УДК: 004.45

*[1]M.B. Polat, [2]Zh.E. Temirbekova

*[1]Kazakh-British Technical University, Almaty, Kazakhstan;*
*[2]Al-Farabi Kazakh National University, Almaty, Kazakhstan.*
**E-mail: m_polat@kbtu.kz*

# A COMPARATIVE STUDY OF SIGN GESTURE RECOGNITION MODELS IN MEDICAL CONTEXT WITH MEDIA PIPE

**Polat Moldir Bakhytzhankyzy,** master student.
E-mail: m_polat@kbtu.kzORCID ID: 0009-0003-6227-4891
**Temirbekova Zhanerke Erlanovna,** PhD, assistant-professor.
E-mail: temyrbekovazhanerke2@gmail.com; ORCID ID: 0000-0003-3909-0210

By 2050, approximately 2.5 billion people worldwide will experience some degree of hearing loss, with over 700 million requiring hearing rehabilitation. This growing prevalence emphasizes the urgent need to address communication barriers faced by the deaf and hard-of-hearing population, particularly in healthcare settings. In Kazakhstan, the lack of effective tools for interpreting Kazakh Sign Language (KSL) complicates communication between patients and medical professionals. This study aims to improve healthcare accessibility by developing and evaluating dynamic sign gesture recognition models using MediaPipe for preprocessing. A key contribution of the research is the creation of the KMSG11 dataset, which includes health-related vocabulary in KSL. The proposed Long Short-Term Memory (LSTM)-based model was compared with other existing models, all trained and tested on two datasets—KMSG11 and the Argentinian Sign Language dataset (LSA20)—under the same preprocessing conditions. The results demonstrate the potential of the model to enhance communication, diagnosis, and medical services for deaf individuals in Kazakhstan.

***Keywords**: Kazakh sign language, sign gesture recognition, hard-of-hearing people, doctor-patient communication, neural networks, long short-term memory (LSTM), MeidaPipe, keypoints extraction.*

*[1]М.Б. Полат, [2]Ж.Е. Темирбекова

*[1]Казахско-Британский технический университет, Алматы, Казахстан;*
*[2]Казахский национальный университет имени аль-Фараби, Алматы, Казахстан.*
*E-mail: m_polat@kbtu.kz

# СРАВНИТЕЛЬНОЕ ИССЛЕДОВАНИЕ МОДЕЛЕЙ РАСПОЗНАВАНИЯ ЖЕСТОВ В МЕДИЦИНСКОМ КОНТЕКСТЕ С MEDIAPIPE

**Полат Молдир Бахытжанкызы, м**агистрант;
E-mail: m_polat@kbtu.kz ; ORCID ID: 0009-0003-6227-4891
**Темирбекова Жанерке Ерлановна,** PhD, доцент,
E-mail: temyrbekovazhanerke2@gmail.com; ORCID ID: 0000-0003-3909-0210

Ожидается, что к 2050 году около 2,5 миллиарда человек будут иметь ту или иную степень потери слуха, при этом не менее 700 миллионов из них будут нуждаться в реабилитации слуха. Увеличение числа таких пациентов подчеркивает актуальность решения проблем коммуникации, с которыми сталкиваются глухие и слабослышащие люди, особенно в таких критически важных сферах, как здравоохранение.

В Казахстане пациенты с нарушениями слуха испытывают трудности при общении с медицинским персоналом из-за ограниченности или отсутствия эффективных инструментов перевода казахского языка жестов (KSL). Настоящее исследование направлено на устранение этой проблемы путем сравнения различных моделей распознавания динамических жестов с применением технологии MediaPipe для предварительной обработки данных. Основным вкладом работы стало создание нового набора данных KMSG11, содержащего термины, связанные со здоровьем и медициной на казахском языке жестов. Разработанная модель, основанная на архитектуре долговременной кратковременной памяти (LSTM), была сопоставлена с моделями из других научных работ. Все модели были обучены и протестированы на двух наборах данных – KMSG11 и аргентинском наборе данных языка жестов (LSA20) – с использованием одинаковой методики предварительной обработки через MediaPipe, что обеспечило объективность сравнения. Процесс включал подготовку видеоданных, обучение моделей и оценку их эффективности. Полученные результаты демонстрируют потенциал предложенного подхода для улучшения коммуникации, диагностики и лечения, а также повышения доступности качественной медицинской помощи для глухих пациентов в Казахстане.

***Ключевые слова:*** *казахский язык жестов, распознавание жестов, люди с нарушениями слуха, коммуникация врача и пациента, нейронные сети, долговременная кратковременная память (LSTM), MediaPipe, извлечение ключевых точек.*

***[1]М.Б. Полат, [2] Ж.Е. Темирбекова**

*[1]Қазақ-Британ техникалық университеті, Алматы, Қазақстан;*
*[2]әл-Фараби атындағы Қазақ ұлттық университеті, Алматы, Қазақстан.*
*E-mail: m_polat@kbtu.kz

# MEDIAPIPE КӨМЕГІМЕН МЕДИЦИНАЛЫҚ КОНТЕКСТЕ ҚИМЫЛДЫ ТАНУ ҮЛГІЛЕРІН САЛЫСТЫРМАЛЫ ЗЕРТТЕУ

**Полат Молдір Бахытжанқызы, м**агистрант;
E-mail: m_polat@kbtu.kz ; ORCID ID: 0009-0003-6227-4891
**Темирбекова Жанерке Ерланқызы,** PhD, доцент,
E-mail: temyrbekovazhanerke2@gmail.com; ORCID ID: 0000-0003-3909-0210

2050 жылға қарай әлемде шамамен 2,5 миллиард адамда есту қабілетінің белгілі бір дәрежеде жоғалуы байқалады деп күтілуде, олардың кем дегенде 700 миллионы есту қабілетін қалпына келтіруді қажет етеді. Мұндай көрсеткіштердің өсуі саңырау және нашар еститін адамдардың, әсіресе денсаулық сақтау сияқты маңызды салаларда, кездесетін коммуникациялық қиындықтарын шешудің өзектілігін көрсетеді. Қазақстанда есту қабілеті бұзылған науқастардың дәрігерлермен тиімді қарым-қатынас орнатуы көбіне қиындық тудырады, себебі қазақ ым тілі (KSL) үшін аударма құралдары шектеулі немесе мүлдем жоқ. Бұл зерттеу пациент пен дәрігер арасындағы байланысты жақсарту мақсатында MediaPipe технологиясын қолдану арқылы динамикалық ым-қимылдарды тануға арналған түрлі модельдерді салыстыруға бағытталған. Зерттеудің маңызды нәтижесі – қазақ ым тілінде денсаулық пен медицинаға қатысты терминдерді қамтитын KMSG11 атты жаңа деректер жиынын әзірлеу. Ұсынылған Long Short-Term Memory (LSTM) моделінің тиімділігі басқа ғылыми жұмыстардағы үлгілермен салыстырылды. Барлық модельдер бірдей алдын ала өңдеу әдісімен – MediaPipe көмегімен – екі деректер жиынында: KMSG11 және аргентиналық ым тілі деректер жиынында (LSA20) оқытылып, тестілеуден өтті. Зерттеу бейне деректерді дайындау, модельдерді оқыту және олардың нәтижелерін бағалауды қамтыды. Жұмыстың нәтижелері Қазақстандағы есту қабілеті бұзылған азаматтардың сапалы медициналық қызметке

қолжетімділігін арттыруға, сондай-ақ денсаулық сақтау жүйесіндегі коммуникацияны, диагностиканы және емдеуді жетілдіруге бағытталған.

***Түйін сөздер:*** *қазақ ымдау тілі, ым қимылдарын тану, нашар еститін адамдар, дәрігер мен пациент байланысы, нейрондық желілер, ұзақ қысқа мерзімді жады (LSTM), MeidaPipe, түйінді нүктелерді шығару.*

## INTRODUCTION

It is estimated that by 2050, around 2.5 billion individuals will have some form of hearing loss, and at least 700 million of them would need hearing rehabilitation. Furthermore, because of risky listening habits, nearly 1 billion young individuals could experience irreversible, avoidable hearing loss [20]. The increasing number of people experiencing hearing loss highlights the importance of paying close attention to the communication obstacles that deaf or hard-of-hearing people encounter, particularly in important contexts like healthcare.

One of the main problems in Kazakh healthcare settings is the communication gap between deaf or hard-of-hearing patients and healthcare professionals. The available sign language interpretation possibilities are sometimes insufficient or nonexistent, which does not satisfy the demands of Kazakh sign language users. The lack of efficient communication tools makes it more difficult for patients to receive high-quality healthcare by preventing optimal diagnosis, treatment, and overall patient satisfaction.

Recognizing the urgent need for a reliable solution, this study aims to provide a thorough comparative analysis of dynamic sign gesture recognition models customized to the unique needs of patient-doctor communication in Kazakhstan. We try to find the most accurate and effective way to close the communication gap by analyzing several models. Our mission is to improve deaf and hard-of-hearing patients' healthcare experiences and make Kazakhstan's healthcare system more welcoming.

The field of sign language recognition has witnessed substantial advancements over the past few years, encouraged by the integration of deep learning technologies and the increasing need for accessible communication solutions across various domains, including healthcare. Researchers mainly focused on developing models that can accurately recognize sign language gestures, translating them into spoken language to facilitate communication between the deaf or hard-of hearing individuals and the hearing population. These efforts have spanned across different sign languages, including Saudi sign language (SSL) [5], Arabic sign language [13], Mexican sign language (MSL) [17], Chinese sign language (CSL) [4], [8] and Indian sign language (ISL) [10],[15], each presenting unique challenges and insights.

From an implementation perspective scientists have come up with different models in hand gesture recognition systems. From these works [1], [11], [3], [4] it is notable that one of the core methods is Convolutional Neural Networks (CNN) and Motion Networks, because it is possible to extract and learn from spatial features in images and videos. The integration of Dynamic and Accumulative Motion Networks further enhances the recognition process by focusing on the temporal aspects of gestures, making it possible to interpret continuous sign language sequences with higher accuracy [11]. Scale Invariant Feature Transform (SIFT) alongside CNNs has been used for feature extraction in sign language images in this work [3]. Some researchers employed advanced architecture like Convolutional Block Attention Module (CBAM)-ResNet and 3D Residual Networks (3D-ResNet) [4], [14] for enhancing recognition accuracy. These models have shown promising results in various sign languages, emphasizing the potential for broad application across different linguistic contexts.

Studies such as the development of the Saudi deaf companion system (SDCS) [5] and a prototype for MSL recognition in healthcare settings [17] highlight the evolution towards two-way communication systems. These systems not only recognize sign language but also provide mechanisms for translating spoken language into sign language, thus enabling bidirectional communication between deaf and hearing individuals. The opposite direction of translating dialogues between doctors and patients is a crucial part.

The introduction of transformer models, specifically through the SIGNFORMER architecture [10], represents a novel approach in the field. By employing a vision transformer to recognize static signs, this model demonstrates the efficacy of transformer architecture in handling the complexity of sign language, offering an alternative to traditional CNN approaches. The SIGNFORMER's use of positional embedding patches and self-attention layers allows for high accuracy with fewer training epochs, proving the potential for rapid and efficient model training on sign language datasets.

There are several studies that focus on Kazakh sign language (KSL) recognition as well. This paper [1] presents a method using Convolutional Neural Networks (CNNs) to recognize dynamic gestures in Kazakh sign language in real-time. Researchers in this study [2] explore a system for continuous KSL recognition, translating signs into intonation-colored speech. Focusing on static gesture recognition, this research [9] authors developed a cost-effective prototype using an RGB mono camera for Kazakh sign language. It offers a novel approach to recognizing static gestures without the need for expensive equipment, demonstrating promising results for broader application.

While referenced studies demonstrate significant results in sign language recognition and its application, several gaps remain, particularly concerning the adaptation of these models to Kazakh sign language and the specific context of patient-doctor communication in Kazakhstan. First, many of the developed systems are adjusted for specific languages, such as Saudi sign language or Mexican sign language, without consideration for adaptation to other sign languages like Kazakh sign language. This limitation raises questions about the ability to transfer and scalability of these models across different linguistic contexts. The second aspect is that although some studies have begun to explore the application of sign language recognition in healthcare settings, there is a need for deeper integration of medical terminologies and scenarios specific to patient-doctor communication. The models need to be trained on domain-specific datasets to improve accuracy and effectiveness in real-world medical consultations. Finally, practical challenges, such as sensor heating, environmental constraints, and the need for continuous hours of operation, exist as limiting factors in the effective deployment of these technologies in healthcare settings. Moreover, we assume that the necessity to modify the recording environment to minimize disturbances from natural lighting and maintain sensor performance indicates a gap in the robustness and user-friendliness of current solutions.

Our study offers two significant contributions to solve these limitations. To meet the unique needs of patient-doctor communication in Kazakhstan, we create and assess a thorough comparative analysis of dynamic sign gesture recognition models. We provide a new dataset of medical sign gestures in Kazakh that includes medical terminologies and health related wordsю

The methodology figure is shown in Figure 1, that describes all stages of this paper's work.

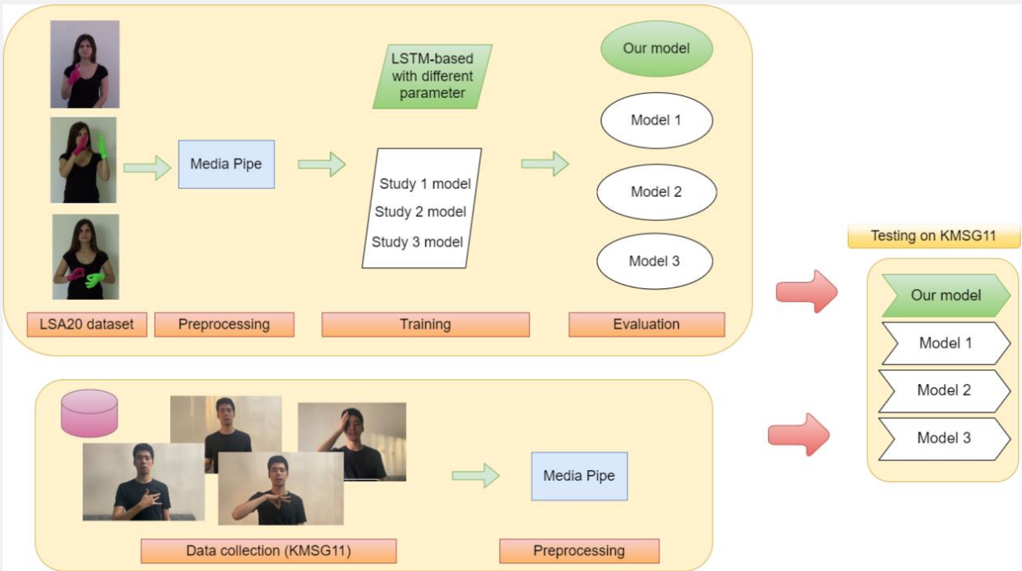



**Figure 1** - Methodology figure

Starting from preprocessing video recordings using Media Pipe, we trained different recognition models, collecting Kazakh sign language dataset in parallel. Afterwards, we conducted a comparison experiment of all final models on our dataset.

## MATERIALS AND METHODS

Dataset

Current research's experiments are carried on 2 datasets.

LSA64: A Dataset for Argentinian Sign Language [16] consists of 3200 video recordings in which 10 people performed 5 iterations of 64 distinct sign gestures. The most often used signs - both verbs and nouns - in the LSA vocabulary were chosen for the selection. In order to provide environment variations, one part of recording was made outside under natural lightning, and the other indoors under artificial lightning. All sign performers were in black clothes, they wore fluorescent-colored gloves to make the task of hand segmentation within a picture easier to understand. They were sitting or standing with a white wall background. We decreased a number of signs to 20 gestures (LSA20), and listed them in Table 1.

| № | List of words |
|---|---|
| 1 | Accept |
| 2 | Appear |
| 3 | Born |
| 4 | Buy |
| 5 | Call |
| 6 | Catch |
| 7 | Copy |
| 8 | Dance |
| 9 | Deaf |
| 10 | Food |
| 11 | Give |
| 12 | Help |
| 13 | Learn |
| 14 | Music |
| 15 | Name |
| 16 | Run |
| 17 | Shut down |
| 18 | Thanks |
| 19 | Trap |
| 20 | Water |

**Table 1** - LSA20: Argentinian medical sign gestures dataset

The second dataset is Kazakh medical sign gestures (KMSG11), that we collected specifically for current research. It includes 11 signs that are listed in the website Surdo.kz [18] in the section "Human" – "Health, medicine", because we are interested in doctor-patient communication, please see Table 2.

| № | Kazakh | English |
|---|---|---|
| 1 | Artikuliatsiia | Articulation |
| 2 | Dene | Body |
| 3 | Auru | Disease |
| 4 | Betjuzı | Face |
| 5 | Bas | Head |
| 6 | Emdeu | Heal |

| | | |
|---|---|---|
| 7 | Densaulyq | Health |
| 8 | Jurek | Heart |
| 9 | Auruhana | Hospital |
| 10 | Juqpaly auru | Infectious disease |
| 11 | Darı | Remedy |

**Table 2** - KMSG11: Kazakh medical sign gestures dataset

Overall, four non-signers in black clothes participated in dataset collection, everyone recorded one sign twenty times in natural as well as indoor artificial lightning environment, resulting in 880 video recordings in total.

## PREPROCESSING

We employed a Media Pipe to extract key points from video data in preprocessing stage. This stage is crucial for transforming raw video inputs into a structured format suitable for model training and evaluation. Our approach involved several key stages, including data loading, key point extraction, and sequence padding to ensure uniformity across the dataset.

To extract key points from the video files, we used Media Pipe's Pose solution [6]. We can identify landmarks of human bodies in video to classify the gesture and capture important hand movements using 33 body landmark locations (Figure 2).

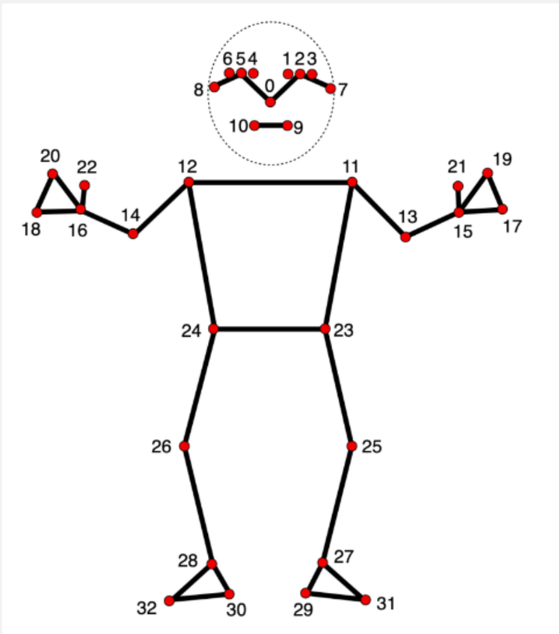



**Figure 2** - Media Pipe pose landmarks [6]

The inclusion of 3D coordinates $(x, y, z)$ for each keypoint provides a richer and more accurate representation of performed sign. Besides, this solution is widely applicable because it is built to work well on a variety of hardware platforms, including desktop computers and mobile devices.

Thus, the key points were saved in JSON format. During data preparation we read JSON files to flatten the key points of each frame into a single list, the resulting sequences in turn were padded and normalized for further training models. The key points and labels were split into two subsets: 80% for training and 20% for testing.

Sign gesture recognition models

1) Recurrent Neural Network models: We built a different RNN models to figure out the optimized method of recognition. Specifically, we compared two RNN models: Long Short-term Memory (LSTM) and Gated Recurrent Unit (GRU). For dynamic sign gesture recognition, the selection of LSTM and GRU models makes sense due to their capacity to efficiently capture long-term relationships and ability to capture temporal dependencies and patterns within sequential data.

Further we dived into one model - LSTM, and experimented with various parameters such as LSTM units, a number of layers, dropout rate, L2 regularization factor testing on LSA20 dataset of Argentinian sign language. Figure 3 captures the proposed LSTM architecture for dynamic sign gesture recognition.

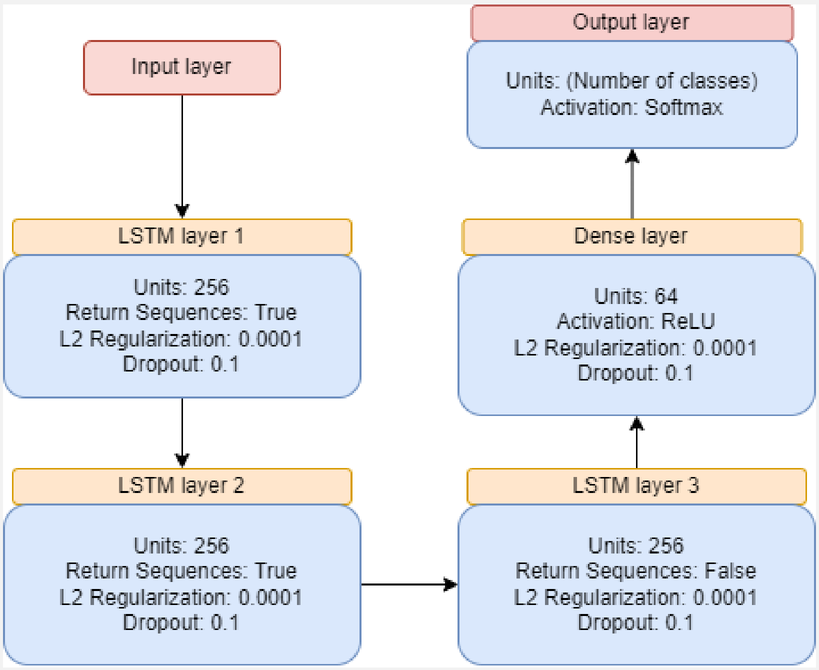



**Figure 3** - LSTM model architecture

The architecture consists of an input layer followed by three LSTM layers with dropout, a dense layer with dropout, and an output layer with Soft-max activation. The use of L2 regularization and dropout helps to prevent overfitting, making the model more robust.

Moreover, we conducted comparison between two optimizers: Adam and RMSprop.

2) Comparison with other models: To validate the effectiveness of our proposed LSTM-based dynamic sign gesture recognition model, we compared our results with those from three notable studies in the field testing on Argentinian sign language dataset. Important note is that a preprocessing stage of each study differs, and current research uses one method of data preparation for all models.

First study [1], that was chosen for comparison introduced a methodology for automatically collecting spatio-temporal features of gestures by calculating coordinates and normalizing them. They constructed an optimal multilayer perception for multiclass classification. The second study [7] found that LSTMs outperformed CNNs in recognizing dynamic gesture sign languages by monitoring hands, faces, and poses. Their LSTM model was used for experiments on LSA20 dataset. Last paper's [12] proposed system performs real-time gesture recognition on 5 static gestures from the American Sign Language (ASL) and gives precise, accurate, and efficient results, and it is LSTM-based improved model.

Evaluating models on Kazakh sign language dataset

During the last stage we finalized our model and 3 other models from a previous phase testing them on our dataset of Kazakh words that are related to human and medicine.

The main evaluation metrics of recognition systems were accuracy, precision, recall, and F1 score [19]. These metrics are commonly utilized in a machine learning domain to evaluate performance.

Precision: represents the proportion of predicted positive cases that are correctly identified as positive.

$$Precision = \frac{TP}{TP + FP}$$

where TP is the number of true positives, and F P is the number of false positives.

Recall: measures the accuracy with which the system correctly identifies positive cases among the actual positive cases The formula for the recall is:

$$Recall = \frac{TP}{TP + FN}$$

where FN is the number of false negatives.

F1-Score: expresses the balance between precision and recall. A higher F-score value signifies improved performance. The formula is:

$$F1\ Score = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

Confusion Matrix: gives us a thorough analysis of the model's predictions, helping us to see where the model is accurate and where it is inaccurate. The true classes are represented by rows in the confusion matrix, while the predicted classes are represented by columns. Every component in the matrix represents a number of instances categorized under a specific group.

## Results and Discussion

Initial testing of two RNN models are shown in Table 3. The LSTM model slightly outperformed the GRU model, thus it was the base architecture for our further experiments.

| | LSTM | GRU |
|---|---|---|
| Test accuracy | 71% | 70.5% |

**Table 3** - Initial testing of LSTM-GRU

Next, we carried out the training with various parameters results displayed in Table 4 and Table 5. The best result 91% was with these parameters: 256 LSTM units, three layers, a dropout rate of 0.3, a regularization factor L2 of 0.0001.

| LSTM units | Number of layers | Dropout rate | Test accuracy |
|---|---|---|---|
| 64 | 2 | 0.2 | 70.5% |
| 32 | 1 | 0.1 | 46% |
| 128 | 3 | 0.3 | 71.5% |
| 256 | 4 | 0.4 | 63.5% |
| 128 | 3 | 0.4 | 68.5% |
| 128 | 3 | 0.2 | 69.5% |
| 128 | 2 | 0.2 | 73% |
| 128 | 2 | 0.1 | 73% |
| 128 | 1 | 0.1 | 66.5% |
| 256 | 2 | 0.2 | 75.5% |
| 256 | 2 | 0.1 | 78% |
| 256 | 3 | 0.2 | 82% |
| **256** | **3** | **0.1** | **83%** |
| 256 | 4 | 0.2 | 74.5% |
| 256 | 4 | 0.1 | 79% |

**Table 4** - Results of experiments on 3 parameters

| LSTM units | Number of layers | Dropout rate | L2 regularization factor | Test accuracy |
|---|---|---|---|---|
| 256 | 3 | 0.1 | 0.01 | 82.5% |
| 256 | 3 | 0.1 | 0.005 | 86.5% |
| 256 | 3 | 0.1 | 0.001 | 87% |
| 256 | 3 | 0.001 | 0.005 | 84.5% |
| 256 | 3 | 0.3 | 0.001 | 86.5% |
| **256** | **3** | **0.1** | **0.0001** | **91%** |
| 256 | 4 | 0.1 | 0.001 | 87% |

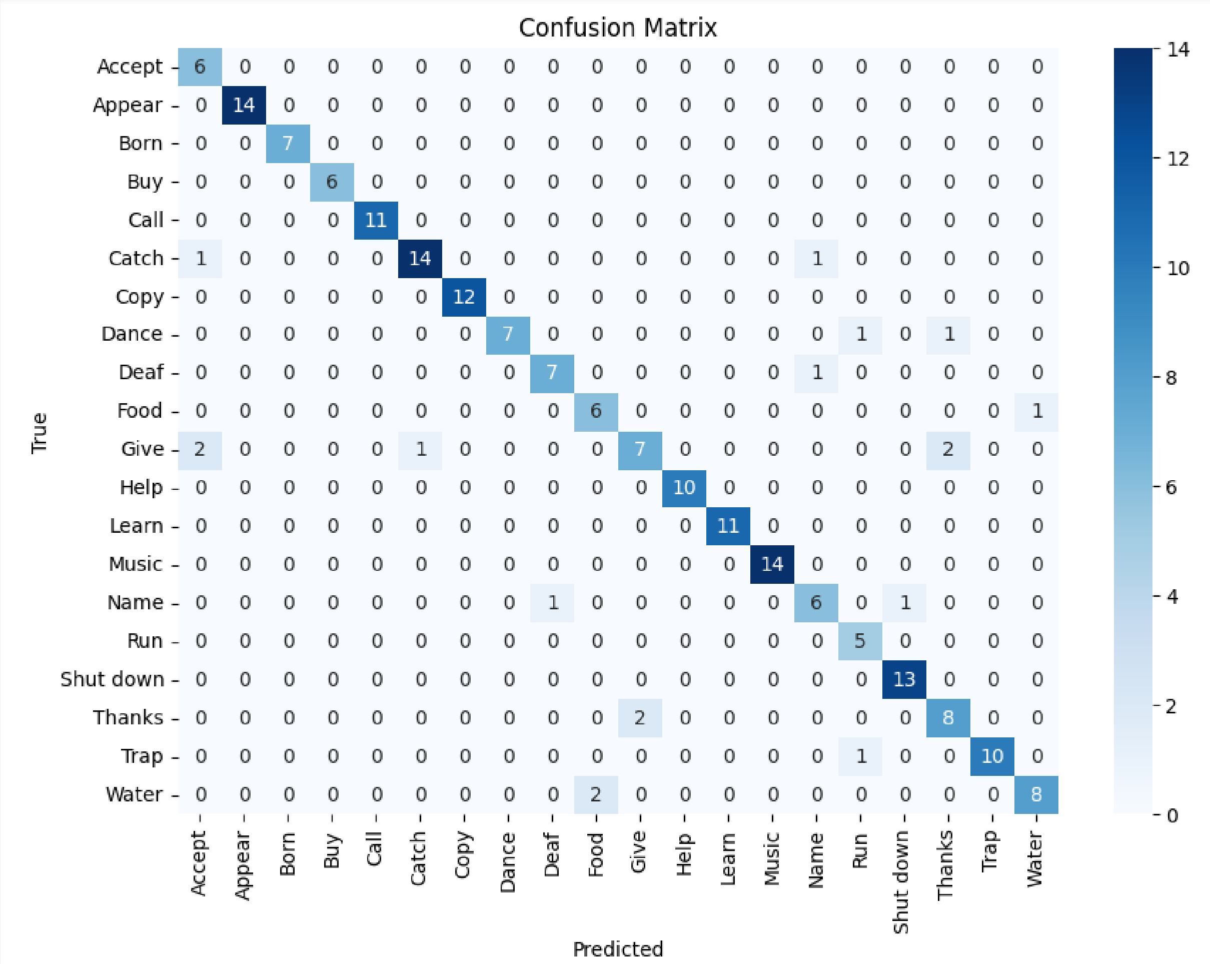
**Table 5** - Results of experiments on 4 parameters

Please see Table 6 to see the comparison between two optimizers, where Adam optimizer gave the better accuracy.

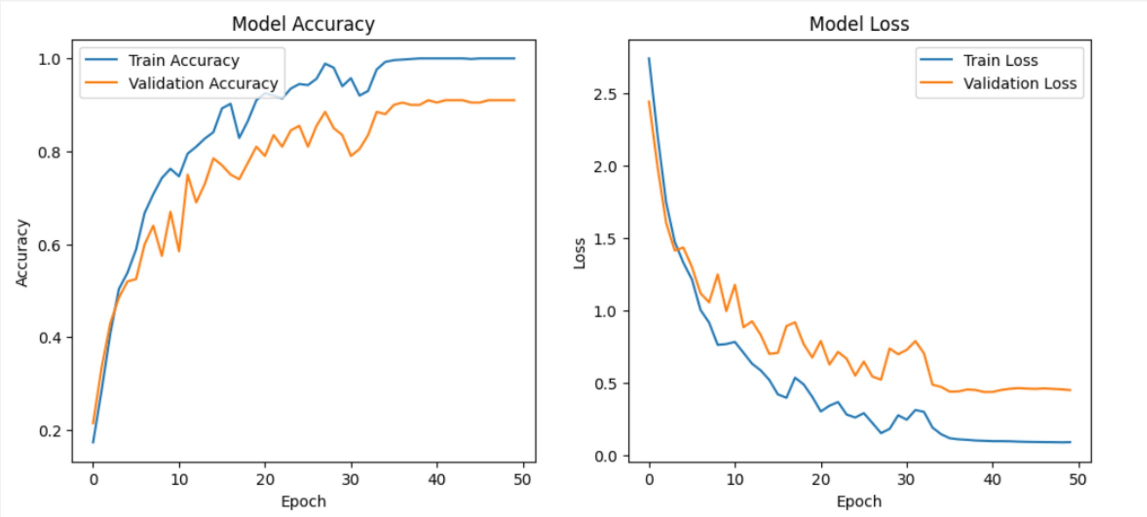
| Optimizer | Test accuracy |
|---|---|
| **Adam** | **91%** |
| RMSprop | 88.5% |

**Table 6** - Results of experiment on optimizers

Model loss, model accuracy graphs and a confusion matrix of our method trained on LSA20 dataset are depicted in Figure 4 and on Kazakh sign language in Figure 5.

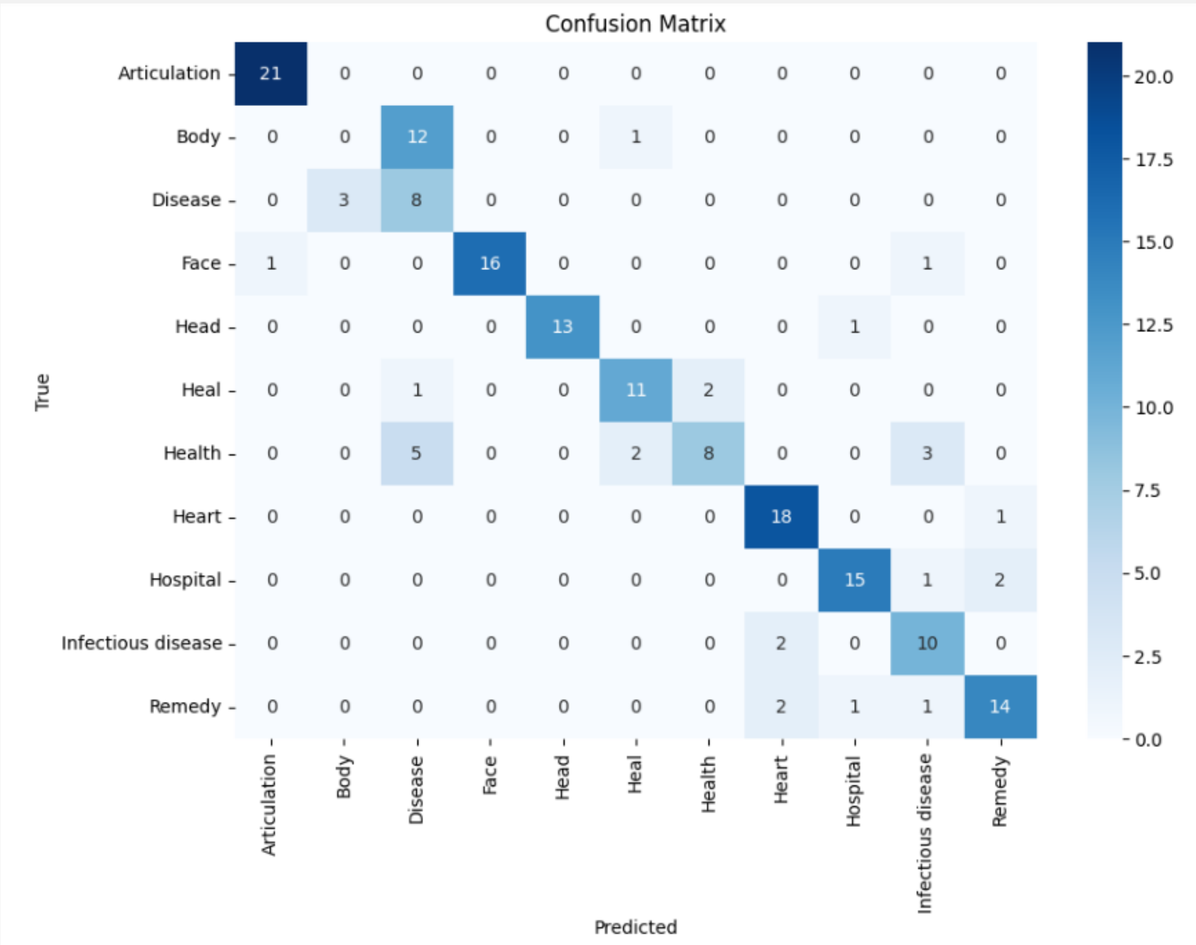



a. Confusion matrix

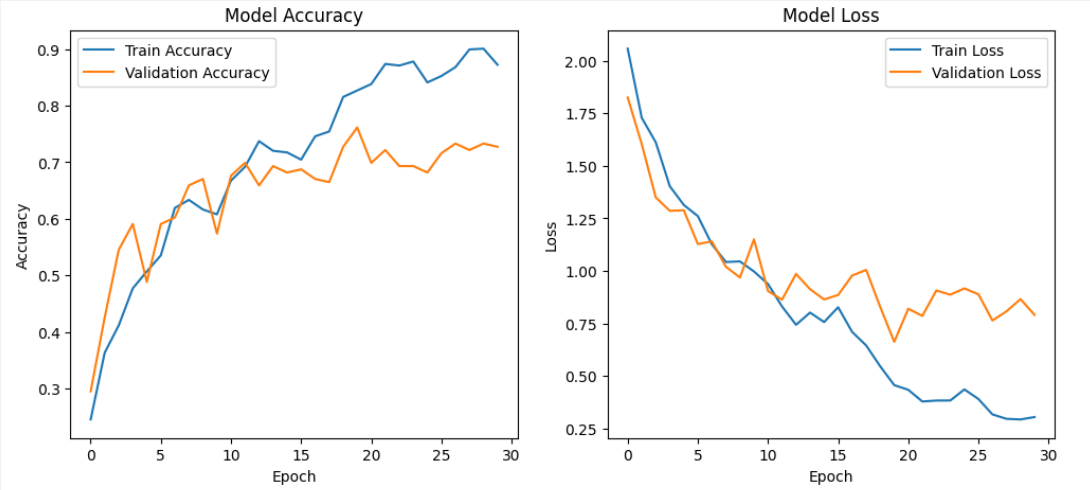



b. Accuracy and loss graphs

**Figure 4** - Results for LSA20




a. Confusion matrix




b. Accuracy and loss graphs

**Figure 5** - Results for KMSG11

It is notable that all models managed our KMSG dataset with lower precision in general. The possible reasons could include several aspects. Video recordings' resolution was comparatively low, and a number of frames per second was twice smaller than in LSA20 videos (60 frames per second). Next is that participants in our dataset did not wear fluorescent-colored gloves, which significantly helped to avoid a skin color variation issue during key points extraction in videos of Argentinian gestures. Another important factor is the dataset volume, the more data we have, the more accurate results we can get.

From Figure 6 that concludes the final comparison of 4 sign language recognition models in terms of test accuracy, precision, recall, and F1-score we can see that among 4 models the highest results belong to Model 2 Amangeldy et al.

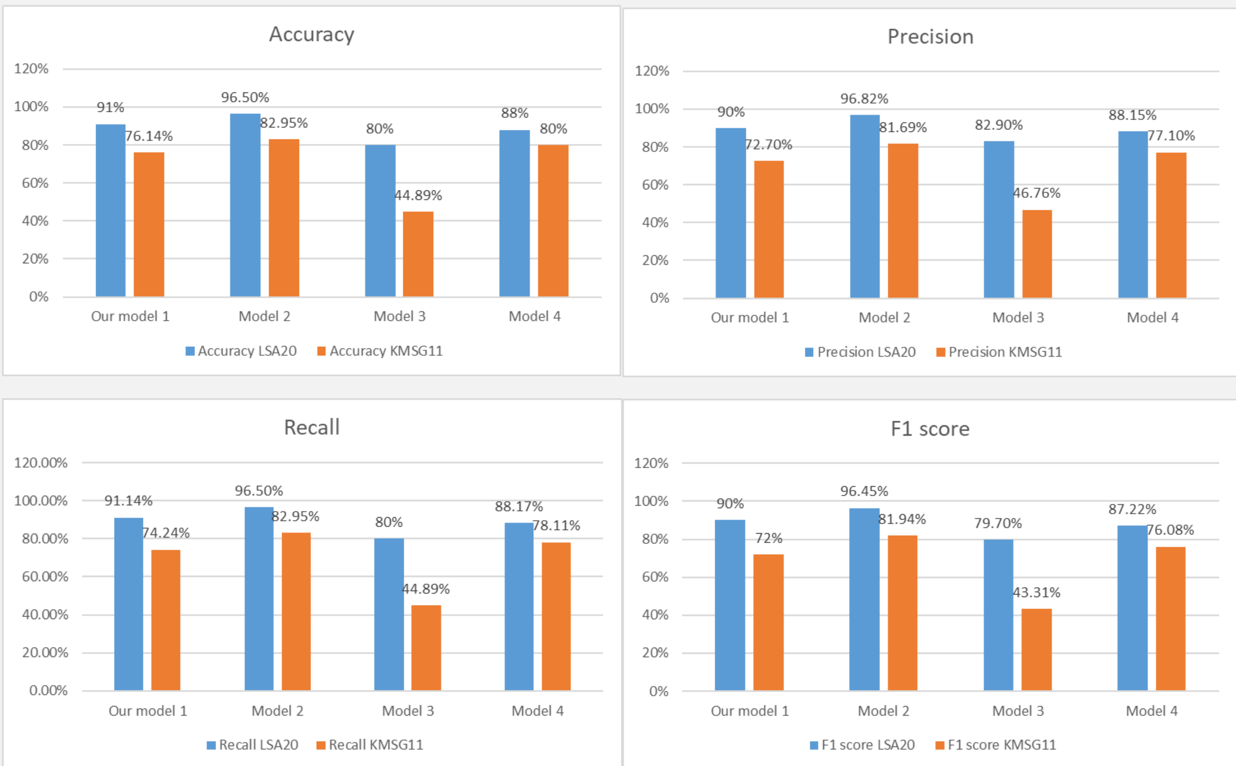



**Figure 6** - Four metrics comparison of 4 models (blue - LSA20, red - KMSG11)

Study [1] showing 96.5% and 82.95% accuracy for LSA20 dataset and KMSG11 dataset respectively. Our LSTM-based model with parameters performed well for Argentinian sign language, resulting in accuracy of 91%, however for our dataset it was inferior with 76.14% to the LSTM model of Model 4 Mirikar et al. [12] that achieved 80% accuracy. Model 3 was Goyal et al. study [7].

Regarding the application of the models trained on KMSG11 dataset, models can be integrated into the tools that facilitate hard-of-hearing people in hospitals when they want to get consultation or simply to see a doctor. Increasing dataset with more common words used in medicine and health state description, it is possible to get a promising recognition system for patient-doctor communication scenarios.

## CONCLUSION

In this research we have provided a thorough examination and comparison of dynamic sign gesture recognition model for Kazakh sign language in a medical setting. We examined the performance of different recurrent neural network models using Media Pipe during preprocessing stage and testing them on two datasets: the newly generated Kazakh Medical Sign Gestures (KMSG11) dataset and the Argentinian Sign Language dataset. On the smaller LSA20 dataset, the LSTM model with improved hyper-parameters demonstrated noteworthy accuracy of 91%. It's relatively poor performance on the KMSG11 dataset, however, emphasizes the critical role of dataset quality and volume, use of fluorescent-colored gloves in achieving high recognition accuracy.

The significance of this study lies in its focus on comparing and creating an effective recognition model that can be used in practice for improving communication in medical settings for the hard-of-hearing community in Kazakhstan. In addition to pointing out the advantages and disadvantages of the existing models, the comparative analysis provides a road map for future improvements, guaranteeing that the systems that are created are reliable and flexible enough for use in actual applications.

## REFERENCES

1 Amangeldy N., Milosz M., Kudubayeva S., Kassymova A., Kalakova G., Zhetkenbay L. "A real-time dynamic gesture variability recognition method based on convolutional neural networks," *Applied Sciences (Switzerland)*, vol. 13, Oct. 2023.

2 Amangeldy N., Ukenova A., Bekmanova G., Razakhova B., Milosz M., and Kudubayeva S. "Continuous sign language recognition and its translation into intonation-colored speech," *Sensors*, vol. 23, Jul. 2023.

3 Arooj S., Altaf S., Ahmad S., Mahmoud H., and Mohamed A. S. N. "Enhancing sign language recognition using CNN and SIFT: A case study on Pakistan sign language," *Journal of King Saud University – Computer and Information Sciences*, vol. 36, Feb. 2024.

4 Chao H., Fenhua W., and Ran Z. "Sign language recognition based on CBAM-ResNet," *Association for Computing Machinery*, Oct. 2019.

5 Faisal M., Alsulaiman M., Mekhtiche M., Abdelkader B. M., Algabri M., Alrayes T. B. S., Muhammad G., Mathkour H., Abdul W., Alohali Y., Al-Hammadi M., Altaheri H., and Alfakih T. "Enabling two-way communication of deaf using Saudi sign language," *IEEE Access*, vol. 11, pp. 135423–135434, 2023.

6 Google Developers, "MediaPipe Pose Landmarker," 2024. Available: https://ai.google.dev/edge/mediapipe/solutions/vision/pose_landmarker?hl=en

7 Goyal K. and Velmathi G., "Indian sign language recognition using MediaPipe Holistic," 2023.

8 Hu H., Zhou W., Pu J., and Li H. "Global-local enhancement network for NMF-aware sign language recognition," *ACM Transactions on Multimedia Computing, Communications and Applications*, vol. 17, Aug. 2021.

9 Imashev A. "Sign Language Static Gestures Recognition Tool Prototype," *2017 IEEE 11th International Conference on Application of Information and Communication Technologies (AICT)*, Moscow, Russia, 2017, pp. 1-4. doi: 10.1109/ICAICT.2017.8687032.

10 Kothadiya D. R., Bhatt C. M., Saba T., Rehman A., and Bahaj S. A. "Signformer: Deep vision transformer for sign language recognition," *IEEE Access*, vol. 11, pp. 4730-4739, 2023.

11 Luqman H. "An efficient two-stream network for isolated sign language recognition using accumulative video motion," *IEEE Access*, vol. 10, pp. 93785-93798, 2022.

12 Mirikar M., Singh K., and Dhole S. “Continuous sign language recognition using LSTM and MediaPipe Holistic,” *International Journal of Scientific Research and Engineering Development*, vol. 6. Available: www.ijsred.com
13 Mohandes M., Deriche M., and Liu J. “Image-based and sensor-based approaches to Arabic sign language recognition,” *IEEE Transactions on Human-Machine Systems*, vol. 44, pp. 551-557, 2014.
14 Pei X., Guo D., and Zhao Y. “Continuous sign language recognition based on pseudo-supervised learning,” *Association for Computing Machinery, Inc.*, Oct. 2019, pp. 33-39.
15 Reshna S. and Jayaraju M. “Spotting and recognition of hand gesture for Indian sign language recognition system with skin segmentation and SVM,” *2017 International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET)*, 2017, pp. 386-390.
16 Ronchetti F., Quiroga F., Estrebou C., Lanzarini L., and Rosete A. “LSA64: A dataset of Argentinian sign language,” *XXII Congreso Argentino de Ciencias de la Computación (CACIC)*, 2016.
17 Sosa-Jimenez C. O., Rios-Figueroa H. V., and Solis-Gonzalez-Cosio A. L. “A prototype for Mexican sign language recognition and synthesis in support of a primary care physician,” *IEEE Access*, vol. 10, pp. 127620–127635, 2022.
18 Surdo.kz, “Kazakh sign language dictionary,” 2024. Available: http://www.surdo.kz/kaz/category/1
19 Wikipedia contributors, “F-score,”2024. Accessed February 24, 2024. Available: https://en.wikipedia.org/wiki/F-score
20 World Health Organization. “Deafness and hearing loss,” 2023. Available: https://www.who.int/newsroom/fact-sheets/detail/deafness-and-hearing-loss