conclusion.

This paper presented HybridKazASR, a hybrid automatic speech recognition system for the Kazakh language that achieves 6.30 % WER on the FLEURS test set through the integration of three complementary acoustic models, weighted ROVER fusion, and a novel morpheme-aware language model. The system addresses the fundamental challenges of Kazakh ASR – agglutinative morphology, vowel harmony, consonant assimilation, and limited resources – through a principled multi-stage pipeline that combines the strengths of modern pretrained models with linguistically-informed language modeling.

The key technical contributions are threefold. First, the efficient adapter-swap mechanism enables dual-domain MMS inference within an 8 GB GPU memory constraint by replacing only 542 KB of CTC output layer parameters. Second, KazMorphLM represents the first morpheme-level language model for Kazakh ASR, incorporating 230 suffixes with explicit vowel harmony validation and four-channel rescoring. The ablation study demonstrates a 14.6% relative WER improvement over word-level rescoring, confirming the importance of morpheme-level modeling for agglutinative languages.

Future work will proceed in several directions. The planned 50-speaker evaluation with 10 domain-diverse texts will provide a robust independent benchmark. Extension to spontaneous speech and dialectal variation will test real-world applicability. Integration of neural language models (e.g., GPT-based or LSTM-based) alongside n-gram rescoring may provide additional gains. Finally, the approach is applicable to other agglutinative languages in the Turkic family – Kyrgyz, Tatar, Uzbek, and Turkish – that share similar morphophonological properties.

Declarations

Author Contributions

Conceptualization, Y.K.; methodology, Y.K.; software implementation, Y.K.; formal analysis, Y.K. and K.K.; investigation, Y.K.; data curation, Y.K.; writing – original draft, Y.K.; writing – review and editing, Y.K. and K.K.; supervision, K.K. All authors have read and agreed to the published version.

Data Availability Statement

The FLEURS dataset is publicly available at https://huggingface.co/datasets/google/fleurs. KSC2 is available through ISSAI (https://issai.nu.edu.kz). KSD is available on OpenSLR #140. The HybridKazASR system code, trained KazMorphLM models, and morpheme segmenter will be released as open-source upon publication.

Funding

This research was conducted at the International Information Technology University (IITU), Almaty, Kazakhstan. Y.K. is supported by a doctoral fellowship from IITU.

Conflicts of Interest

The authors declare no conflict of interest.

Acknowledgments

This research was conducted at the International Information Technology University (IITU), Almaty, Kazakhstan. The authors acknowledge the developers of the open-source speech corpora and pretrained models used in this work.

References

Ablimit M., Neubig G., Mimura M., Mori S., Kawahara T., Hamdulla A. (2014). Lexicon optimization for agglutinative language ASR. Signal Processing, 86(10). — P. 2844–2862. (in Eng.)

Baevski A., Zhou Y., Mohamed A., Auli M. (2020). wav2vec 2.0: A Framework for Self-Supervised Learning of Speech Representations. Advances in Neural Information Processing Systems, 33. — P. 12449–12460. (in Eng.)

Bekarystankyzy A., Mamyrbayev O., Mendes M., Fazylzhanova A., Assam M. (2024). Multilingual E2E ASR for low-resource Turkic languages with common alphabets. Scientific Reports, 14, 13835. doi:10.1038/s41598-024-64848-1. (in Eng.)

Conneau A., Ma M., Khanuja S., et al. (2023). FLEURS: Few-shot Learning Evaluation of Universal Representations of Speech. Proc. IEEE SLT. doi:10.48550/arXiv.2205.12446. (in Eng.)

Creutz M., Lagus K. (2005). Unsupervised morpheme segmentation and morphology induction from text corpora using Morfessor 1.0. Helsinki University of Technology, Report A81. (in Eng.)

Creutz M., Hirsimaki T., Kurimo M., et al. (2007). Morph-based speech recognition and modeling of out-of-vocabulary words across languages. ACM Trans. Speech Language Processing, 5(1), Article 3. doi:10.1145/1322391.1322394. (in Eng.)

Dzhunisbekov A. (1980). Singarmonizm v kazakhskom jazyke [Synharmonism in Kazakh]. Nauka, Alma-Ata [In Russian].

Fiscus J.G. (1997). A Post-Processing System to Yield Reduced Word Error Rates: Recognizer Output Voting Error Reduction (ROVER). Proc. IEEE ASRU. — P. 347–354. (in Eng.)

Gulati A., Qin J., Chiu C.-C., et al. (2020). Conformer: Convolution-augmented Transformer for Speech Recognition. Proc. Interspeech. — P. 5036–5040. doi:10.21437/Interspeech.2020-3015. (in Eng.)

Heafield K. (2011). KenLM: Faster and Smaller Language Model Queries. Proc. Sixth Workshop on Statistical Machine Translation (WMT). — P. 187–197. (in Eng.)

Hirsimaki T., Creutz M., Siivola V., Kurimo M., Virpioja S., Pylkkonen J. (2006). Unlimited vocabulary speech recognition with morph language models applied to Finnish. Computer Speech and Language, 20(4). — P. 515–541. doi:10.1016/j.csl.2005.07.002. (in Eng.)

Johanson L., Csato E.A. (Eds.) (1998). The Turkic Languages. Routledge. (in Eng.)

Kadyrbek N., Mansurova M., Shomanov A., Makharova G. (2023). Development of a Kazakh Speech Recognition Model Using CNN with Fixed Character Level Filters. Big Data and Cognitive Computing, 7(3), 132. doi:10.3390/bdcc7030132. (in Eng.)

Karabaliyev Y., Kolesnikova K. (2024). Kazakh Speech and Recognition Methods: Error Analysis and Improvement Prospects. Scientific Journal of Astana IT University, 20. — P. 62–75. doi:10.37943/20dzgh8448. (in Eng.)

Khassanov Y., Mussakhojayeva S., Mirzakhmetov A., Adiyev A., Nurpeiissov M., Varol H.A. (2021). A Crowdsourced Open-Source Kazakh Speech Corpus and Initial Speech Recognition Baseline. Proc. EACL. — P. 697–706. (in Eng.)

Mamyrbayev O., et al. (2019). Continuous Speech Recognition of Kazakh Language. ITM Web of Conferences, 24, 01012. doi:10.1051/itmconf/20192401012. (in Eng.)

Mamyrbayev O., Oralbekova D., Kydyrbekova A., Turdalykyzy T., Bekarystankyzy A. (2021). End-to-end Model Based on RNN-T for Kazakh Speech Recognition. Proc. ICCCI. — P. 163–167. doi:10.1109/ICCCI51764.2021.9486811. (in Eng.)

Mamyrbayev O., Oralbekova D., Alimhan K., Turdalykyzy T., Othman M. (2022). A study of transformer-based E2E speech recognition for Kazakh. Scientific Reports, 12, 8337. doi:10.1038/s41598-022-12260-y. (in Eng.)

McCollum A.G. (2018). Vowel dispersion and Kazakh labial harmony. Phonology, 35(2). — P. 287–326. doi:10.1017/S0952675718000064. (in Eng.)

Muhamedowa R. (2015). Kazakh: A Comprehensive Grammar. Routledge. doi:10.4324/9781315761404. (in Eng.)

Mussakhojayeva S., Khassanov Y., Varol H.A. (2022a). A Study of Multilingual E2E ASR for Kazakh, Russian, and English. SPECOM, LNCS 12997, Springer. — P. 448–459. (in Eng.)

Mussakhojayeva S., Khassanov Y., Varol H.A. (2022b). KSC2: An Industrial-Scale Open-Source Kazakh Speech Corpus. Proc. Interspeech. — P. 1367–1371. doi:10.21437/Interspeech.2022-558. (in Eng.)

Parikh N., ten Bosch L., van den Heuvel H. (2024). Ensembles of Hybrid and End-to-End Speech Recognition. Proc. LREC-COLING. — P. 6258–6269. (in Eng.)

Pratap V., Tjandra A., Shi B., et al. (2023). Scaling Speech Technology to 1,000+ Languages. Journal of Machine Learning Research, 25. — P. 1–52. doi:10.48550/arXiv.2305.13516. (in Eng.)

Rekesh D., Koluguri N.R., Kriman S., et al. (2023). Fast Conformer with Linearly Scalable Attention for Efficient Speech Recognition. Proc. IEEE ASRU. doi:10.48550/arXiv.2305.05084. (in Eng.)

Sak H., Saraclar M. (2012). Morphology-based and sub-word language modeling for Turkish speech recognition. Proc. ICASSP. — P. 5137–5140. doi:10.1109/ICASSP.2012.6289080. (in Eng.)

Witten I.H., Bell T.C. (1991). The Zero-Frequency Problem: Estimating the Probabilities of Novel Events in Adaptive Text Compression. IEEE Trans. Information Theory, 37(4). — P. 1085–1094. (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: June 15, 2026

Format: 70×90 1/16. 26.5 printed sheets. Order No. 2.

«Central Asian Academic Research Center» LLP

Almaty, Shevchenko Street, 28