

INNOVATION SCIENCE AND TECHNOLOGY

ISSUE 9, PART I / SEPTEMBER 2026

 Acceptance of papers



**Acceptance of
papers**

Published monthly



Topics

economics,
technology, social
sciences



EDITOR-IN-CHIEF:

Mirzaliyev Sanjar Makhmatjon ugli

DEPUTY EDITOR-IN-CHIEF:

Makhmudov Nosir Makhmudovich
DSc., Prof., Academician

DEPUTY EDITOR-IN-CHIEF:

Ochilov Bobur Bakhtiyor ugli – Senior
lecturer at TSUI

THE SCIENTIFIC-POPULAR ELECTRONIC
JOURNAL **"INNOVATION SCIENCE AND
TECHNOLOGY"** HAS BEEN REGISTERED
UNDER THE NUMBER **C-5669633** BY THE
AGENCY FOR INFORMATION AND MASS
COMMUNICATIONS (AOKA) OF THE
REPUBLIC OF UZBEKISTAN, EFFECTIVE
FROM OCTOBER 9, 2024.

CONTACTS

Phone: **+998 50 737 87 88**

Website: <https://ist-journal.uz>

Email: innovationist2025@gmail.com

The scientific electronic journal "Innovation Science and Technology" has been included in the list of scientific publications recommended for the publication of main scientific results of dissertations for the award of PhD and DSc degrees in economics and technical sciences, in accordance with the Resolution No. 370 of the Presidium of the Higher Attestation Commission of the Republic of Uzbekistan, dated May 8, 2025.

Editorial board:



Sharipov Kongiratbay Avezimbetovich,
Doctor of Technical Sciences (DSc), Professor



Abdurakhmanova Gulnora Kalandarovna, Doctor of
Economic Sciences (DSc), Professor



Cham Tat Huei,
Doctor of Philosophy (PhD), Professor (Malaysia)



Muhammad Imran Sadiq
Doctor of Philosophy in Economics (PhD), Professor,
Malaysia



Ahmed Aziz Ismail
Doctor of Technical Sciences (DSc),
Professor (Egypt)



Lee Chin
Doctor of Philosophy in Economics (PhD), (Malaysia)



Asongu Simplice
Doctor of Philosophy in Economics (PhD), Cameroon



Rui Dang
Doctor of Chemistry (DSc), Professor, China



Zahoor Ahmed
Doctor of Philosophy in Economics (PhD), Turkey



Shujaat Abbas
Doctor of Philosophy in Economics (PhD), Russia



Tina A Coffelt
Doctor of Philosophy in Educational Sciences (PhD),
USA



Abdikarimova Dinara Rustamxanovna
Doctor of Economic Sciences (DSc), Professor

Kurbonbekova Mohichehra Turobjonovna
Doctor of Economic Sciences (DSc), Professor

Alimardonov Ilkhom Muzrabshokovich
Doctor of Economic Sciences (DSc), Professor



Razakova Barno Sayfieva
Doctor of Philosophy in Economics (PhD)



Khasanov Sarvar Ulugbek ugli
Doctor of Philosophy in Economics (PhD)



Kholikova Rukhsora Sanjarovna
Associate Professor (PhD)

CONTENTS

ECONOMIC ESSENCE AND CLASSIFICATION OF FINANCIAL ERRORS AND IRREGULARITIES IN BUDGET ORGANIZATIONS AND CRITERIA FOR THEIR INTERNAL AUDIT ASSESSMENT.....	8
Mulayev Farxod Alisherovich	
THE IMPACT OF FISCAL DISCIPLINE AND PUBLIC DEBT ON NATIONAL CURRENCY STABILITY: THEORETICAL APPROACHES AND MODELING	14
G'afurov Olimjon G'olib o'g'li	
THE ROLE OF REMOTE BANKING SERVICES IN ENHANCING FINANCIAL INCLUSION IN UZBEKISTAN	20
Sharipov Tohir Mirakhimovich	
THE IMPORTANCE OF STANDARDIZATION REQUIREMENTS IN THE ACCREDITATION OF NON-DESTRUCTIVE TESTING LABORATORIES.....	26
Ahmedov G'ayratjon Mahmudjonovich	
THEORETICAL ASPECTS OF ASSESSING THE IMPACT OF CORPORATE GOVERNANCE ON ECONOMIC EFFICIENCY IN CONSTRUCTION ENTERPRISES.....	32
Giyosov Ilkhom Karimovich	
FOREIGN EXPERIENCE IN STATE REGULATION OF SUSTAINABLE TOURISM DEVELOPMENT AND THE POSSIBILITIES OF ITS APPLICATION IN UZBEKISTAN.....	39
O'sarov Abdulla Davronqulovich	
THE CURRENT STATE AND DEVELOPMENT TRENDS OF THE SMALL BUSINESS SECTOR IN THE REPUBLIC AND ITS IMPACT ON HUMAN CAPITAL.....	46
Ulugmuradova Nodira Berdimuradovna	
INCREASING ECONOMIC EFFICIENCY BASED ON DIGITAL PLATFORMS IN THE IMPLEMENTATION OF ALTERNATIVE ENERGY TECHNOLOGIES (THE CASE OF RURAL AREAS).....	53
Ashurov Azizbek Ergash ugli	
WAYS TO IMPROVE THE QUALITY OF ASSETS AND STABILITY OF THE CREDIT PORTFOLIO IN COMMERCIAL BANKS	58
Baymuratova Zina Aqilbekovna	
MEASURING GREEN EFFICIENCY IN THE MINING INDUSTRY THROUGH AN INTEGRATED INDICATOR-BASED APPROACH FROM NAVOI MINING AND METALLURGICAL COMPANY.....	63
Kurbanova Mehriniso, Ergashev Hamidulla	
CORRELATION BETWEEN MACROECONOMIC INDICATORS AND TAX REVENUES.....	69
Khakimov Ulugbek Abdullaevich	
ENHANCING ECONOMIC EFFICIENCY AND FORECAST INDICATORS OF SCALING BIG DATA SOLUTIONS IN THE TOURISM INDUSTRY	74
E. I. Temirov	
IMPROVING THE EFFICIENCY OF STATE FINANCIAL SUPPORT FOR FARMS: A RESULTS- AND RISK-ORIENTED APPROACH.....	81
Rashidov Baxadir Yusupovich	
THE ROLE OF REMOTE BANKING SERVICES IN ENHANCING FINANCIAL INCLUSION IN UZBEKISTAN	86
Sharipov Tohir Mirakhimovich	
ORGANIZATIONAL AND LEGAL FOUNDATIONS FOR ESTABLISHING A PERSONNEL RESERVE FOR LOCAL GOVERNMENT AUTHORITIES IN UZBEKISTAN.....	92
G'aniyev Elyor Sobirjonovich	
METAEXTRACT: A HYBRID METADATA EXTRACTION SYSTEM COMBINING RULE-BASED AND STATISTICAL APPROACHES	100
Rashid Muhtorovich Turgunbaev	
DIRECTIONS FOR USING FINANCIAL ANALYSIS RESULTS IN THE DEVELOPMENT OF E-COMMERCE PLATFORMS IN UZBEKISTAN.....	105
Amankulova Mukhlisa Narsafar kizi	

IMPROVING THE ORGANISATIONAL-ECONOMIC MECHANISMS FOR ENHANCING THE PROFITABILITY OF ENTREPRENEURIAL ACTIVITY	111
Inobatov Abror Boshlarovich	
THE CURRENT STATE AND DYNAMICS OF FINANCING ENTREPRENEURIAL ACTIVITIES THROUGH BANK LOANS.....	117
Ergashev Otamurod Tashtemirovich	
IMPROVING THE METHODOLOGY FOR ASSESSING THE EFFICIENCY OF LAND TAX ADMINISTRATION FOR INDIVIDUALS	123
Ramazonov Shamil Ulugbek ogli	
IMPROVING THE EFFICIENCY OF STRATEGIC RESOURCE ALLOCATION IN BANKING ACTIVITIES THROUGH ARTIFICIAL INTELLIGENCE.....	130
Shukhrat Komilovich Roziboyev	
STRATEGIC DIRECTIONS FOR ENHANCING THE COMPETITIVE POTENTIAL OF ENTERPRISES THROUGH INNOVATION-DRIVEN DEVELOPMENT	138
Karima Samatovna Tashmukhamedova	
FUNCTIONAL FEATURES OF SEMANTIC PHENOMENA IN ENGLISH AND UZBEK METALLURGICAL TERMINOLOGY	143
O'tkirova Shahlo Asqar qizi	
DECENTRALIZED TECHNOLOGIES AND ARTIFICIAL INTELLIGENCE AS DRIVERS OF CROWDFUNDING EVOLUTION	151
Karimova Aziza A'zamiddin qizi	
THEORETICAL FOUNDATIONS FOR USING ARTIFICIAL INTELLIGENCE IN ASSESSING CORPORATE FINANCIAL STABILITY	159
Jokhongir Hasanovich Dagarov	
A MODEL FOR ASSESSING THE ECONOMIC EFFICIENCY OF INTRODUCING DIGITAL PLATFORMS	163
Nabiev Zafar Irkinovich	
CHARGING OF SYNTHETIC FIBERS AND IMPROVEMENT OF EMULSION.....	169
Arabov Jamoliddin Sadriddinovich	
ECONOMETRIC ANALYSIS OF FACTORS AFFECTING THE FINANCING OF INNOVATIVE ACTIVITIES OF TOURIST ENTERPRISES IN THE SAMARKAND REGION	177
Ruzibaeva Nargiza Khakimovna	
IMPLEMENTATION OF QUALIFIED PERSONNEL TRAINING CLUSTERS IN THE TOURISM SECTOR OF THE REGIONAL ECONOMY	181
Raxmatov Adxam Itolmasovich	
ASSESSING THE DIGITAL MATURITY OF ENTERPRISES AND IMPROVING THE MECHANISMS FOR STRENGTHENING THEIR COMPETITIVENESS: AN INDEX-BASED FRAMEWORK FOR UZBEK INDUSTRIAL AND SERVICE FIRMS	189
Kungratov Ilmurod Kuzibay ugli	
DIRECTIONS FOR DEVELOPING EMPLOYEE COMPETENCIES IN THE DIGITAL ECONOMY	197
Musaev Ozodjon Shavkat Ogli	
SPEAKER DIARIZATION OF KARAKALPAK SPEECH USING A PRETRAINED SORTFORMER MODEL UNDER LOW-RESOURCE CONDITIONS.....	202
Kudaybergenov Tangirbergen Kadirbergenovich	
ANALYSIS OF THERMAL AND ELECTRICAL PARAMETERS OF A PHOTOVOLTAIC-THERMAL ELECTROBIOMEMBRANE REACTOR IN WASTEWATER TREATMENT FOR MICROPOLLUTANT REMOVAL	208
Sevara U. Mirzayorova	
MODELING MAGNETOELASTIC VIBRATIONS OF COMPLEX-SHAPED TRANSVERSELY ISOTROPIC PLATES USING THE R-FUNCTION METHOD.....	212
Faxriddin Nuraliyev, Nuriddin Tojiyev, Bobir Adilov	

OPTIMAL CONDITIONS FOR EXTRACTION OF VALUABLE COMPONENTS FROM RICE HUSK.....	221
Umarov Faxriddin, Rahimov Baxodir, Blsmanov Ne'matullo, Ataxo'jayev Abdubositxo'ja	
IMPROVING THE ORGANIZATIONAL AND ECONOMIC MECHANISM FOR THE DEVELOPMENT OF THE SERVICES SECTOR: PROBLEMS AND PROSPECTIVE DIRECTIONS	227
Yo'ldoshev Shuxratmirza Tursunboy o'g'li	
INVESTIGATION OF THE KINETICS AND OPTIMAL CONDITIONS FOR THE PRODUCTION OF $FeCl_3$ COAGULANT FROM TEBINBULOQ ORES.....	235
Nurmurodov To'ikin, Oljayev Dilshod, Mansurova Shabnam	
ENVIRONMENTAL PROTECTION THROUGH RESOURCE-EFFICIENT WASTE UTILIZATION.....	241
Toirova Nuriya, Urinova Xulkar	
TECHNOLOGICAL AND DIGITAL DIVERSIFICATION IN INDUSTRIAL ENTERPRISES: DIRECTIONS FOR ADAPTING FOREIGN EXPERIENCES TO THE CONDITIONS OF UZBEKISTAN	245
Asadov Elyorbek Kandiyarovich	
A PROCESS-AWARE DATA PREPROCESSING AND TIME-DELAY IDENTIFICATION FRAMEWORK FOR CEMENT CLINKER ENERGY MONITORING	250
Abror Turakulov, Maksadkhon Yakubov, Zafar Turakulov	
DEVELOPMENT OF KPIS TO MEASURE THE SUCCESS OF INNOVATIVE ENVIRONMENTAL PROJECTS WITHIN THE COMPANY.....	260
Musadjanova Nargiza Abduvoxid qizi	
ANALYSIS OF THE EFFICIENT USE OF FINANCIAL RESOURCES IN ENSURING EFFECTIVE ENTREPRENEURIAL ACTIVITY IN THE REGION (ANALYSIS OF THE STATE OF ENTREPRENEURSHIP DEVELOPMENT IN FERGANA REGION).....	266
Bozorov Ilyos Isomiddinovich	
TIJORAT BANKLARIDA ICHKI AUDIT AMALLARINI BAJARISH.....	271
Yuldashev Xurshidbek Mamirjonovich	
THE ROLE OF FASHION ELEMENTS IN THE FORMATION OF A NATION BRAND: A CONCEPTUAL FRAMEWORK AND RESEARCH METHODOLOGY	277
Yusupova Mokhinur Farkhod qizi	
THE ORGANISATIONAL AND ECONOMIC MECHANISM FOR EVALUATING THE EFFECTIVENESS OF MARKETING PROMOTION OF EDUCATIONAL SERVICES.....	287
Umarov Shuhrat Bahodir o'g'li	
COMPUTER VISION TECHNOLOGIES TO IMPROVE THE IMAGE QUALITY OF THE IRIS	297
Makhmudova Shakhzoda, Sharopova Barno, Saydazimov Javlonbek	
SPECIFIC ASPECTS OF ENHANCING INDUSTRIAL PRODUCTION COMPETITIVENESS.....	301
Olimov Maksudjon, Habibullayev Oybek	
COMPARATIVE EVALUATION OF CNN-LSTM AND SQUEEZEFORMER-TRANSFORMER ARCHITECTURES FOR REAL-TIME UZBEK SIGN LANGUAGE RECOGNITION	307
Jasurbek Ergashev, Munira Rakhmatova, Mukhtor Eshimov, Ibragim Atadjonov, Sajid Khan	

COMPARATIVE EVALUATION OF CNN-LSTM AND SQUEEZEFORMER-TRANSFORMER ARCHITECTURES FOR REAL-TIME UZBEK SIGN LANGUAGE RECOGNITION

Jasurbek Ergashev

Researcher at Department of Computer Science, School of Engineering
Central Asian University

Munira Rakhmatova

Researcher at Department of Computer Science, School of Engineering
Central Asian University

Mukhtor Eshimov

Researcher at Department of Computer Science, School of Engineering
Central Asian University

Supervisor: Ibragim Atadjonov

Researcher at Department of Computer Science, School of Engineering
Central Asian University

Co-Supervisor: Sajid Khan

Researcher at Department of Computer Science, School of Engineering
Central Asian University

Abstract: Uzbek Sign Language (UzSL) remains underrepresented in machine-learning research, in part because annotated datasets and deployable recognition systems are scarce. This paper develops and evaluates a camera-based, isolated-sign recognition pipeline for UzSL using a proprietary landmark dataset, and compares a CNN-LSTM baseline against a Squeezeformer-style Transformer architecture. Over 500 gesture classes were recorded with HD webcams (1280×720, 30 fps) under varied lighting, background, clothing, and camera conditions, converted to normalized 130-point landmark sequences, and split 70/15/15 into training, validation, and test sets. The Transformer achieved 91% accuracy versus 87% for CNN-LSTM, with lower latency (0.58 s vs. 0.72 s), higher throughput (78 vs. 65 gestures/min), and stronger low-light accuracy (83% vs. 76%). A small controlled usability study (n = 10) returned a mean satisfaction score of 4.6/5. These results support attention-based sequence modeling as a promising direction for real-time UzSL recognition, though the system is currently limited to isolated-sign recognition and has not been evaluated with a signer-independent split. We present the dataset and prototype as a reproducible starting point for future continuous UzSL translation research.

Keywords: Uzbek Sign Language; sign language recognition; CNN-LSTM; Transformer; Squeezeformer; computer vision; landmark sequences; assistive technology.

Annotatsiya: O'zbek imo-ishora tili (UzSL) mashinaviy o'rganish sohasidagi tadqiqotlarda yetarlicha o'rganilmagan bo'lib qolmoqda. Bunga qisman annotatsiyalangan ma'lumotlar to'plamlari va amaliyotga joriy etishga tayyor imo-ishoralarni tanib olish tizimlarining kamligi sabab bo'lmoqda. Ushbu maqolada xususiy landmark ma'lumotlar to'plamidan foydalan-gan holda UzSL uchun kameraga asoslangan, alohida imo-ishoralarni tanib olish konveyeri ishlab chiqiladi va baholanadi hamda CNN-LSTM bazaviy modeli Squeezeformer uslubidagi Transformer arxitekturasini bilan taqqoslanadi. 500 dan ortiq imo-ishora sinfi HD veb-kameralar (1280×720, 30 kadr/soniya) yordamida turli yorug'lik, fon, kiyim-kechak va kamera sharoitlarida yozib olindi, 130 ta nuqtadan iborat normallashtirilgan landmark ketma-ketliklariga aylantirildi hamda o'qitish, validatsiya va test to'plamlariga 70/15/15 nisbatida ajratildi. Transformer modeli CNN-LSTM modelining 87 foizlik aniqli-giga nisbatan 91 foizlik aniqlikka erishdi, shuningdek, kechikish vaqti pastroq (0,58 soniya, 0,72 soniyaga nisbatan), o'tka-zuvchanligi yuqoriroq (daqiqasiga 78 ta, 65 ta imo-ishoraga nisbatan) va kam yorug'lik sharoitidagi aniqligi yuqoriroq (83 %, 76 % ga nisbatan) bo'ldi. Foydalanish qulayligi bo'yicha o'tkazilgan kichik nazoratli tadqiqot (n = 10) natijasida o'rtacha

qoniqish darajasi 4,6/5 ni tashkil etdi. Ushbu natijalar real vaqt rejimida UzSLni tanib olish uchun diqqat mexanizmiga asoslangan ketma-ketlik modellashtirish istiqbolli yo'nalish ekanligini ko'rsatadi. Biroq tizim hozirda alohida imo-ishoralarni tanib olish bilan cheklangan va imo-ishoralarni bajaruvchilardan mustaqil ma'lumotlarni ajratish usuli asosida baholanmagan. Biz ma'lumotlar to'plami va prototipni UzSLning uzluksiz tarjimasini bo'yicha kelgusidagi tadqiqotlar uchun qayta ishlab chiqish mumkin bo'lgan boshlang'ich asos sifatida taqdim etamiz.

Kalit so'zlar: O'zbek imo-ishora tili; imo-ishora tilini tanib olish; CNN-LSTM; Transformer; Squeezeformer; kompyuter ko'rishi; landmark ketma-ketliklari; yordamchi texnologiyalar.

Аннотация: Узбекский жестовый язык (UzSL) остается недостаточно представленным в исследованиях в области машинного обучения, отчасти из-за нехватки аннотированных наборов данных и готовых к внедрению систем распознавания. В данной статье разрабатывается и оценивается основанный на использовании камеры конвейер распознавания отдельных жестов UzSL с применением собственного набора данных о ключевых точках, а также проводится сравнение базовой архитектуры CNN-LSTM с архитектурой Transformer в стиле Squeezeformer. Более 500 классов жестов были записаны с помощью HD-веб-камер (1280×720, 30 кадров/с) в условиях различного освещения, фона, одежды и положения камеры, преобразованы в нормализованные последовательности из 130 ключевых точек и разделены на обучающую, валидационную и тестовую выборки в соотношении 70/15/15. Transformer продемонстрировал точность 91 % по сравнению с 87 % у CNN-LSTM, меньшую задержку (0,58 с против 0,72 с), более высокую пропускную способность (78 против 65 жестов/мин) и более высокую точность при слабом освещении (83 % против 76 %). Небольшое контролируемое исследование удобства использования ($n = 10$) показало среднюю оценку удовлетворенности 4,6/5. Полученные результаты свидетельствуют о перспективности моделирования последовательностей на основе механизма внимания для распознавания UzSL в режиме реального времени, однако система в настоящее время ограничена распознаванием отдельных жестов и не оценивалась с использованием разделения данных по принципу независимости исполнителей жестов. Набор данных и прототип представлены в качестве воспроизводимой основы для дальнейших исследований в области непрерывного перевода UzSL.

Ключевые слова: узбекский жестовый язык; распознавание жестового языка; CNN-LSTM; Transformer; Squeezeformer; компьютерное зрение; последовательности ключевых точек; вспомогательные технологии.

INTRODUCTION

Communication between deaf or hard-of-hearing people and hearing people is constrained when the two groups do not share a sign language. Computer vision and deep learning have enabled substantial progress in sign-language recognition, but research and public resources remain concentrated on comparatively well-resourced languages such as American Sign Language (ASL). Uzbek Sign Language (UzSL) is considerably less represented in machine-learning datasets and practical recognition systems.

The underlying challenge is not model selection alone. A working recognition system needs representative annotated data, temporal modeling of gestures, robustness to changes in signer behavior and recording conditions, and an interface that can deliver predictions with sufficiently low latency. The present study, based on a bachelor's thesis conducted at Central Asian University [1], addressed these requirements by building an in-house UzSL dataset and a real-time recognition prototype. This manuscript reorganizes that work as a focused empirical comparison of two model families: CNN-LSTM and a Squeezeformer-based Transformer, evaluated on the same landmark data.

The central research question is: which of the two evaluated architectures offers a better balance of recognition accuracy and real-time performance for the collected UzSL landmark sequences? The study's contributions are threefold: (1) a proprietary dataset spanning 500+ gesture classes recorded under varied conditions; (2) a head-to-head comparison of a CNN-LSTM baseline and a Squeezeformer-Transformer model on the same data; and (3) an integrated prototype demonstrating real-time camera input, gesture-to-text output, and optional speech synthesis.

The work is deliberately scoped as isolated-sign recognition rather than continuous-language translation. Natural sign-language communication is contextual and sequential, and the system described here does not solve sentence segmentation, continuous translation, or broad signer-independent generalization - these are identified as priorities for future work.

REVIEW OF LITERATURE ON THE SUBJECT

Sign languages are the primary language of an estimated 70 million people worldwide, yet unlike spoken language they are not universal and vary substantially across linguistic, grammatical, and dialectal lines. Early sign-language recognition (SLR) systems relied on handcrafted features and hardware such as data gloves or depth sensors; the shift to standard RGB cameras combined with deep learning has since made SLR more scalable and affordable.

CNN-based models have been widely used for extracting spatial features in sign-language recognition, while LSTM-based models can capture temporal relationships in dynamic gestures [2]. Hybrid CNN-LSTM designs combine spatial feature extraction with sequential modeling and have been applied to dynamic sign-language recognition [3], which is consistent with earlier CNN-based approaches to sign-language recognition more broadly [4].

Transformer architectures represent a distinct mechanism: self-attention models relationships across distant positions in a sequence and allows parallel processing during training, in contrast to the sequential computation of RNNs and LSTMs [5]. Attention-based models have also been successfully applied to isolated sign-language recognition, including approaches based on Transformer architectures [6]. Newer variants such as Conformer and Squeezeformer, originally developed for speech recognition, combine self-attention with convolutional modules to capture both long-range dependencies and local temporal patterns, and are increasingly explored for sign-language translation.

Across this literature, three practical gaps are relevant to UzSL. First, annotated datasets for low-resource sign languages are scarce, in contrast to well-resourced datasets such as RWTH-PHOENIX-Weather or MS-ASL for ASL; broader signer diversity and signer-independent evaluation are important for assessing generalization across different signing styles and recording conditions [7]. Second, most systems, including this one, concentrate on isolated signs rather than continuous, sentence-level translation. Third, deployment is often constrained by dependence on specialized sensors, high-resolution processing, or server-side computation, which limits accessibility in lower-resource settings. These gaps motivate the combined dataset-construction, model-comparison, and prototype-deployment approach taken here; the study does not claim to close the continuous-translation gap, but evaluates whether a landmark-based Transformer offers a measurable advantage over a CNN-LSTM baseline within the collected vocabulary.

RESEARCH METHODOLOGY

Because no public UzSL dataset was available, we recorded a proprietary dataset specifically for this project. More than 500 unique gesture classes were captured, covering greetings, personal pronouns, basic questions, common verbs, emotions, and numbers. Each gesture was recorded multiple times per signer using HD webcams at 1280×720 resolution and 30 frames per second. The collection protocol varied natural versus artificial lighting, background contrast, clothing, and camera angle; recording was primarily indoors, with a limited set of outdoor samples included to assess generalization. Participants were briefed on the recording protocol and consented to academic and research use of their recordings (Figure 1, 2).

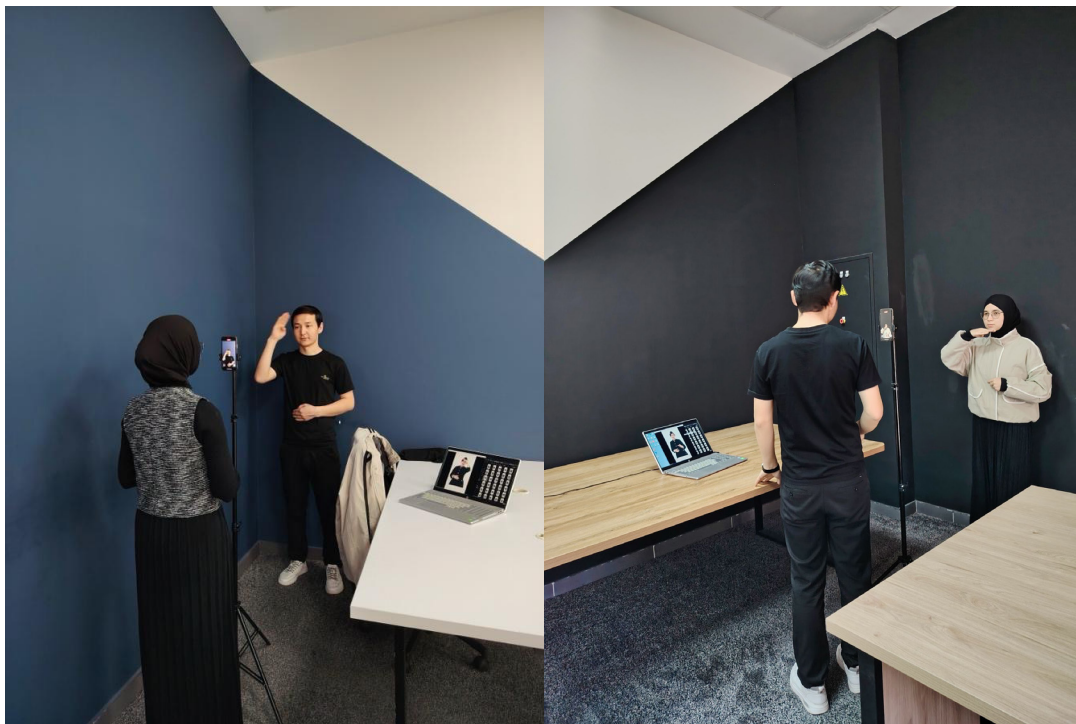


Figure 1. Data-collection setup: signers were recorded against contrasting backgrounds using a tripod-mounted

camera and a companion laptop for prompt display and logging¹

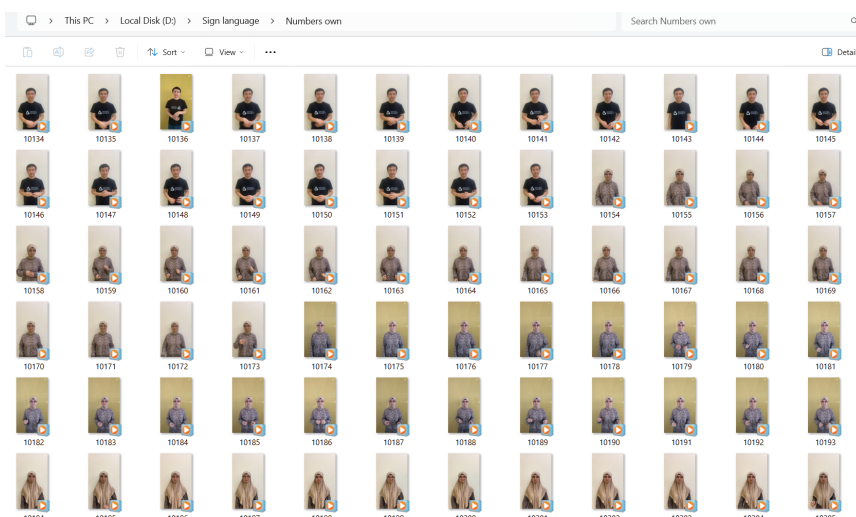


Figure 2. Sample of recorded gesture videos as organized on disk, illustrating per-signer, per-gesture file naming used for annotation²

Videos were annotated manually with Python scripts, storing labels in structured JSON with frame-level segmentation to support sequence-model training. Metadata for each clip, including signer identity, lighting condition, and recording angle, was retained to support balanced validation sampling. Each frame was reduced to 130 normalized landmark keypoints (x, y, z), extracted using a MediaPipe-based perception pipeline [8].

Data augmentation was applied primarily on landmark sequences: temporal resizing, shifting, time-warping, and temporal masking; spatial left-right flips, affine transforms (scale, shear, shift, rotation), and spatial keypoint masking; small-magnitude coordinate noise; and a CutMix variant adapted for sequences, which mixes segments from two landmark sequences along with their target text. The Squeezeformer-based model additionally used FingerDropout, Face/PoseDropout, and decoder-input masking as further robustness measures. The dataset was partitioned 70% training / 15% validation / 15% test (Figure 3).

Numbers				Pronouns				Family			
id	video_id	text_uz	text_en	id	video_id	text_uz	text_en	id	video_id	text_uz	text_en
1	10001	bir	one	199	20001	men	I	436	30001	inson	person
2	10002	bir	one	200	20002	sen	you	437	30002	inson	person
3	10003	ikki	two	201	20003	sen	you	438	30003	erkak	man
4	10004	ikki	two	202	20004	u	he/she/it	439	30004	erkak	man
5	10006	uch	three	203	20005	biz	we	440	30005	yigit	boy
6	10007	to'rt	four	204	20006	biz	we	441	30006	ayol	woman
7	10008	to'rt	four	205	20007	siz	you	442	30007	ayol	woman
8	10009	besh	five	206	20008	siz	you	443	30008	qiz	girl
9	10010	besh	five	207	20009	ular	they	444	30009	qiz	girl
10	10011	olti	six	208	20010	ular	they	445	30010	chagaloq	baby
11	10012	olti	six	209	20011	mening	my	446	30011	chagaloq	baby
12	10013	yetti	seven	210	20012	mening	my	447	30012	oila	family
13	10014	yetti	seven	211	20013	sizing	your	448	30013	oila	family
14	10015	sakkiz	eight	212	20014	sizing	your	449	30014	ota	father
15	10016	sakkiz	eight	213	20015	qaerda	where	450	30015	ota	father
16	10017	to'qqiz	nine	214	20016	qaerda	where	451	30016	ona	mother
17	10018	to'qiz	nine	215	20017	nimaga	why	452	30017	ona	mother
18	10019	o'n	ten	216	20018	nimaga	why	453	30018	buvi	grandmother
19	10020	o'n	ten	217	20019	qaysi	which	454	30019	buvi	grandmother
20	10021	o'n bir	eleven	218	20020	qaysi	which	455	30020	bobo	grandfather
21	10022	o'n bir	eleven	219	20021	qachon	when	456	30021	bobo	grandfather
22	10023	o'n ikki	twelve	220	20022	qachon	when	457	30022	opa	older_sister
23	10024	o'n ikki	twelve	221	20023	kim	who	458	30023	opa	older_sister
24	10025	o'n uch	thirteen	222	20024	kim	who	459	30024	singil	little_sister
25	10027	o'n to'rt	fourteen	223	20025	nima uchun	why	460	30025	singil	little_sister
26	10028	o'n to'rt	fourteen	224	20026	nima uchun	why	461	30026	aka	older_brother
27	10029	o'n besh	fifteen	225	20029	tez tez	often	462	30027	aka	older_brother
28	10030	o'n besh	fifteen	226	20030	tez tez	often	463	30028	uka	little_brother

Figure 3. Excerpt of the annotated dataset, showing UzSL and English gloss pairs across three example vocabulary

¹ Source: Author's own work.
² Source: Author's own work.

groups (numbers, pronouns, family terms)³

The baseline follows a two-stage design: convolutional (or, for landmark input, 1D-convolutional/dense per-timestep) processing first transforms frame-level features, after which LSTM layers model temporal dependencies; a final dense/softmax layer produces the predicted sign. CNN-LSTM was chosen as a comparatively simple, well-established temporal baseline against which the attention-based model could be measured, following prior demonstrations of CNN-LSTM effectiveness for gesture recognition tasks involving structured movement such as sign language [4].

The primary model uses multi-stream landmark feature extraction, a Squeezeformer-style encoder, a Transformer decoder, and an auxiliary confidence head. An “all-landmarks” branch processes all 130 keypoints through 2D convolution, batch normalization, SiLU activation, and a fully-connected layer, producing a 208-dimensional feature vector per frame. Four region-specific branches (face, left hand, right hand, pose) each apply the same operations to their respective keypoint subsets, each yielding a 52-dimensional vector; their concatenation (208-D) is added element-wise to the all-landmarks output to form the 208-dimensional per-frame encoder input, with sequences padded or resized to a maximum training length of approximately 384 frames (Figure 4).

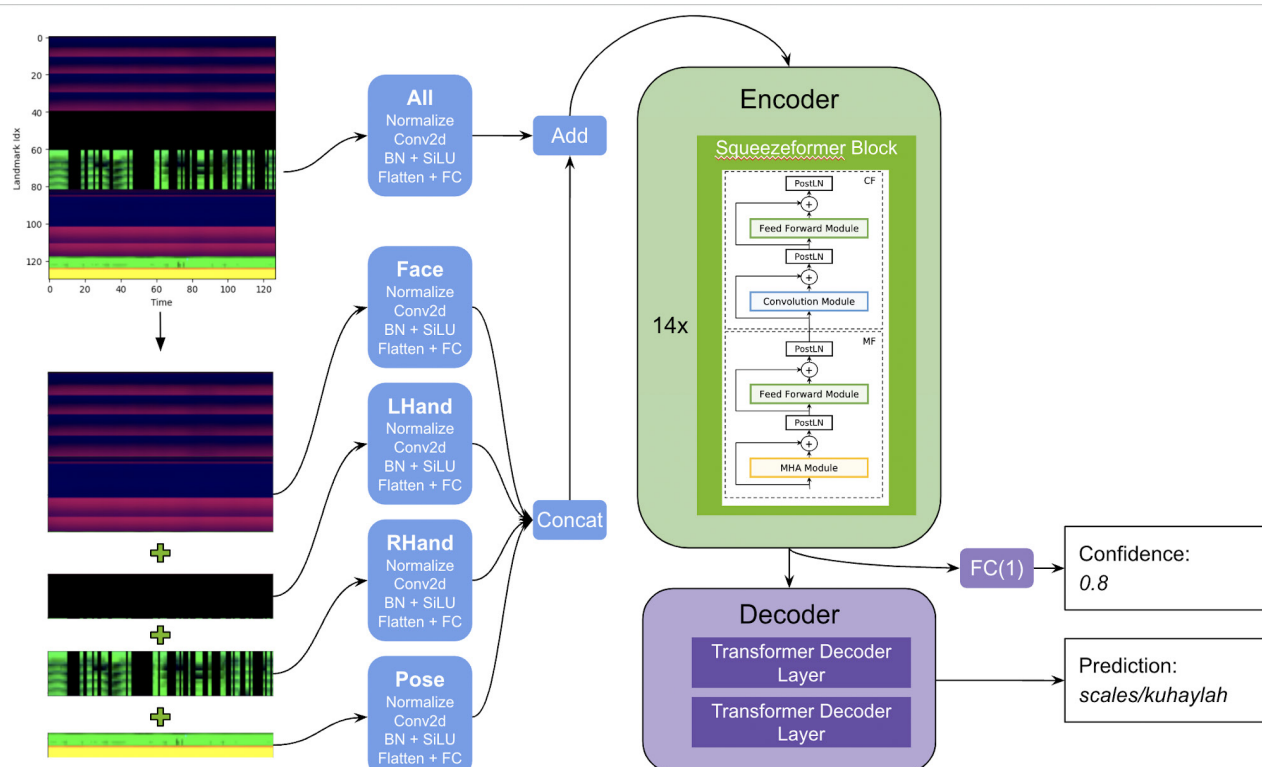


Figure 4. Multi-Stream Squeezeformer-Transformer architecture, taken from the source thesis⁴

The encoder stacks 14 Squeezeformer blocks, each combining multi-head self-attention (using rotary positional embeddings), a depth-wise convolutional module for local temporal patterns, and position-wise feed-forward layers in a macaron-like arrangement with residual connections and post-layer normalization. A two-layer Transformer decoder generates the output text via masked self-attention and cross-attention to the encoder output. An auxiliary linear head predicts a confidence score from the encoder representation.

Training used an Adam/AdamW optimizer with a cosine learning-rate schedule (approximately 400 epochs, a peak learning rate of 0.0045, and a 10-epoch warm-up), weight decay of approximately 0.08, dropout of approximately 0.1 in the encoder and decoder, mixed-precision (fp16) training, and attention masking for variable-length sequences. The primary loss was cross-entropy on the decoder output, with an auxiliary reversed-sequence cross-entropy term, and a regression loss (mean squared error or L1) trained the confidence head against a normalized edit-distance target.

Performance was assessed using accuracy, inference latency, throughput, low-light accuracy, confusion-

³ Source: Author's own work.

⁴ Source: Author's own work.

matrix analysis, and where applicable to sequence output - edit distance / CER / WER and F1, following standard practice for classification and sequence-evaluation reporting [9]. The prototype was tested on 120 pre-recorded sign sequences under controlled lighting and 300 real-time samples under varied conditions. Latency was measured with Python's `timeit` and JavaScript's `performance.now()`; accuracy and F1 were computed with `scikit-learn` [10]. Table 1 summarizes the evaluation setup (Table 1).

Table 1. Summary of dataset, representation, and evaluation protocol⁵

Item	Value / Description
Gesture classes	500+ isolated UzSL signs (greetings, pronouns, questions, verbs, emotions, numbers)
Recording setup	HD webcams, 1280×720 px, 30 fps, mixed indoor/outdoor, varied lighting and clothing
Landmark representation	130 keypoints/frame; 5-stream extraction (all, face, left hand, right hand, pose)
Split	70% train / 15% validation / 15% test
Test sets	120 pre-recorded sequences (controlled lighting) + 300 real-time samples (varied conditions)
Usability study	n = 10 (project team + coursemates); mean satisfaction 4.6 / 5

A web client (Next.js, Tailwind CSS, Radix UI) captured live webcam input in-browser and streamed frames asynchronously to a FastAPI backend over WebSockets; the server hosted the recognition model and returned predictions to the client in real time. The front-end was evaluated for screen-reader and keyboard-only navigation against the Web Content Accessibility Guidelines (WCAG) 2.1 [11]. Text-to-speech used the Google Cloud Text-to-Speech API with English and Uzbek voice options [12], and the server was containerized with Docker for deployment consistency. A companion Flutter mobile application, Signly.uz, offered gamified UzSL learning with offline recognition for a subset of signs via TensorFlow Lite.

ANALYSIS AND RESULTS

Table 2 reports the comparative results from the source thesis. The Transformer model outperformed the CNN-LSTM baseline on every reported metric: higher accuracy, lower inference latency, higher throughput, and higher low-light accuracy (Table 2; Figure 5; 6).

Table 2. Comparative performance reported in the source thesis⁶

Metric	CNN-LSTM	Squeezeformer-Transformer
Accuracy	87%	91%
Reported inference latency	0.72 s	0.58 s
Throughput	65 gestures/min	78 gestures/min
Low-light accuracy	76%	83%

⁵ Source: Author's own work.

⁶ Source: Author's own work.

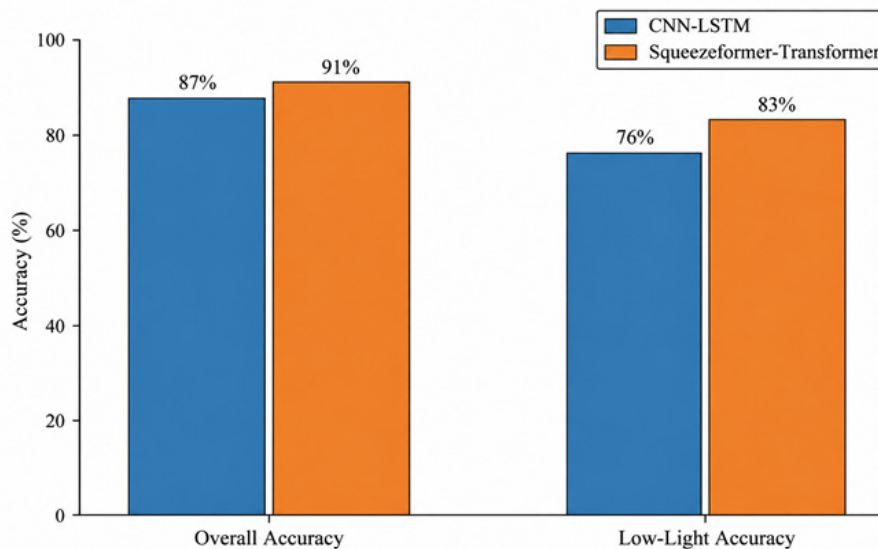


Figure 5. Overall and low-light recognition accuracy by architecture⁷

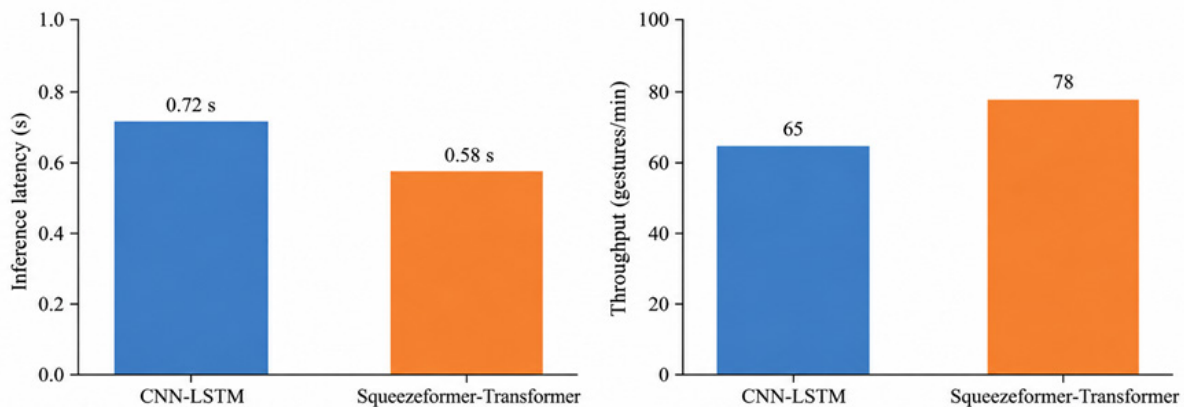


Figure 6. Reported inference latency and throughput by architecture⁸

Low-light conditions reduced model confidence for both architectures, motivating additional nighttime recordings and integration of MediaPipe-based hand segmentation [8]. Visually similar gestures, for example, the UzSL signs for “thank you” and “sorry” produced confusion in both models; this was partially mitigated by adding training examples for the affected classes and introducing temporal confidence thresholds. Both architectures degraded when a signer’s hands left the webcam’s field of view, pointing to a need for improved hand tracking and on-screen user guidance.

Ten participants took part in a controlled usability study: project team members who had learned a subset of UzSL signs, and coursemates who were instructed on a set of predefined signs the model could recognize. Participants rated ease of use, responsiveness, and translation output; mean satisfaction was 4.6 out of 5, and the interface and voice-feedback features were generally well received. A few participants described the synthesized voice as somewhat robotic and suggested that a more expressive TTS model, such as Tacotron 2, could sound more natural [13]. Because the participant group was small and drawn largely from people connected to the project rather than fluent UzSL signers or deaf and hard-of-hearing community members, this result is better read as a prototype usability signal than a representative usability claim.

The prototype demonstrated an end-to-end path from webcam capture through landmark processing, model inference, text output, and optional speech synthesis. Caching repeated text-to-speech phrases reduced average synthesis latency by 40%. The companion mobile app achieved acceptable recognition on its tested subset of signs but showed longer initial load times on lower-end Android devices; its gamified structure (XP, streaks, quizzes) was associated with repeated daily engagement during testing, consistent with literature on

⁷ Source: Author’s own work.

⁸ Source: Author’s own work.

gamification's motivational effects on learning [14].

The main finding is that the Squeezeformer-Transformer architecture outperformed the CNN-LSTM baseline across every reported metric: a four-percentage-point accuracy gain (91% vs. 87%), a latency reduction from approximately 0.72 s to 0.58 s, a throughput increase from 65 to 78 gestures/min, and a low-light accuracy improvement from 76% to 83%. These results are consistent with the general rationale for attention-based sequence modeling, in which self-attention captures relationships across a sequence without relying solely on recurrent state propagation.

The results should nonetheless be read within the boundaries of the experiment. Both models were evaluated on a proprietary, project-collected dataset whose exact per-class video counts, signer-level split strategy, and hardware configuration are not fully specified in the source thesis. If frame or sequence splitting was performed randomly rather than by signer, samples from the same signer may appear in both training and test subsets, which can overestimate generalization to unseen signers. A stronger benchmark for a future version of this work would enforce signer-independent train/validation/test partitions and report class-balanced precision, recall, macro-F1, confidence intervals, and per-signer breakdowns.

The dataset itself is a contribution independent of model architecture, since the shortage of structured UzSL data is a bottleneck for the field regardless of which model is used. However, 500+ gesture classes does not by itself guarantee linguistic or demographic coverage: regional signing variation, signer-specific movement, occlusion, and specialized vocabulary remain open limitations, and the current system addresses isolated words rather than continuous, sentence-level signing. The reported 91% accuracy should therefore be understood as performance within this isolated-sign evaluation setting, not as a claim about unrestricted natural UzSL conversation.

From a deployment standpoint, relying on ordinary webcams and software-only processing avoids specialized gloves or depth sensors, and separating capture from inference in the web architecture supports future scaling. At the same time, server-side inference introduces a dependency on connectivity and compute resources; the TensorFlow Lite mobile path points toward an edge-deployment alternative, but additional benchmarking is needed to establish whether the full pipeline can run efficiently on low-end devices.

The 4.6/5 usability result is encouraging for an early prototype but, should not be generalized. A more representative study would include fluent UzSL signers, deaf and hard-of-hearing users, hearing users with varying sign-language knowledge, and participants from different regions of Uzbekistan, together with task-completion time, per-participant error rate, and standardized usability instruments.

Relative to other camera-based sign-language tools, such as commercial hand-tracking SDKs (e.g., SignAll) [15], the system's main point of differentiation is language coverage: it is purpose-built for UzSL rather than adapted from an ASL-first product, which matters for a language with essentially no existing commercial support. This localization value should be weighed against the fact that commercial SDKs have typically been validated on far larger and more diverse datasets than the one used here.

Several limitations should be addressed before broader publication or deployment:

Isolated-sign scope: the system recognizes individual signs rather than continuous, sentence-level UzSL, which is linguistically and grammatically distinct from a word-by-word gloss.

Dataset availability and documentation: the dataset is proprietary and not publicly released; exact class distribution, per-class sample counts, and signer demographics require fuller documentation.

Usability sample: the study included only ten participants, drawn largely from the project team and coursemates rather than the deaf and hard-of-hearing community the system is intended to serve.

Robustness: background interference and hands leaving the camera frame remain sources of error for both architectures.

Generalization: regional and signer-specific variation can reduce accuracy for signers outside the training distribution, and this has not been evaluated with a signer-disjoint split.

Future work should prioritize: (1) signer-independent evaluation with disjoint train/validation/test splits by signer identity; (2) dataset expansion with regional variants and specialized (e.g., medical, legal, educational) vocabulary; (3) continuous sign segmentation and sentence-level translation, drawing on continuous sign-language translation research [16] and CTC-based alignment techniques [17]; (4) more rigorous reporting of CER/WER, macro-F1, and confidence intervals; (5) stronger 2D/3D hand and pose tracking (e.g., OpenPose) to reduce field-of-view errors [18], together with facial-expression recognition (e.g., via OpenFace) to capture the non-manual component of signing [19]; (6) model compression and quantization for efficient edge deployment [20]; and (7) sustained collaboration with UzSL linguists and deaf and hard-of-hearing community members on annotation quality, linguistic validity, and external evaluation, in line with best practice in deaf-education research [21].

CONCLUSIONS AND SUGGESTIONS

This study presents a camera-based UzSL recognition pipeline and a direct comparison of CNN-LSTM and Squeezeformer-Transformer architectures on a proprietary, 500+ class landmark dataset. Within the reported experimental setting, the Transformer model achieved 91% accuracy versus 87% for CNN-LSTM, together with lower latency, higher throughput, and stronger low-light accuracy, and the accompanying prototype demonstrated a complete pipeline from webcam capture to spoken output. These findings support attention-based sequence modeling as a promising direction for low-resource sign-language recognition, but they do not establish readiness for unrestricted conversational translation. The most important next step is not simply increasing model complexity, but building a larger, better-documented, signer-independent, and linguistically validated UzSL dataset, and evaluating continuous signing under realistic conditions for which this study provides a practical, reproducible starting point.

Participants in the source project were briefed on the recording protocol and consented to academic and research use of their recordings. The dataset is proprietary to the project team and is not released with this manuscript.

References:

1. J. Ergashev, M. Rakhmatova and M. Eshimov, "Sign Language Translation Model," B.S. thesis, Dept. of Computer Science, Central Asian University, Tashkent, Uzbekistan, 2025.
2. W. Tak, V. Arora, V. Chambyal, V. Kumari, Sumitra, and S. Chug, "A Dual-Path Deep Learning Framework for Static and Dynamic Sign Language Recognition Using CNN and LSTM Architectures," *Procedia Computer Science*, vol. 283, pp. 2557–2566, 2026, doi: 10.1016/j.procs.2026.06.325.
3. H. Luqman and E. Elalfy, "Utilizing motion and spatial features for sign language gesture recognition using cascaded CNN and LSTM models," *Turkish Journal of Electrical Engineering and Computer Sciences*, vol. 30, no. 7, pp. 2508–2525, 2022, doi: 10.55730/1300-0632.3952.
4. L. Pigou, S. Dieleman, P.-J. Kindermans, and B. Schrauwen, "Sign Language Recognition Using Convolutional Neural Networks," in *Computer Vision – ECCV 2014 Workshops*, 2015, pp. 572–578, doi: 10.1007/978-3-319-16178-5_40.
5. Vaswani, N. Shazeer, N. Parmar, et al., "Attention Is All You Need," in *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 30, 2017.
6. M. De Coster, M. Van Herreweghe, and J. Dambre, "Sign Language Recognition with Transformer Networks," in *Proceedings of the Twelfth Language Resources and Evaluation Conference (LREC 2020)*, Marseille, France, 2020, pp. 6018–6024.
7. M. Mukushev, A. Ubingazhibov, A. Kydyrbekova, A. Imashev, V. Kimmelman, and A. Sandygulova, "FluentSigners-50: A signer independent benchmark dataset for sign language processing," *PLOS ONE*, vol. 17, no. 9, e0273649, 2022, doi: 10.1371/journal.pone.0273649.
8. C. Lugesesi, J. Tang, H. Nash, et al., "MediaPipe: A Framework for Building Perception Pipelines," *arXiv preprint arXiv:1906.08172*, 2019.
9. M. Sokolova and G. Lapalme, "A Systematic Analysis of Performance Measures for Classification Tasks," *Information Processing & Management*, vol. 45, no. 4, pp. 427–437, 2009.
10. F. Pedregosa, G. Varoquaux, A. Gramfort, et al., "Scikit-learn: Machine Learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.
11. World Wide Web Consortium (W3C), "Web Content Accessibility Guidelines (WCAG) 2.1," 2018.
12. Google Cloud, "Text-to-Speech Documentation," accessed May 2025.
13. J. Shen, R. Pang, R. J. Weiss, et al., "Natural TTS Synthesis by Conditioning WaveNet on Mel Spectrogram Predictions," *arXiv preprint arXiv:1712.05884*, 2018.
14. J. Hamari, J. Koivisto, and H. Sarsa, "Does Gamification Work? A Literature Review of Empirical Studies," in *Proc. Hawaii Int. Conf. System Sciences (HICSS)*, pp. 3025–3034, 2014.
15. SignAll, "SignAll SDK for Sign Language Recognition," 2023.
16. N. C. Camgoz, S. Hadfield, O. Koller, H. Ney, and R. Bowden, "Neural Sign Language Translation," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, 2018, pp. 7784–7793.
17. Graves, S. Fernández, F. Gomez, and J. Schmidhuber, "Connectionist Temporal Classification: Labelling Unsegmented Sequence Data with Recurrent Neural Networks," in *Proc. ICML*, pp. 369–376, 2006.
18. Z. Cao, G. Hidalgo, T. Simon, S.-E. Wei, and Y. Sheikh, "OpenPose: Realtime Multi-Person 2D Pose Estimation Using Part Affinity Fields," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 1, pp. 172–186, 2021.
19. T. Baltrušaitis, P. Robinson, and L.-P. Morency, "OpenFace: An Open Source Facial Behavior Analysis Toolkit," in *Proc. WACV*, 2016, pp. 1–10.
20. TensorFlow, "Post-training quantization," *TensorFlow Lite Documentation*. [Online]. Available: TensorFlow documentation.
21. B. Woll and G. Morgan, *Language Learning and Language Teaching in Deaf Education*, Cambridge University Press, 2012.

Proofreader: Zokir ALIBEKOV

Layout and Designer: Oloviddin Sobir ugli

2026. № 9

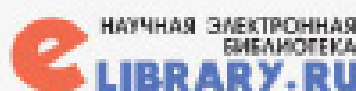
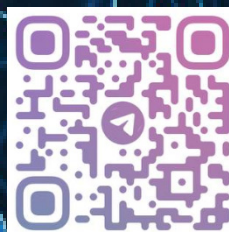
© When materials are reproduced, the INNOVATION SCIENCE AND TECHNOLOGY journal must be cited as the source. Authors are responsible for the accuracy of the information in materials and advertisements published in the journal. Editorial opinions may not always align with those of the authors. Submitted materials will not be returned to the editorial office.

To publish articles in this journal, you may submit articles, advertisements, stories, and other creative materials through the following links. Materials and advertisements are published on a paid basis.

You may subscribe to the journal at any time using the following details. Once subscribed, please send a screenshot or photo of your payment confirmation to our Telegram page @iqtisodiyot_77. Based on this, we will send the latest issue of the journal to your address each month.

“The journal “INNOVATION SCIENCE AND TECHNOLOGY” has been registered by the Agency for Information and Mass Communications under the Administration of the President of the Republic of Uzbekistan from 09.10.2024 under the registration number №390637. License number: C-5669633. PNFL: 30407832680027

Our address: Tashkent city, Yunusobod district, 19th block,
House 17.



Acceptance of articles

Published every
monthly



Directions

Social, economic, political,
technological, scientific



Scopus || Scientific electronic journal specializing in Scopus

CERTIFICATE NUMBER: №390637

**ORDER NUMBER ACCORDING TO
THE LICENSE REGISTER: C-5669633**

CONTACT:



Contact us
+998 50 737 87 88



Telegram channel
t.me/scopus_IST2100



Journal official website
<https://ist-journal.uz/index.php/IST>