Journal of Robotics and Control (JRC)
Volume 6, Issue 1, 2025
ISSN: 2715-5072, DOI: 10.18196/jrc.v6i1.23879

Design of QazSL Sign Language Recognition System for Physically Impaired Individuals

Lazzat Zholshiyeva [1], Tamara Zhukabayeva [2*], Dilaram Baumuratova [3], Azamat Serek [4]
[1, 2] Department of Information Systems, L.N. Gumilyov Eurasian National University, Astana, Kazakhstan
[1, 3] Department of Education, Astana International University, Astana, Kazakhstan
[2] Department of Computer Engineering, Astana IT University, Astana, Kazakhstan
[4] School of Information Technology and Engineering, Kazakh-British Technical University (KBTU), Almaty, Kazakhstan
Email: [1] lazzat.zhol.81@gmail.com, [2] tamara_kokenovna@mail.ru, [3] baumuratova.d@gmail.com, [4] a.serek@kbtu.kz
*Corresponding Author

Abstract*—Automating real-time sign language translation through deep learning and machine learning techniques can greatly enhance communication between the deaf community and the wider public. This research investigates how these technologies can change the way individuals with speech impairments communicate. Despite advancements, developing accurate models for recognizing both static and dynamic gestures remains challenging due to variations in gesture speed and length, which affect the effectiveness of the models. We introduce a hybrid approach that merges machine learning and deep learning methods for sign language recognition. We provide new model for the recognition of Kazakh Sign Language (QazSL), employing five algorithms: Support Vector Machine (SVM), Long Short Term Memory (LSTM), Gated Recurrent Unit (GRU), Convolutional Neural Networks (CNN) with VGG19, ResNet-50, and YOLOv5. The models were trained on a QazSL dataset of more than 4,400 photos. Among the assessed models, the GRU attained the highest accuracy of 100%, followed closely by SVM and YOLOv5 at 99.98%, VGG19 at 98.87% for dynamic dactyls, LSTM at 85%, and ResNet-50 at 78.61%. These findings illustrate the comparative efficacy of each method in real-time gesture recognition. The results yield significant insights for enhancing sign language recognition systems, presenting possible advancements in accessibility and communication for those with hearing impairments.

***Keywords*—Sign Language Recognition; Kazakh Sign Language; Machine Learning; Deep Learning; Physically Impaired Individuals.**

I. INTRODUCTION

Sign languages vary across countries and possess distinct structures, utilizing manual articulation such as hand shapes, movements, and orientations to convey words or concepts [1]. Additionally, non-manual elements like facial expressions and body language play a crucial role in enhancing and emphasizing meaning, aiding in the conveyance of tone and emotion [2], [3]. The visual-spatial modality enables effective communication through the use of body position and movement [4].

Sign languages play a key role in communication within deaf communities, but they still face challenges such as recognition, support and lack of universal standardization as they differ across regions [5]. Also, many deaf children miss language opportunities because their hearing parents are unfamiliar with sign language [6]. Studies indicate that providing early access to sign language for hearing parents can help avoid developmental delays [7]. Understanding sign language is essential for the healthy development of children from deaf families. It facilitates the combination of gestures with verbal skills, improving their communication ability and navigating challenges associated with learning to speak [8]. A significant challenge for individuals with hearing impairments is their isolation from the world. This feeling is intensified by mobility limitations, difficulties communicating with peers and adults, and restricted access to various cultural events and educational opportunities. This problem is particularly significant in Kazakhstan, where more than 30,800 (according to the data from the Astana Deaf Society) individuals experience hearing impairments [9] and over 17,000 children with disabilities have been enrolled in general education schools [10]. Sign language specialists and media interpreters bridge the information gap for the deaf; nevertheless, in Kazakhstan, the government allocates just 60 hours of interpreting services per year, which is insufficient. Thus, automating sign language recognition is essential. However, barriers persist, such as variations in sign language and limited tech accessibility [11], [12]. Furthermore, not everyone is comfortable with modern technology due to insufficient education and development. Inclusive education in Kazakhstan promotes intellectual growth and learning access tailored to individual needs [13]. Therefore, this paper explores solutions to these challenges through research on SLR systems.

The systems designed for sign language recognition play a key role in enhancing communication for individuals with hearing impairments, allowing for improved interaction with their surroundings. Thanks to advancements in DL, and ML, effective solutions have been developed to aid communication for the deaf and hard of hearing, mainly by converting sign language into text or speech through images and videos.

Sign language recognition systems have evolved from static gesture classification to understanding dynamic actions, with visually oriented approaches yielding better results [14], [15]. Developing static and dynamic gesture recognition models remains challenging due to differences in gesture speed and duration [16].

Hybrid technologies can address this issue [17]. Significant improvements have been made in hybrid gesture recog- nition, where combining multiple techniques has enhanced the accuracy and reliability of systems [18], [19].

Recently, there has been growing interest in artificial neural networks as highly effective classifiers [20].

As DL is an evolving field, architectures such as CNNs, RNNs, and video encoders are actively applied [21]–[25]. Advanced hybrid models have the potential to overcome the limitations of traditional approaches and are increasingly popular due to their improved accuracy in recognizing complex gestures [26], [27]. Combining artificial intelligence and machine learning technologies should reconcile the ethical aspects of such conglomerate with user privacy to provide inclusive yet safe communication solutions among the deaf community [28]. Mixed Reality (MR) technologies have been explored for educating of with hearing impairments through Kazakh Sign Language [29]. The effectiveness of MR technologies relies on several factors, including the curriculum, the preparedness of the instructor, and the technical resources.

Given the significance of hybrid methods, this research focuses on studying and comparing various ML, and DL methods for recognizing Kazakh sign language. Recognition of Kazakh signs is still in its early stages and is not widely discussed, although there are some existing studies on this theme [30], [31]. Research confirms that Kazakh sign language can exist as a distinct language based on demonstration forms compared to Russian, English, and Turkish sign languages. ML and DL methods are optimal for classifying Kazakh dactyls. Despite some progress in Kazakh SLR, challenges still exist in dynamic recognition and educational integration. These advancements enhance communication among the deaf community and promote inclusive education in Kazakhstan.

Considering the mentioned shortcomings and issues in the literature, the research questions addressed in this paper are:

1. What are the main challenges in automating the recognition of Kazakh sign language?
2. How can ML and DL algorithms used to improve the accuracy of real-time SLR using computer vision?
3. Which ML and DL algorithms effectively recognise static and dynamic gestures using a custom dataset?

This study aims to develop machine and deep learning models to enhance communication between people with hearing impairments and those who do not know sign language. It also explores methods for converting sign language into text to identify the most effective and accurate model for this translation.

Contributions:

- We developed ML, DL-based models for classifying and recognizing Kazakh dactyls using the custom dataset.
- Applied ML and DL models, such as SVM, LSTM, GRU, CNN, and YOLOv5, for recognizing static and dynamic gestures.
- Compared our models with results from other studies that used these algorithms for recognizing various sign languages with different datasets.
- Evaluated the effectiveness of the applied models in recognizing Kazakh signs with our dataset.
- Created the QazSL platform and a Telegram bot for learning Kazakh sign language, and obtained a copyright certificate.

This paper is structured as follows: Section 1 begins with an introduction, outlining the research questions, objectives, and aim of the paper. Section 2 reviews the literature, and relevant studies, and selects papers for further research. Section 3 describes the research methodology. Section 4 presents the proposed model, while Section 5 discusses the results. Finally, Section 6 provides the conclusion and future work.

II. LITERATURE REVIEW

Conducting a systematic literature review can summarize current scientific research, identify knowledge gaps, and determine future research directions. In this paper, we performed a systematic literature review to address the research questions posed. Initially, we collected results from two databases, Scopus, Google Scholar, IEEE Xplore, and PubMed covering the period from 2020 to 2024, and utilized the following set of keywords (Table I).

From the first set, all 12 articles were selected, while from the second set, 1,082 sources in English were chosen. After filtering by field, 479 articles from the computer science and engineering sectors with open access remained, and 141 articles were selected.

TABLE I. SET OF KEYWORDS

Year	Count	Databases	Keywords
2020-2024	12	Scopus, Google Scholar, IEEE Xplore, PubMed	(TITLE-ABS-KEY (Kazakh) AND TITLE- ABS KEY (“Sign Language”)) AND PUBYEAR > 2019 AND PUBYEAR < 2025
2020-2024	1082	Scopus, Google Scholar, IEEE Xplore, PubMed	(TITLE-ABS-KEY (“Sign Language”) OR TITLE- ABS-KEY (“sign language recognition”) AND TITLE-ABS-KEY (“deep learning”) OR TITLE- ABS-KEY (“machine learning”) OR TITLE-ABS-KEY (computer AND vision)) AND PUBYEAR > 2019 AND PUBYEAR < 2025

A. Inclusion and Exclusion Criteria

From the Scopus, Google Scholar, IEEE Xplore, and PubMed databases, covering the period from 2020 to 2024, 1,094 articles related to sign language recognition using machine learning and computer vision methods were selected for further review. Exclusion criteria included non-English studies, non-relevant and non-open-access papers, and du- plicate publications. Duplicates (n = 14) and non-relevant papers were removed, leaving 141 papers for screening.

All remaining articles were screened, and 95 papers were selected for further research. Subsequently, 42 papers (see Fig. 1) were excluded after full-text review due to irrelevance to the key terms. The complete paper selection process for the study is illustrated in Fig. 2.

Fig. 1. Relevant papers selection

Fig.2 presents the algorithm for the systematic selection and comparison of literature for further research.

Machine translation of sign language is still in its early stages of development but has the potential to transform how information is accessed and communication is facilitated for deaf and hard-of-hearing individuals. However, the technology faces challenges, including accurate gesture recognition and the transmission of grammatical and syntactical nuances of sign language and recognition systems must consider the cultural context for effective communication [32].

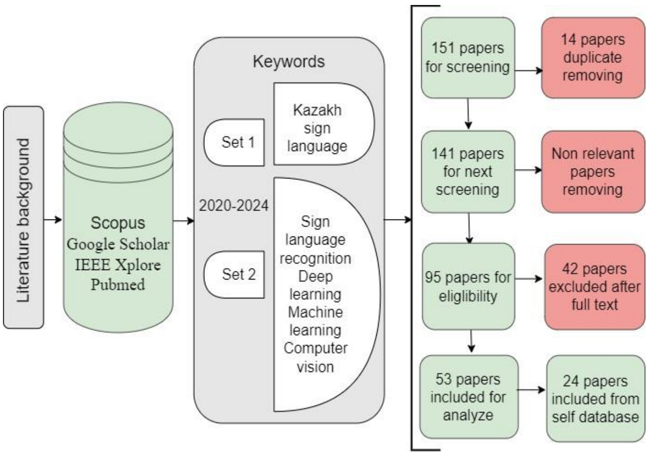

Fig. 2. Complete paper selection process

Recent reviews of methods and algorithms for detecting and recognizing hand gestures in computer vision have led to the development of effective gesture recognition methods [33]. As mentioned, SLR is a complex field utilizing machine learning and computer vision to interpret gestures, including Kazakh sign language. Various models, including Convolutional Neural Networks (CNNs), are employed to enhance the accuracy of SLR. CNNs are effective in processing images and extracting spatial features from video, which is crucial for recognizing complex gestures [34], [35]. Challenges in gesture recognition include the difficulty of recognizing hand poses due to similar gestures and issues with hand size, position, shape, lighting, and background [36], [37]. Transfer learning, where a pre-trained model is adapted to a new task, improves CNN performance in SLR [38]. Recurrent Neural Networks (RNNs), such as LSTM and GRU, capture temporal dependencies in sign language sequences well, making them suitable for recognizing dynamic gestures [39]. A preliminary quantitative analysis of eyebrow position in Kazakh-Russian Sign Language (KRSL) has been presented, exploring the impact of grammar and emotions. The study highlights the importance of OpenPose for sign language analysis [40].

B. Kazakh Sign Language

The historical aspects of sign languages play a significant role in developing recognition systems, presenting both challenges and new opportunities for technological advancements. Historical ideologies and language policies play a significant role in how sign languages are recognized and integrated into technology, along with the advancement of recognition systems [41].

Kazakh Sign Language has been in use for a long time and is employed across all regions. In addition to the primary language, there are also dactyl signs corresponding to the letters of the Kazakh alphabet, which allows for the representation of words without a direct sign equivalent [42], making it a specialized language for the deaf and hard-of-hearing.

Kazakh Sign Language, like many sign languages, developed organically within the deaf community in Kazakhstan, evolving from indigenous sign systems. Over time, it has developed a unique grammar, syntax, and vocabulary, reflecting the needs and culture of the local community. Kazakh Sign Language is closely related to Russian Sign Language [43]. Table II illustrates the history of Kazakh Sign Language's development.

TABLE II. SIGN LANGUAGES ESTABLISH

Sign Languages	Country/Region	Characteristics	Year Established
American Sign Language (ASL)	USA	Influenced by French Sign Language	1817
British Sign Language (BSL)	UK	Two-handed alphabet, regional variations	1760
French Sign Language (LSF)	France	Basis for many other sign languages	1760
Chinese Sign Language (CSL)	China	Regional dialects, evolving standard	1950
Japanese Sign Language (JSL)	Japan	Extensive use of mouth shapes	1908
Russian Sign Language (RSL)	Russia	Incorporates some aspects of oral language	1800s
Kazakh Sign Language (QazSL)	Kazakhstan	Developed from Russian Sign Language	1937s

Sign language is a distinct language with unique grammar, often differing from the country's spoken language, and is typically based on a finger-spelling or dactyl alphabet [44]. Kazakh Sign Language includes gestures specific to Kazakh culture and language, such as the dactyl alphabet used for spelling Kazakh letters. It is employed in specialized schools and institutions for the deaf in Kazakhstan, as well as in public life to promote inclusivity. Although Kazakh Sign Language is an independent language system, it shares connections and similarities with other sign languages. The sign language used in Kazakhstan is closely related to Russian Sign Language (RSL) [43], though they are not identical.

The Kazakh alphabet, which consists of 42 letters-33 borrowed from the Russian alphabet and 9 unique to Kazakh-presents challenges for developing sign language recognition systems due to its unique orthographic features [45]. Additionally, Kazakh, a Turkic language, also uses a Latin variant consisting of 31 letters. While a sign language recognition system for Turkish Sign Language has been developed [46], the lexical differences between Kazakh and Turkish languages necessitate the exploration of recognition for 31 dactyls in KSL using machine learning methods [47]. Non-manual components, such as facial expressions and head orientation, are crucial for distinguishing similar signs and impact recognition accuracy. Improvements in Kazakh Sign Language recognition are needed, presenting opportunities for further research. MediaPipe Holistic and OpenFace were tested for analyzing eyebrow movements in Kazakh-Russian Sign Language, revealing that both models require additional adjustments for accurate linguistic analysis [48].

C. Dataset

The FluentSigners-50 dataset [45] serves as a crucial resource for researching Kazakh-Russian Sign Language (KRSL), offering a variety of data related to age, gender, and sign language usage. This dataset is particularly useful for tasks involving sign language processing because of its extensive linguistic diversity and varied participant backgrounds. Participants' ages span from 8 to 57 years, with a gender breakdown of 18 men and 32 women, showcasing a broad demographic spectrum.

The MediaPipe technology has several potential applications in gesture recognition and dataset creation [49]. MediaPipe models and SVM have been utilized for recognizing movements, demonstrating effectiveness in recognizing Kazakh 42 dactyls [50] and other sign languages [51]–[53]. This highlights how MediaPipe can be used for real-time sign language recognition, translation, and improving accuracy in dynamic and multimodal situations. Creating specialized datasets for different sign languages underscores the importance of targeted data for training and evaluating SLR models.

Previous research on Kazakh Sign Language often employed Russian Sign Language models due to alphabet similarities, allowing Kazakh dactyl language to be considered a distinct sign language with its vocabulary, leading to the creation of datasets for 42 dactyls [54]. Existing gesture recognition methods primarily focus on manual gestures, but facial expressions and body movements also play a crucial role. Consequently, the first dataset for Kazakh-Russian Sign Language (K-RSL) has been developed for use in computer vision and sign language linguistics. This dataset includes 28,250 annotated video clips aimed at enhancing gesture recognition accuracy [55]. Fluent Signers-50 dataset for Kazakh-Russian Sign Language gestures, featuring signers of varying ages, genders, and other characteristics to ensure high linguistic variability. This dataset is utilized for training ML models. Additionally, a corpus created in 2015 for Kazakh-Russian Sign Language is being updated to address gaps in data for machine learning, with ongoing efforts to improve it [56].

III. Proposed Model

This section details the structure of the proposed model "Fig. 4" for creating a more reliable system for converting video gestures into text for recognizing Kazakh dactyls. The model incorporates the efficiency of using machine learning and deep learning algorithms, such as SVM, RNNs (LSTM, GRU), CNNs (VGG16, ResNet-50), and YOLOv5, in recognizing Kazakh Sign Language.

Additionally, the QazSL platform [57] has been developed and is being refined to benefit individuals interested in sign language.

A. Dataset

For the analysis, a dataset of sign language videos is required, which will be compiled from our dataset. The first step involves creating this dataset, which will form the basis for developing the proposed model. The data will be cleaned to remove incomplete and incorrectly formatted records to ensure accuracy, as each video contains both spatial and temporal features.

The dataset was created using a laptop camera and MediaPipe Hands, capturing video streams for recognizing 42 Kazakh dactyls and including 4,450 images. This dataset also features 13 dynamic dactyls of Kazakh Sign Language, such as Ғ, Д, Е, З, Й, К, Қ, Ң, Ұ, Ц, Щ, ь and ъ. To enhance the dataset's applicability, various lighting conditions and distances from the camera were considered. Each image is labelled with the class name from its folder, and the function processes the image to extract 21 key points [58]. Before training the model, the data is classified into 42 classes according to the Kazakh alphabet. The dataset has been processed and divided into training, validation, and test sets with ratios of 70:20:10 for YOLOv5 and 80:20 for SVM, LSTM, GRU, and CNN.

B. Training

The application of SVM, RNN, CNN, and YoloV in SLR takes advantage of their capabilities in managing intricate, dynamic gestures, improving accuracy and facilitating real-time use. SVM is effective with small datasets [59]. LSTM and GRU are effective at capturing temporal dependencies, which makes them well-suited for recognizing movement sequences in dynamic gestures [60]. CNN models such as VGG16 and ResNet-50 are highly effective at extracting spatial features, which play a vital role in recognizing intricate shapes and hand movements in sign language [61], [62]. YoloV5 is designed for real-time object detection, making it perfect for applications that need immediate interpretation of sign language gestures [63].

Because of these benefits, SVM, LSTM, GRU, VGG16, ResNet-50, and YOLOv5 were chosen and trained for the study. The primary YOLOv5 model will be compared with various classification models such as SVM, CNN, and RNN. The choice of these models considered various factors, including the combination of static and dynamic gestures, visual characteristics, real-time needs, and available computational resources.

First, we set up the modules, then import the libraries, create a directory to store the training results and load the data for classification. Next, we preprocess the data, which includes resizing images, normalizing them, augmenting the data to increase diversity, removing noise to improve quality, and splitting the data into training and testing sets. After preprocessing, the model is trained and tested.

C. Evaluation

Evaluation metrics will include time complexity and classification accuracy. In the next stage, hyperparameters of the model will be tuned to enhance its predictive accuracy. To select the most reliable and effective model, its performance will be compared with other algorithms. Metrics such as accuracy, precision, recall, and F1-score will be used to evaluate the model's performance. The effectiveness of the model will be assessed by comparing its results with those of other models.

To enable a more comprehensive analysis and evaluation of classification models, it is beneficial to incorporate additional metrics such as the confusion matrix and ROC curves. The confusion matrix provides a detailed view of the classifier's performance by showing the counts of true positives, false positives, true negatives, and false negatives. This facilitates an assessment of classification errors across various categories. The ROC curve, which illustrates the trade-off between the false positive rate and the true positive rate, is especially useful for evaluating a model's ability to distinguish between classes. The area under the ROC curve (AUC) is a crucial metric for assessing the overall quality of a model, regardless of the classification threshold, and is essential for comparing different models. These additional metrics offer a more nuanced perspective on the model's performance, aiding in the identification of issues like class imbalance or poor class discrimination. The proposed model Fig. 3 is designed to recognize Kazakh dactyls, independent of the performer.

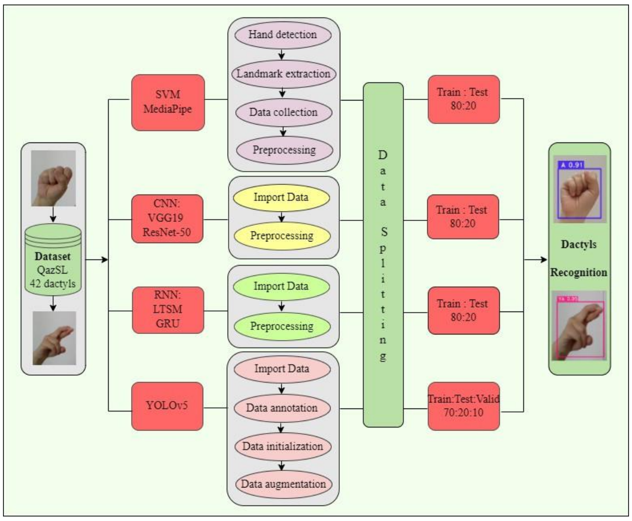

Fig. 3. Proposed model for Kazakh dactyl recognition

IV. Results

This section describes the results of the proposed model. The research results aim to demonstrate the effectiveness of the models in recognizing sign language.

A. MediaPipe and SVM

The SVM method is a versatile set of algorithms used for processing various types of data. It is applied not only for classification and regression tasks but also for anomaly detection. It employs kernel functions to transform data into a higher-dimensional space to improve classification [64]. SVM can be adapted for multi-class tasks, such as gesture classification. This article examines an SVM model using MediaPipe for recognizing Kazakh sign language, including 42 dactyls, focusing on real-time processing speed. MediaPipe utilizes the SVM model to analyze data from the camera, determine the presence of a hand in the image, estimate the 3D position of its joints considering factors such as lighting and camera angle, and output the coordinates of these points. Models for palm detection, highlighting the hand in the image, and hand orientation, which returns 21 3D keypoint coordinates, are included. These functions were applied to extract the hand skeleton from our QazSL dataset. For prediction, we use all 42 classifiers on the input data and select the gesture corresponding to the classifier. The detection achieves a high 99% accuracy [65]. The full algorithm of the model, along with dactyl recognition results and quality assessment metrics, is presented in Fig. 4. As shown in Fig. 4, the results demonstrate excellent performance across all 42 classes with 99% accuracy on all metrics. The F1-score reflects the balance between precision and recall, highlighting the model's high effectiveness in accurate classification. Fig. 5 also presents the confusion matrix for the QazSL dataset. Based on the confusion matrix, it is noted that the proposed SVM model made some errors in recognizing dynamic gestures such as Қ, Ғ, Қ, and З.

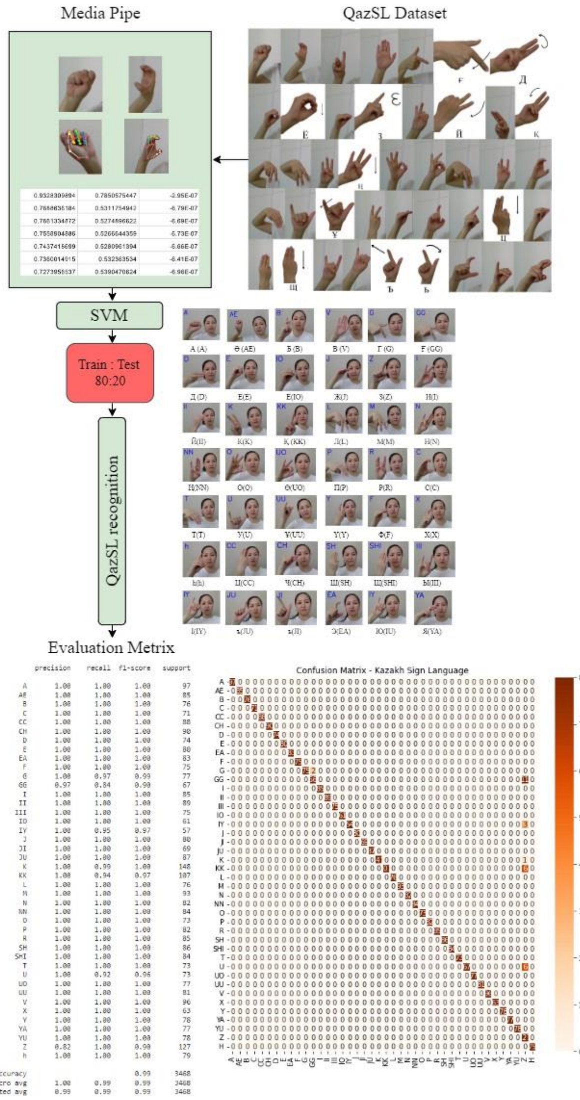

Fig. 4. Algorithm of proposed SVM model

Examining classification errors, particularly in recognizing dynamic gestures like К, Ғ, Қ, and З, can uncover underlying patterns tied to the model's characteristics or the limitations of the data utilized. These errors might stem from various issues, such as challenges in feature extraction, inadequate distinction between similar gestures, and variations in gesture dynamics, including movement speed or amplitude, which can influence classification accuracy.

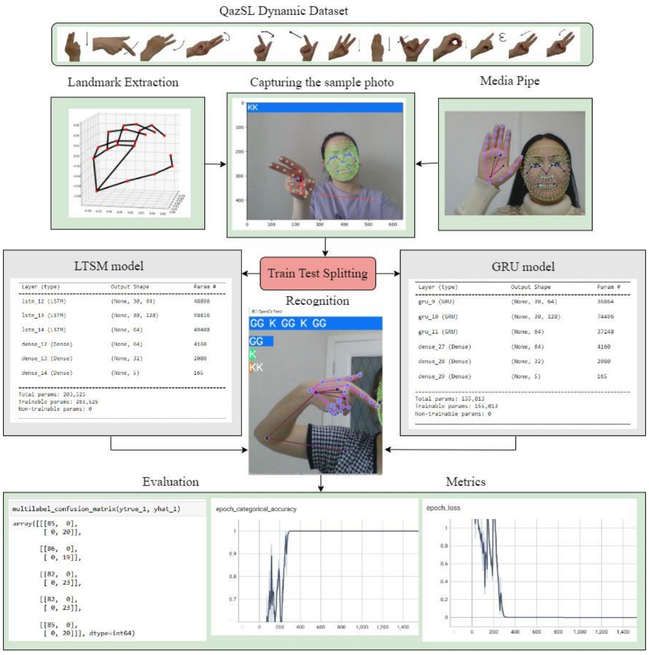

Fig. 5. Algorithm of LTSM, GRU methods

B. LSTM and GRU

LSTM networks were initially used to account for longterm dependencies, such as in handwritten text recognition. An extended application of LSTM is in natural language processing [66]. Subsequently, LSTM was modified, including simplified versions like GRU (Gated Recurrent Unit), which has fewer parameters but retains many key features of LSTM.

Using this feature to address and resolve the issues of the SVM method for recognizing dynamic gestures in QazSL, we explored and presented a dynamic dactyl recognition model based on two architectures: GRU and LSTM. Parameters include node counts ranging from 32 to 128, activation functions such as "relu" or "softmax," and the "Adam" optimizer. GRU is similar to LSTM but has fewer parameters due to the absence of forgetting gates.

The custom dataset used contains 390 videos of dynamic gestures (QazSL), including gestures Ғ, Д, Е, З, Й, К, Қ, Ң, Ұ, Ц, Щ, ь and ъ. The models are neural networks for multi-class classification, where each class corresponds to a single dactyl gesture. The network comprises six layers: three LSTM (or GRU) layers and three fully connected layers. The first, second, and third layers include 64, 128, and 64 neurons, respectively, while the fully connected layers consist of 64 and 32 neurons with the activation function "relu." Accuracy rates of 85% and 100% were achieved for the LSTM and GRU models, respectively. Subsequently, the models were evaluated for recognition accuracy of 13 dactyls of the Kazakh sign language [67].

The LSTM and GRU models achieved impressive accuracy rates of 85% and 100%, respectively, in dynamic gesture recognition. However, their computational complexity must be considered for real-time applications. LSTM is more resource-intensive due to its complex architecture, while GRU is more efficient Fig. 5. In practical scenarios, response time and energy consumption are crucial, especially for mobile or embedded systems. Therefore, further research is needed to optimize these models for effective integration, balancing processing speed and energy efficiency.

C. VGG16 and ResNet-50

The CNN architecture is designed for processing structured data such as images and includes convolutional, pooling, and fully connected layers. CNNs are effective in tasks like image classification, object detection, and segmentation due to their hierarchical feature extraction. VGG16 is an example of such a network, known for its simplicity and high performance in computer vision tasks such as classi- fication and object recognition, owing to its architecture. VGG16 stands out with its use of 3×3 convolutional layers with a unit stride and fixed padding, along with two fully connected layers and a softmax output layer. ResNet50, which consists of 50 layers, uses residual blocks to avoid the vanishing gradient problem and enhance the training of deep networks by adding residual connections, allowing for the creation of a 50-layer model. This model utilizes 2 scaling methods and contains 3.8 billion FLOPs.

In the proposed model, which describes the process of image processing and model building based on CNN architectures such as VGG16 and ResNet50, the process starts with image processing: resizing, filtering, and segmentation. A data generator is then created with image preprocessing, pixel scaling, and data split into training and validation sets. Both models are trained on the training data and evaluated for accuracy on the validation data. Recognition accuracy was achieved at 98.867% (test) and 91.323% (validation) with VGG16, and 78.612% (test) and 62.69% (validation) with ResNet50 respectively [68]. The results and architecture of this model are shown in Fig. 6.

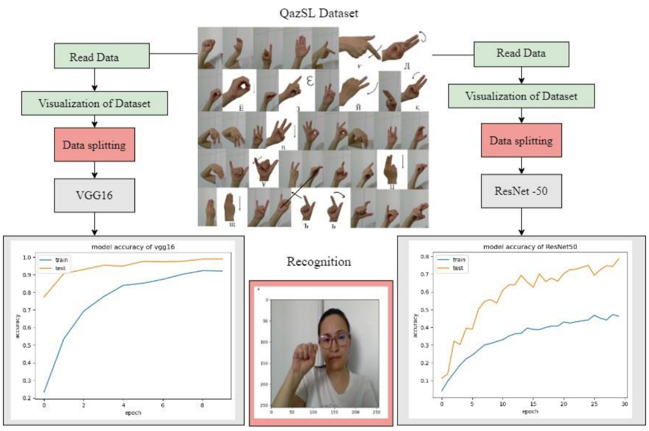

Fig. 6. Architecture of CNN methods

The results indicate that for recognizing Kazakh sign language, the VGG16 architecture outperforms the ResNet50 architecture in terms of accuracy. Although ResNet50 has a more intricate design with residual blocks aimed at preventing vanishing gradients and enhancing the training of deeper networks, its performance might be hindered by the specific structure of the data and the complexity of the features it needs to extract. In contrast, VGG16, with its simpler design and shallower depth, could be more effective for tasks that involve simpler and less complex data. This analysis should take into account how each model deals with variability, noise, and feature complexity, as well as how these elements influence classification accuracy. Additionally, it's crucial to consider the learnability of the models; ResNet50 may excel with larger datasets, while VGG16 might perform better with smaller, less complex datasets.

D. YOLOv5

1) Data Annotation

In the YOLOv5 model, data annotation was initially performed Fig. 7. However, during the review process, it was found that dynamic gestures could not be labelled effectively in photo format. As a result, the classes for dynamic dactyls such as ” ,” ” ,” ” ,” and ” ” remained unannotated. For training the YOLOv5 model, the dataset must include proper annotations. Annotations are defined by drawing bounding boxes around objects in the images The coordinates of these annotations must be normalized to a range of 0 to 1.

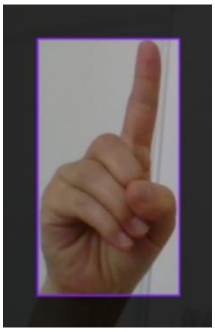

Fig. 7. Data annotation process

Data splitting after that, the dataset is split into 70% for training, 20% for testing, and 10% for validation. Thus, the training set includes 3014 images, the test set includes 861 images, and the validation set includes 430 images.

A 'dataset_params' dictionary is created with information about the classes and the data, as shown in Fig. 8.

```
dataset_params = {
    'data_dir':'/content/sign-lang-3',
    'train_images_dir':'/content/sign-lang-3/train/images',
    'train_labels_dir':'/content/sign-lang-3/train/labels',
    'val_images_dir':'/content/sign-lang-3/valid/images',
    'val_labels_dir':'/content/sign-lang-3/valid/labels',
    'test_images_dir':'/content/sign-lang-3/test/images',
    'test_labels_dir':'/content/sign-lang-3/test/labels',
    'classes': ['A', 'AE', 'B', 'C', 'CH', 'E', 'EA', 'F', 'GG', 'I', 'II', 'III', 'IO', 'IY', 'J', 'JU', 'K', 'KK',
                'L', 'M', 'O', 'P', 'R', 'SH', 'SHI', 'T', 'U', 'UO', 'UU', 'Y', 'YA', 'YU', 'z', 'h', 'v', 'x' ]
}
```

Fig. 8. Classes of QazSL

2) Data Augmentation

Data augmentation is then applied. The 'dataset_params' dictionary, containing information about the classes and the data, is initialized, and the 'train_data' object is created with parameters 'batch_size = 32'. Data augmentation and normalization are performed (see Fig. 9). After defining these parameters, the model is trained.

Fig. 9. Training data with augmentations

Based on the testing results of the model with 50 epochs, max_epoch=50, batch_size=32, and nms_threshold=0.5, the YOLO-trained gesture detection model shows the effectiveness of classifying object classes. The loss metric, which measures the overlap between predicted and actual detection areas, demonstrates the model's accuracy in defining object boundaries. The model achieves a precision of around 88%. Recall, representing the proportion of true positive results among all actual positive objects, indicates that the model detects all real positive objects with 100% accuracy. The value 0.9185264110565186 shows that the F1-score is approximately 92%, reflecting a good balance between precision and recall. The best threshold value for determining a positive prediction, found during testing, is 0.53.

Thus, the results in Table III show that the model demonstrates high performance in gesture detection, achieving excellent precision, recall, and overall average accuracy in the sixth approach. Fig. 10 shows several classification examples obtained using the trained model.

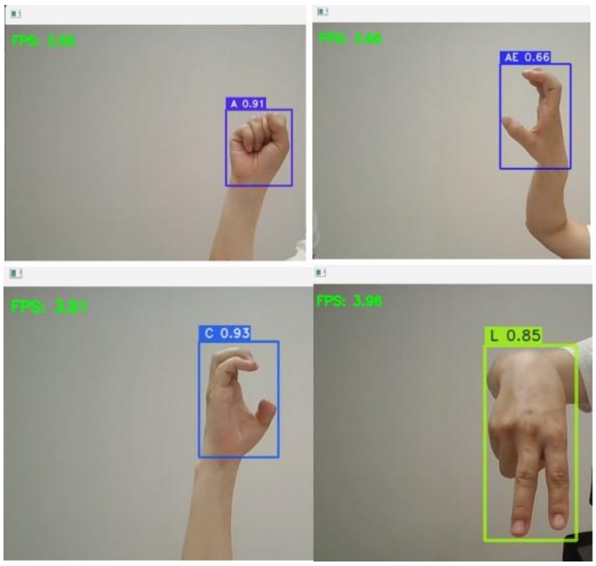

Fig. 10. Recognition of QazSL

TABLE III. TESTING RESULTS

Results	Number of samples	mAP	Precision	Recall	F1-score
max_epoch=20	batch_size=16	1.0	0.35	1.0	0.49
max_epoch=25	batch_size=16	1.0	0.55	1.0	0.68
max_epoch=25	batch_size=32	1.0	0.42	1.0	0.57
max_epoch=50	batch_size=32	1.0	0.27	1.0	0.39
max_epoch=100	batch_size=32	1.0	0.51	1.0	0.64
max_epoch=50	batch_size=16 nms_treshold= 0.5	1.0	0.87	1.0	0.91

3) Model Performance Evaluation

The performance of the proposed system was evaluated in three stages. In the first stage, the model was trained over several epochs. In the second stage, the model was assessed by testing it on data. In the third stage, the performance of the proposed model was compared with other models using accuracy metrics to evaluate its effectiveness. For assessing the quality of the proposed model and analyzing results, metrics such as mean Average Precision (mAP), recall, F1-score, and precision were applied. Table IV presents the values for accuracy, recall, and F1-score for each method, with all average scores being 99%.

The effect of data augmentation on model performance needs more in-depth analysis. Techniques like rotation, scaling, and flipping contribute to enhancing the model's applicability by introducing data diversity and reducing the risk of overfitting. Rotating images can improve the model's ability to handle variations in orientation while flipping helps it become more resilient to shifts in perspective. A closer look at the changes in metrics can shed light on their impact on accuracy and balance.

TABLE IV. SIGN LANGUAGES ESTABLISH

Model	Sign language	Dataset	Accuracy	Precision	Recall	F1-score
SVM [49]	Kazakh Sign Language	31 dactyls	0.98	0.90	0.79	0.84
CNN, LTSM [68]	TSL	AUTSL and Montalbano datasets	0.96	0.96	0.96	0.96
CNN, VGG-16Net [69]	ISL	ISL datasets	0.98	0.96	0.98	0.98
CNN [31]	Kazakh Sign Language	Dynamic gestures	0.98	0.98	0.98	0.98
YOLOv4-CSP [70]	TSL	0-9 figures	mAP 0.99	0.98	0.98	0.98
YOLO, LTSM [71]	ASL	ASLYset	mAP 0.82	0.82	0.82	0.82
Yolov5 [72]	ASL	26 A-Z	mAP 0.95	0.95	0.95	0.95
YOLOv5-m [73]	TSL	288 samples	mAP 0.98	0.91	0.90	0.91
YOLO, LTSM [74]	DSL	DSL10 dataset	mAP 0.90	0.86	0.92	0.89
YOLO [75]	ISL	ISL dataset	mAP 0.99	0.99	0.99	0.99
Proposed SVM	QazSL	42 dactyls	0.99	0.99	0.99	0.99
Proposed LTSM-GRU	QazSL	13 dactyls	0.87-1.00	0.84-1.00	0.84-1.00	0.84-1.00
Proposed VGG16-ResNet50	QazSL	42 dactyls	0.98-0.78	0.98-0.78	0.98-0.78	0.98-0.78
Proposed Yolov5	QazSL	42 dactyls	mAP 1.00	0.88	1.00	0.92

Table IV summarizes that the proposed real-time model for recognizing Kazakh sign language, which includes 42 dactyl letters, is focused on real-time processing speed and achieving a high level of recognition accuracy compared to other studies.

The YOLO model demonstrated the highest accuracy and reliability in recognizing static gestures. To expand the model's capabilities and enhance its effectiveness, increasing the number of signals and iterations could be beneficial. While the LSTM model using skeleton data successfully classifies continuous gestures, its accuracy decreases as the number of gestures increases. Therefore, a hybrid approach combining static and continuous gestures has been chosen [76]. SVM and Extreme Gradient Boosting offer high performance in real-time, while Random Forest shows high accuracy for Kazakh dactyl sign language. CNNs achieve about 94% accuracy for static signs but are less accurate for dynamic ones, which necessitates hybrid methods for video data. LSTM and CNN are popular for gesture recognition: CNNs excel in recognizing and classifying images, while RNNs handle sequential data efficiently. Convolutional neural networks demonstrated accuracy around 94% for static signs, while recognizing dynamic signs poses challenges that may require hybrid methods to address issues with video datasets [74]. Despite difficulties in direct comparison with existing research, our study shows both real-time processing capability and superior results compared to other approaches.

The results indicate a clear trend towards using hybrid methods for automating sign language translation. Limitations of the proposed model:

– Small dataset

– Recognition of only Kazakh sign language

Subsequent studies will concentrate on augmenting the dataset, modifying the model for more sign languages, and enhancing the recognition of dynamic motions.

V. FUTURE WORK

The model demonstrates high accuracy but is limited by a small dataset, which reduces its universality. To improve the model's applicability, it is necessary to expand the dataset with various examples and consider adapting the model to other sign languages.

Future research could focus on developing more comprehensive and diverse datasets to enhance the completeness and accuracy of the results. Moreover, there should be more investigation into hybrid models, as they have the potential to combine the advantages of different methods for recognizing both static and dynamic gestures, thus boosting system efficiency and improving real-time gesture processing.

Several attempts have been made to develop machine translation systems for dynamic sign language. One approach involves using computer vision techniques to track the movements of the signer's hands, and then employing natural language processing (NLP) methods to translate the signs into written or spoken language [78]. Another approach focuses on creating a database of signed phrases and their corresponding translations [79]. Sign language translation can be implemented through isolated sign recognition (ISLR), continuous sign language recognition (CSLR), and sign language translation (SLT) "Fig. 11".

With the advancement of DL, continuous sign language recognition (CSLR) systems have significantly improved the accuracy of gesture recognition without pauses between signs. The recent surge in interest in CSLR can be attributed to the rapid development of machine learning techniques and the increasing number of individuals with hearing impairments who use sign language for communication [79], [80].

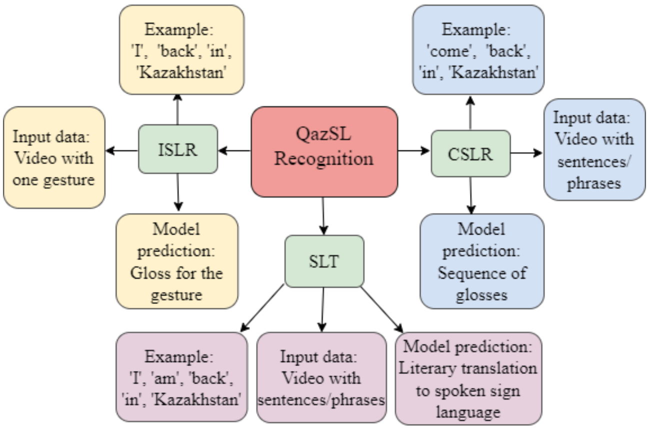

Fig. 11. Types of sign language translation

VI. Conclusion

This paper examined the automation of real-time sign language recognition using ML and DL methods, which can greatly enhance communication between the deaf population and the general public. A hybrid methodology that amalgamates diverse algorithms for gesture recognition, encompassing both machine learning and deep learning techniques, was developed. We presented a hybrid model for Kazakh Sign Language (QazSL) recognition, using SVM, LSTM, GRU, CNN with VGG19 and ResNet-50, and YOLOv5. We also created the QazSL dataset of dactyls and a platform for learning Kazakh Sign Language. SVM, YOLOv5, and GRU achieved the highest accuracy of over 99%. We compared our models with results from other studies and assessed how effective the proposed model is in recognizing Kazakh signs using our dataset.

While the results appear encouraging, the study is limited by a lack of diversity in the dataset, which hinders the model's ability to generalize. Future studies should focus on augmenting the dataset, investigating supplementary methodologies, and refining the analysis of dynamic movements to enhance the suggested system. Additionally, future studies may investigate the adaption of the model for a wider range of sign languages and the creation of resources for sign language acquisition.

This work makes a theoretical contribution by presenting a hybrid strategy that combines the strengths of different methodologies, leading to high accuracy in real-time gesture detection. This discovery could lay the groundwork for future advancements in sign language technology, improving accessibility and communication for those with hearing impairments.

This research significantly improves sign language recognition, creating new possibilities for more inclusive solutions for individuals with hearing impairments.

Acknowledgment

This research has been funded by the Science Committee of the Ministry of Education and Science of the Republic of Kazakhstan (Grant No. AP22686869).

References

[1] M. Moradi, D. D. Kannan, S. Asadianfam, H. Kolivand, and O. Aldhaibani, "A Review of Sign Language Systems," *2023 16th International Conference on Developments in eSystems Engineering (DeSE)*, pp. 200–205, Dec. 2023, doi: 10.1109/dese60595.2023.10468964.

[2] B. A. Dabwan *et al*., "Hand Gesture Classification for Individuals with Disabilities Using the DenseNet121 Model," *2024 International Conference on Advancements in Power, Communication and Intelligent Systems (APCI)*, pp. 1–5, Jun. 2024, doi: 10.1109/apci61480.2024.10616504.

[3] Ș. Takır, B. Bilen, and D. Arslan, "Sentiment Analysis in Turkish Sign Language Through Facial Expressions and Hand Gestures," *2024 32nd Signal Processing and Communications Applications Conference (SIU)*, pp. 1–4, May 2024, doi: 10.1109/siu61531.2024.10601084)

[4] T. N. Fitria, "The use of sign language as a media for delivering information on national television news broadcasts" *ELP (Journal of English Language Pedagogy)*, vol. 9, no. 1, pp. 118–131, Jan. 2024, doi: 10.36665/elp.v9i1.764.)

[5] R. Rastgoo, K. Kiani, S. Escalera, V. Athitsos, and M. Sabokrou, "A survey on recent advances in Sign Language Production," *Expert Systems with Applications*, vol. 243, p. 122846, Jun. 2024, doi: 10.1016/j.eswa.2023.122846.

[6] K. Emmorey, "Ten Things You Should Know About Sign Languages," *Current Directions in Psychological Science*, vol. 32, no. 5, pp. 387–394, May 2023, doi: 10.1177/09637214231173071.

[7] N. K. Caselli, K. Emmorey, and A. M. Cohen-Goldberg, "The signed mental lexicon: Effects of phonological neighbourhood density, iconicity, and childhood language experience," *Journal of Memory and Language*, vol. 121, p. 104282, Dec. 2021, doi: 10.1016/j.jml.2021.104282.

[8] F. Fitriyani, L. Q. Ainii, R. Jannah, and S. Maryam, "Analysis of Sign Language Skills in Improving Communication and Learning for Deaf Children," *Continuous Education: Journal of Science and Research*, vol. 5, no. 1, pp. 30–39, Feb. 2024, doi: 10.51178/ce.v5i1.1757.

[9] K. Pravda. *Uslyshte nas*. Kazakhstanskaya Pravda, 2023.

[10] U.N. Office of the High Commissioner for Human Rights, " Experts from the Committee on the Rights of Persons with Disabilities commend Kazakhstan for its commitment," *OHCHR*, Mar. 2024. [Online]. Available: https://www.ohchr.org/en/news/2024/03/experts-committee-rights-persons-disabilities-commend-kazakhstan-its-commitment.

[11] N. S. Alsharif, T. Clifford, A. Alhebshi, S. N. Rowland, and S. J. Bailey, "Effects of Dietary Nitrate Supplementation on Performance during Single and Repeated Bouts of Short-Duration High-Intensity Exercise: A Systematic Review and Meta-Analysis of Randomised Controlled Trials," *Antioxidants*, vol. 12, no. 6, p. 1194, May 2023, doi: 10.3390/antiox12061194.

[12] R. Rastgoo, K. Kiani, and S. Escalera, "Sign Language Recognition: A Deep Survey," *Expert Systems with Applications*, vol. 164, p. 113794, Feb. 2021, doi: 10.1016/j.eswa.2020.113794.

[13] The Concept of Development of Inclusive Education in Kazakhstan. Available online: https://legalacts.egov.kz/application/downloadconceptfile?id=2506747 (accessed on 2 August 2022)

[14] H. Zhou, D. Wang, Y. Yu, and Z. Zhang, "Research Progress of Human-Computer Interaction Technology Based on Gesture

Recognition," *Electronics*, vol. 12, no. 13, p. 2805, Jun. 2023, doi: 10.3390/electronics12132805.

[15] O. M. Herbert, D. Pérez-Granados, M. A. O. Ruiz, R. Cadena Martínez, C. A. G. Gutiérrez, and M. A. Z. Antuñano, "Static and Dynamic Hand Gestures: A Review of Techniques of Virtual Reality Manipulation," *Sensors*, vol. 24, no. 12, p. 3760, Jun. 2024, doi: 10.3390/s24123760.

[16] Ali, D. Jirak, and S. Wermter, "Snapture—a Novel Neural Architecture for Combined Static and Dynamic Hand Gesture Recognition," *Cognitive Computation*, vol. 15, no. 6, pp. 2014–2033, Jul. 2023, doi: 10.1007/s12559-023-10174-z.

[17] A. M. Buttar, U. Ahmad, A. H. Gumaei, A. Assiri, M. A. Akbar, and B. F. Alkhamees, "Deep Learning in Sign Language Recognition: A Hybrid Approach for the Recognition of Static and Dynamic Signs," *Mathematics*, vol. 11, no. 17, p. 3729, Aug. 2023, doi: 10.3390/math11173729.

[18] H. Mohyuddin, S. K. R. Moosavi, M. H. Zafar, and F. Sanfilippo, "A comprehensive framework for hand gesture recognition using hybrid-metaheuristic algorithms and deep learning models," *Array*, vol. 19, p. 100317, Sep. 2023, doi: 10.1016/j.array.2023.100317.

[19] S. Das, Md. S. Imtiaz, N. H. Neom, N. Siddique, and H. Wang, "A hybrid approach for Bangla sign language recognition using deep transfer learning model with random forest classifier," *Expert Systems with Applications*, vol. 213, p. 118914, Mar. 2023, doi: 10.1016/j.eswa.2022.118914.

[20] L. Zulpukharkyzy Zholshiyeva, T. Kokenovna Zhukabayeva, S. Turaev, M. Aimambetovna Berdiyeva, and D. Tokhtasynovna Jambulova, "Hand Gesture Recognition Methods and Applications: A Literature Survey," *The 7th International Conference on Engineering*, pp. 1–8, Oct. 2021, doi: 10.1145/3492547.3492578.

[21] S. F. Ahmed *et al*., "Deep learning modelling techniques: current progress, applications, advantages, and challenges," *Artificial Intelligence Review*, vol. 56, no. 11, pp. 13521–13617, Apr. 2023, doi: 10.1007/s10462-023-10466-8.

[22] T. Tao, Y. Zhao, T. Liu, and J. Zhu, "Sign Language Recognition: A Comprehensive Review of Traditional and Deep Learning Approaches, Datasets, and Challenges," *IEEE Access*, vol. 12, pp. 75034–75060, 2024, doi: 10.1109/access.2024.3398806.

[23] A. Osman Hashi, S. Zaiton Mohd Hashim, and A. Bte Asamah, "A Systematic Review of Hand Gesture Recognition: An Update From 2018 to 2024," *IEEE Access*, vol. 12, pp. 143599–143626, 2024, doi: 10.1109/access.2024.3421992.

[24] X. Zhao, L. Wang, Y. Zhang, X. Han, M. Deveci, and M. Parmar, "A review of convolutional neural networks in computer vision," *Artificial Intelligence Review*, vol. 57, no. 4, Mar. 2024, doi: 10.1007/s10462-024-10721-6.

[25] Q. M. Areeb, Maryam, M. Nadeem, R. Alroobaea, and F. Anwer, "Helping Hearing-Impaired in Emergency Situations: A Deep Learning-Based Approach," *IEEE Access*, vol. 10, pp. 8502–8517, 2022, doi: 10.1109/access.2022.3142918.

[26] N. R and G. Titus, "Hybrid Deep Learning Models for Hand Gesture Recognition with EMG Signals," *2024 International Conference on Advances in Modern Age Technologies for Health and Engineering Science (Amate)*, pp. 1–6, May 2024, doi: 10.1109/amathe61652.2024.10582166.

[27] N. Ashrafi, Y. Liu, X. Xu, Y. Wang, Z. Zhao, and M. Pishgar, "Deep learning model utilization for mortality prediction in mechanically ventilated ICU patients," *Informatics in Medicine Unlocked*, vol. 49, p. 101562, 2024, doi: 10.1016/j.imu.2024.101562.

[28] A. B. Kydyrbekova and Y. Karymsakova, "The Potential of Using Interactive Storytelling in a Mixed Reality Environment in Teaching Kazakh Sign Language," *Iasaýı ýnıversıtet n ń habarshysy*, vol. 132, no. 2, pp. 373–385, Jun. 2024, doi: 10.47526/2024-2/2664-0686.69.

[29] H. ZainEldin *et al*., "Silent no more: a comprehensive review of artificial intelligence, deep learning, and machine learning in facilitating deaf and mute communication," *Artificial Intelligence Review*, vol. 57, no. 7, Jun. 2024, doi: 10.1007/s10462-024-10816-0.

[30] N. Amangeldy, S. Kudubayeva, A. Kassymova, A. Karipzhanova, B. Razakhova, and S. Kuralov, "Sign Language Recognition Method Based on Palm Definition Model and Multiple Classification," *Sensors*, vol. 22, no. 17, p. 6621, Sep. 2022, doi: 10.3390/s22176621.

[31] A. A. Rahim, "Kazakh Sign Language Recognition By Using Machine Learning Methods," *Вестник Аүэс*, vol. 2, no. 65, 2024.

[32] Y. A. Gomaa, "Deciphering linguistic and cultural hurdles in English-Arabic media translation: Insights from the BBC online news articles," *Cadernos de Tradução*, vol. 44, no. 1, pp. 1–21, Mar. 2024, doi: 10.5007/2175-7968.2024.e94510.

[33] B. Ren, M. Liu, R. Ding, and H. Liu, "A Survey on 3D Skeleton-Based Action Recognition Using Learning Method," *Cyborg and Bionic Systems*, vol. 5, Jan. 2024, doi: 10.34133/cbsystems.0100.

[34] G. Halvardsson, J. Peterson, C. Soto-Valero, and B. Baudry, "Interpretation of Swedish Sign Language Using Convolutional Neural Networks and Transfer Learning," *SN Computer Science*, vol. 2, no. 3, Apr. 2021, doi: 10.1007/s42979-021-00612-w.

[35] R. Amer Kadhim and M. Khamees, "A Real-Time American Sign Language Recognition System using Convolutional Neural Network for Real Datasets," *TEM Journal*, pp. 937–943, Aug. 2020, doi: 10.18421/tem93-14.

[36] A. Kasapbaşi, A. E. A. Elbushra, O. lL-Hardanee, and A. Yilmaz, "DeepASLR: A CNN based human-computer interface for American Sign Language recognition for hearing-impaired individuals," *Computer Methods and Programs in Biomedicine Update*, vol. 2, p. 100048, 2022, doi: 10.1016/j.cmpbup.2021.100048.

[37] B. Kareem Murad and A. H. Hassin Alasadi, "Advancements and Challenges in Hand Gesture Recognition: A Comprehensive Review," *Iraqi Journal for Electrical and Electronic Engineering*, vol. 20, no. 2, pp. 154–164, Jul. 2024, doi: 10.37917/ijeee.20.2.13.

[38] M. Iman, H. R. Arabnia, and K. Rasheed, "A Review of Deep Transfer Learning and Recent Advancements," *Technologies*, vol. 11, no. 2, p. 40, Mar. 2023, doi: 10.3390/technologies11020040.

[39] H. Brock, F. Law, K. Nakadai, and Y. Nagashima, "Learning Three-dimensional Skeleton Data from Sign Language Video," *ACM Transactions on Intelligent Systems and Technology*, vol. 11, no. 3, pp. 1–24, Apr. 2020, doi: 10.1145/3377552.

[40] V. Kimmelman, A. Imashev, M. Mukushev, and A. Sandygulova, "Eyebrow position in grammatical and emotional expressions in Kazakh-Russian Sign Language: A quantitative study," *Plos One*, vol. 15, no. 6, p. e0233731, Jun. 2020, doi: 10.1371/journal.pone.0233731.

[41] J. Lo Bianco, "Ideologies of sign language and their repercussions in language policy determinations," *Language Communication*, vol. 75, pp. 83–93, Nov. 2020, doi: 10.1016/j.langcom.2020.09.002.

[42] D. Saparova and A. Kanagatova, "The prefigurative culture of Margaret Mead or the eternal problem of generation gap," *Eurasian Journal of Religious Studies*, vol. 16, no. 4, pp. 34–41, 2018, doi: 10.26577/ejrs-2018-4-186.

[43] A. Imashev, M. Mukushev, V. Kimmelman, and A. Sandygulova, "A Dataset for Linguistic Understanding, Visual Evaluation, and Recognition of Sign Languages: The K-RSL," *Proceedings of the 24th Conference on Computational Natural Language Learning*, pp. 631–640, 2020, doi: 10.18653/v1/2020.conll-1.51.

[44] Y. Obi, K. S. Claudio, V. M. Budiman, S. Achmad, and A. Kurniawan, "Sign language recognition system for communicating to people with disabilities," *Procedia Computer Science*, vol. 216, pp. 13–20, 2023, doi: 10.1016/j.procs.2022.12.106.

[45] 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, p. e0273649, Sep. 2022, doi: 10.1371/journal.pone.0273649.

[46] A. Kasapbaşi and H. Canbolat, "Prediction of Turkish Sign Language Alphabets Utilizing Deep Learning Method," *2023 5th International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA)*, pp. 1–7, Jun. 2023, doi: 10.1109/hora58378.2023.10156705.

[47] C. Kenshimov, Z. Buribayev, Y. Amirgaliyev, A. Ataniyazova, and A. Aitimov, "Sign language dactyl recognition based on machine learning algorithms," *Eastern-European Journal of Enterprise*

Technologies, vol. 4, no. 2(112), pp. 58–72, Aug. 2021, doi: 10.15587/1729-4061.2021.239253.

[48] A. Kuznetsova, V. Kimmelman, "Testing MediaPipe Holistic for Linguistic Analysis of Nonmanual Markers in Sign Languages," *Physics Today*, vol. 2014, no. 12, Dec. 2014, doi: 10.1063/pt.5.028530.

[49] G. Sánchez-Brizuela, A. Cisnal, E. de la Fuente-López, J.-C. Fraile, and J. Pérez-Turiel, "Lightweight real-time hand segmentation leveraging MediaPipe landmark detection," *Virtual Reality*, vol. 27, no. 4, pp. 3125–3132, Sep. 2023, doi: 10.1007/s10055-023-00858-0.

[50] L. Zholshiyeva, T. Zhukabayeva, Sh. Turaev, M. Berdieva, and R. Sengirbayeva, "Real-time Kazakh sign language recognition using Mediapipe and SVM," *News of the academy of sciences of the Re- public of Kazakhstan, Series physics and information technology*, vol. 1, no. 345, pp. 82–93, Mar. 2023, doi: 10.32014/2023.2518-1726.17

[51] S. Katoch, V. Singh, and U. S. Tiwary, "Indian Sign Language recognition system using SURF with SVM and CNN," *Array*, vol. 14, p. 100141, Jul. 2022, doi: 10.1016/j.array.2022.100141.

[52] S. Renjith and R. Manazhy, "Sign language: a systematic review on classification and recognition," *Multimedia Tools and Applications*, vol. 83, no. 31, pp. 77077–77127, Feb. 2024, doi: 10.1007/s11042-024-18583-4.

[53] R. Saravanan and S. Veluchamy, "Sign Language Classification With MediaPipe Hand Landmarks," *2023 International Conference on Energy, Materials and Communication Engineering (ICEMCE)*, pp. 1–6, Dec. 2023, doi: 10.1109/icemce57940.2023.10434034.

[54] Y. Amirgaliyev, A. Ataniyazova, Z. Buribayev, M. Zhassuzak, B. Urmashev, and L. Cherikbayeva, "Application of neural networks ensemble method for the Kazakh sign language recognition," *Bulletin of Electrical Engineering and Informatics*, vol. 13, no. 5, pp. 3275–3287, Oct. 2024, doi: 10.11591/eei.v13i5.7803.

[55] S. Mukhanov, R. Uskenbayeva, Young Im Cho, D. Kabyl, N. Les, and M. Amangeldi, "Gesture recognition of machine learning and convolutional neural network methods for Kazakh sign language," *Scientific Journal of Astana IT University*, pp. 85–100, Sep. 2023, doi: 10.37943/15lpcu4095.

[56] A. Imashev, M. Mukushev, V. Kimmelman, and A. Sandygulova, "A Dataset for Linguistic Understanding, Visual Evaluation, and Recognition of Sign Languages: The K-RSL," *Proceedings of the 24th Conference on Computational Natural Language Learning*, pp. 631–640, 2020, doi: 10.18653/v1/2020.conll-1.51.

[57] L. Zholshiyeva, T. Zhukabayeva, S. Turaev, M. Berdiyeva, and S. R. Kuanysbaevna, "A Real-Time Approach to Recognition of Kazakh Sign Language," *2022 International Conference on Smart Information Systems and Technologies (SIST)*, pp. 1–6, Apr. 2022, doi: 10.1109/sist54437.2022.9945799.

[58] H. Goddard and L. Shamir, "SVMnet: Non-Parametric Image Classification Based on Convolutional Ensembles of Support Vector Machines for Small Training Sets," *IEEE Access*, vol. 10, pp. 24029–24038, 2022, doi: 10.1109/access.2022.3154405.

[59] T. H. Noor *et al*., "Real-Time Arabic Sign Language Recognition Using a Hybrid Deep Learning Model," *Sensors*, vol. 24, no. 11, p. 3683, Jun. 2024, doi: 10.3390/s24113683.

[60] R. K. Pathan, M. Biswas, S. Yasmin, M. U. Khandaker, M. Salman, and A. A. F. Youssef, "Sign language recognition using the fusion of image and hand landmarks through multi-headed convolutional neural network," *Scientific Reports*, vol. 13, no. 1, Oct. 2023, doi: 10.1038/s41598-023-43852-x.

[61] Y. Ma, T. Xu, and K. Kim, "Two-Stream Mixed Convolutional Neural Network for American Sign Language Recognition," *Sensors*, vol. 22, no. 16, p. 5959, Aug. 2022, doi: 10.3390/s22165959.

[62] N. F. Attia, M. T. F. S. Ahmed, and M. A. M. Alshewimy, "Efficient deep learning models based on tension techniques for sign language recognition," *Intelligent Systems with Applications*, vol. 20, p. 200284, Nov. 2023, doi: 10.1016/j.iswa.2023.200284.

[63] H. Sun, L. Wang, H. Liu, and Y. Sun, "Hyperspectral Image Classification with the Orthogonal Self-Attention ResNet and Two-Step Support Vector Machine," *Remote Sensing*, vol. 16, no. 6, p. 1010, Mar. 2024, doi: 10.3390/rs16061010.

[64] G. Makridis, P. Mavrepis, and D. Kyriazis, "A deep learning approach using natural language processing and time-series forecasting towards enhanced food safety," *Machine Learning*, vol. 112, no. 4, pp. 1287–1313, Mar. 2022, doi: 10.1007/s10994-022-06151-6.

[65] L. Zholshiyeva, T. Zhukabayeva, Sh. Turaev, M. Berdieva, "Kazakh sign language recognition based on CNN," *News of the Academy of Sciences of the Republic of Kazakhstan, Series physics and information technology*, vol. 3, no. 347, pp. 76–87, 2023.

[66] L. Zholshiyeva1, T. Zhukabayeva, Sh. Turaev, M. Berdieva, and B. Khu Ven-Tsen, "Development of an intellectual system for recognizing Kazakh dactyl gestures based on LSTM and GRU models," *News of the Academy of Sciences of the Republic of Kazakhstan, Series physics and information technology*, vol. 2, no. 346, pp. 1'41-153, 2023, doi: 10.32014/2023.2518-1726.190.

[67] O. M. Sincan and H. Y. Keles, "AUTSL: A Large Scale Multi-Modal Turkish Sign Language Dataset and Baseline Methods," *IEEE Access*, vol. 8, pp. 181340–181355, 2020, doi: 10.1109/access.2020.3028072.

[68] B. Saini, D. Venkatesh, N. Chaudhari, T. Shelake, S. Gite, and B. Pradhan, "A comparative analysis of Indian sign language recognition using deep learning models," *Forum for Linguistic Studies*, vol. 5, no. 1, p. 197, Jul. 2023, doi: 10.18063/fls.v5i1.1617.

[69] M. Alaftekin, I. Pacal, and K. Cicek, "Real-time sign language recognition based on YOLO algorithm," *Neural Computing and Applications*, vol. 36, no. 14, pp. 7609–7624, Feb. 2024, doi: 10.1007/s00521-024-09503-6.

[70] M. Rivera-Acosta, J. M. Ruiz-Varela, S. Ortega-Cisneros, J. Rivera, R. Parra-Michel, and P. Mejia-Alvarez, "Spelling Correction Real-Time American Sign Language Alphabet Translation System Based on YOLO Network and LSTM," *Electronics*, vol. 10, no. 9, p. 1035, Apr. 2021, doi: 10.3390/electronics10091035.

[71] I. G. A. Poornima, G. S. Priya, C. A. Yogaraja, R. Venkatesh, and P. Shalini, "Hand and Sign Recognition of Alphabets Using YOLOv5," *SN Computer Science*, vol. 5, no. 3, Mar. 2024, doi: 10.1007/s42979-024-02628-4.

[72] B. Hemachandran, C. Pavan Rakesh Reddy, and D. Harsha Vardhan Reddy, "Comparative Study of Classification Algorithms in Sign Language Recognition," *2022 13th International Conference on Computing Communication and Networking Technologies (ICCCNT)*, vol. 1045, pp. 1–5, Oct. 2022, doi: 10.1109/icccnt54827.2022.9984428.

[73] S. Alyami, H. Luqman, and M. Hammoudeh, "Reviewing

[74] 25 years of continuous sign language recognition research: Advances, challenges, and prospects," *Information Processing, Management*, vol. 61, no. 5, p. 103774, Sep. 2024, doi: 10.1016/j.ipm.2024.103774.

[75] Q. M. Areeb, Maryam, M. Nadeem, R. Alroobaea, and F. Anwer, "Helping Hearing-Impaired in Emergency Situations: A Deep Learning-Based Approach," *IEEE Access*, vol. 10, pp. 8502–8517, 2022, doi: 10.1109/access.2022.3142918.

[76] D. Nurgazina, S. Kudubayeva, and A. Ismailov, "Scientific Aspects Of Modern Approaches To Machine Translation For Sign Language," *Scientific Journal of Astana IT University*, pp. 41–54, Jun. 2024, doi: 10.37943/18dqxx2356.

[77] A. Núñez-Marcos, O. Perez-de-Viñaspre, and G. Labaka, "A survey on Sign Language machine translation," *Expert Systems with Applications*, vol. 213, p. 118993, Mar. 2023.

[78] R. Zuo and B. Mak, "Improving Continuous Sign Language Recognition with Consistency Constraints and Signer Removal," *ACM Transactions on Multimedia Computing, Communications, and Applications*, vol. 20, no. 6, pp. 1–25, Mar. 2024.

[79] F. Wei and Y. Chen, "Improving Continuous Sign Language Recognition with Cross-Lingual Signs," *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 23555– 23564, Oct. 2023, doi: 10.1109/iccv51070.2023.02158.

[80] W. Xue, Z. Kang, L. Guo, S. Yang, T. Yuan, and S. Chen, "Continuous Sign Language Recognition for Hearing-Impaired Consumer Communication via Self-Guidance Network," *IEEE Transactions on Consumer Electronics*, vol. 70, no. 1, pp. 535–542, Feb. 2024, doi: 10.1109/tce.2023.3342163.