# COMPUTER SCIENCE
# КОМПЬЮТЕРЛІК ҒЫЛЫМДАР
# КОМПЬЮТЕРНЫЕ НАУКИ

UDC 681.327.12
IRSTI 65.011.56

https://doi.org/10.55452/1998-6688-2025-22-1-10-24

**[1]Othman M.,**
Professor, ORCID ID: 0000-0002-5124-5759,
e-mail: mothmanupm@gmail.com
**[2]Oralbekova D.,**
Senior Lecturer,
e-mail: dinaoral@mail.ru
**[2]*Berzhanova U.G.,**
Doctoral student, ORCID ID: 0009-0000-2467-5721,
e-mail: berzhanovaulmekenn@gmail.com

[1]Universiti Putra Malaysia, Kuala Lumpur, Malaysia
[2]Satbayev University, Almaty, Kazakhstan

## DEVELOPMENT OF A KAZAKH SIGN LANGUAGE RECOGNITION MODEL BASED ON YOLO-NAS

**Abstract**

The development of a reliable model of recognition of Kazakh sign language is an important step towards the development of inclusive communication and assistance to people with hearing impairments. This paper describes in detail the process of collecting and annotating data in which gesture images were used. Special attention is paid to the preparation and preprocessing of data to ensure their compatibility with the model. The process of learning the model involves optimizing hyperparameters and using various techniques to improve recognition accuracy. We also conducted a comprehensive performance assessment of the model based on test data to ensure its effectiveness in real-world conditions. In addition to the main development phase, we are considering testing the YOLO-NAS model on the same dataset to explore potential improvements in accuracy and performance. In conclusion, the results of our research can be used to further develop technologies that facilitate the integration of people with hearing impairments into society, as well as to create educational and communication platforms based on the Kazakh sign language.

**Key words:** You Only Look Once (YOLO); YOLO-NAS, Deep Learning; Convolutional Neural Network (CNN); Artificial Intelligence, Kazakh sign language.

## Introduction

Sign language is an important means of communication for people with hearing and speech impairments, and its development in Kazakhstan is becoming increasingly relevant. Kazakh sign language, like other national sign languages, requires the development of specialized technologies for its full recognition and interpretation. In recent years, Kazakhstan has been actively working on the study and standardization of the Kazakh sign language, which makes our research especially timely and relevant.

The purpose of our research is to develop a modern model based on the YOLO algorithm, capable of effectively recognizing the gestures of Kazakh sign language. We strive to provide high accuracy and speed of recognition, which will allow us to use this model in real-world applications that contribute to improving communication and social integration of people with hearing impairments.

The novelty of this work lies in the use of the latest achievements of YOLO architecture, such as YOLO-NAS small, to create a recognition system for Kazakh sign language. This is the first time we have applied these algorithms to solve this problem, which makes our work unique in the framework of research and development of technologies for Kazakh sign language. Our approach includes a detailed analysis of the development of the YOLO algorithm (Figure 1), its adaptation to the specific requirements of the Kazakh sign language and a comprehensive learning process of the model based on specialized data.

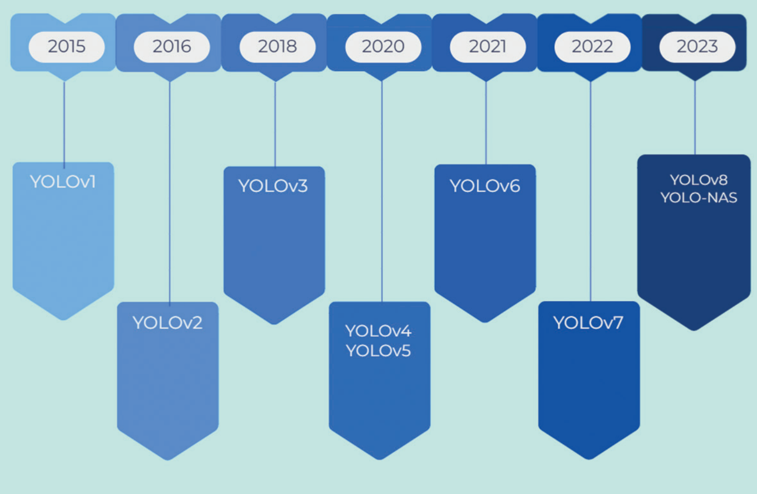


Figure 1 – The chronology of the development
of YOLO models from 2015 to 2023

Thus, the use of the YOLO algorithm in our study opens up new opportunities for the development of advanced gesture recognition technologies, which will have a significant impact on social integration and improving the quality of life of people with disabilities.

## Literature review

Many researchers choose YOLO to create gesture recognition systems due to its high performance and speed. YOLO allows you to detect and classify objects in real time, which is especially important for tasks related to gesture recognition, where instant reaction is required. Its ability to efficiently process images with high precision makes YOLO an ideal choice for applications such as sign language interpretation. Moreover, the ease of use and availability of pre-trained models on various datasets contributes to the widespread adoption of this technology.

In this study [1], a gesture recognition system using the YOLO method has been developed that automates the translation of Indonesian Sign language (Bisindo) into text. A pre-trained YOLOv3 model adapted to gesture recognition tasks was used for the system. A proprietary dataset was collected, including images and video recordings of gestures.

Experiments have shown that when using images, the system achieved 100% accuracy, but when working with video data, the accuracy decreased to 72.97%. The main problem was the difficulty in recognizing transitional frames between gestures, which led to errors. In conclusion, the authors recommend the development of algorithms capable of distinguishing transient frames and gestures, as well as adjusting the frame rate and recognition speed to improve the accuracy of the system in real time.

The article [2] proposes a new technique using Deep Learning for accurate recognition of Arabic sign language (ArSL). The purpose of the study is to promote communication between hearing and deaf–mute people. The system is based on advanced attention mechanisms and the modern architecture of convolutional neural networks (CNN) combined with the YOLO model for object detection. The integration of self-observation, channel and spatial attention units, as well as the cross-convolution module, allows for 98.9% recognition accuracy. The technique demonstrates a significant improvement over traditional methods, which is confirmed by high accuracy and mAP indicators. This model provides an effective solution for ArSL recognition and promotes the social integration of deaf people in the Arab region and beyond.

The following article [3] presents a method of automatic sign language translation that uses the latest advances in computer vision and Deep Learning to improve communication between deaf and hard of hearing people. The method is based on YOLO models used to recognize movements and gesture classes. Experimental results show that the proposed model, especially YOLOv5, achieves high accuracy up to 99.50% in gesture recognition from various datasets. To improve performance, the system uses an algorithm for clustering datasets with minimal manual annotation. The authors also suggest the possibility of further research, including using a combination of different YOLO models for different stages of gesture recognition, which can improve accuracy and reduce computational costs. This method can help overcome communication barriers for deaf and hard of hearing people through messaging or video calls.

The article [4] presents a new method of online hand gesture recognition based on the YOLO object detection model and transformer architecture. The method focuses on the classification and temporal localization of gestures, using the skeleton of the hand as input. The proposed model directly predicts classes and boundaries of gestures, learning based on a new loss function that takes into account gesture centers and boundaries in the time domain. The method was tested on two datasets: SHREC’22 and IPN Hand, where it showed high accuracy and efficiency, significantly improving the results of existing approaches. The model allows you to make early predictions, which is especially important for practical applications.

In this paper [5], a real-time hand gesture recognition system (RTHGR) is presented, which allows interacting with the system using gestures. The webcam captures the user’s gestures, recognizes them and performs the appropriate actions. A neural network based on the YOLO convolution model is used to classify gestures. The preprocessing of gesture data includes skin color recognition, marker-based watershed detection algorithms, and seed filling algorithms, which allows you to obtain clear gesture characteristics and reduce the influence of background. The system achieved 98.66% accuracy in normal lighting for 10 different types of gestures. This method does not require additional training or time for detection.

The article [6] presents a study on the development of a gesture recognition system for deaf people using Indian Sign language (ISL). The development process includes two key steps: creating a dataset and training a model for gesture detection. At the first stage, images are generated for each ISL character, and at the second stage, the YOLOv3 algorithm is used for gesture recognition. The model uses a convolutional neural network based on darknet-53 to extract features. The tests performed on 50 images showed an average accuracy of 82%.

In addition, the article describes a methodology for creating a road network based on GPS trajectories of vehicles. The applied algorithm selects representative points from dense, noisy and stationary data, and then generates a road network. The experimental results show that the method is effective in creating a road network, reducing the number of points and storage requirements, although it faces difficulties when working with sharp curves.

The article [7] proposes a method for automatic recognition of the Malaysian sign language based on the use of a convolutional neural network with the YOLOv3 algorithm. During the research, a dataset was created and annotated, including images and video recordings of gestures. The model was trained using the Darknet platform and demonstrated 63% accuracy after 7000 iterations. During testing, the system achieved 72% accuracy. The results showed that the accuracy of gesture recognition depends on various factors such as background and hand angle. In the future, it is planned to expand the dataset and explore alternative methods, including recurrent and generative adversarial networks, to improve accuracy and improve system performance.

In [8], two methods for recognizing six characteristic hand gestures in various environmental conditions are proposed. The first method is based on highlighting the features of the brush using bulge defects. To do this a skin color transformation is applied to determine the area of the brush, after which contours and bulges defects are highlighted to identify gestures. The second method uses the YOLOv3 model, based on Deep Learning, with the DARKNET-53 (CNN) architecture. The model is trained on a large annotated dataset.

Experimental results show that the Deep Learning method is superior to the hand-based approach with an accuracy of 98.92% versus 95.57%. Both approaches are effective for real-time gesture recognition and static images, while YOLOv3 demonstrates the highest accuracy and ability to work in various lighting and background conditions.

The authors of the article [9] presented a simplified model for hand gesture recognition based on convolutional neural networks YOLOv3 and DarkNet-53. This model has been specially designed to achieve high precision gesture recognition without the need for additional image preprocessing or enhancement. In difficult conditions, including low-resolution images, the model showed excellent results with 97.68% accuracy, 94.88% precision, 98.66% memorization and 96.70% F-1 rating.

Labeled datasets in Pascal VOC and YOLO formats were used to evaluate the model. In comparison with other methods such as Single Shot Detector (SSD) and Visual Geometry Group (VGG16), which show accuracy in the range from 82% to 85%, the proposed model demonstrated significantly better results. It works effectively with both static images of hands and dynamic gestures from video frames, which makes it a very promising tool for practical use.

The authors of the article [10] proposed a model for recognizing sign language based on the YOLOv3 algorithm, designed to facilitate communication with people with hearing and speech impairments. Unlike existing systems that either recognize individual letters or are limited to a small number of words, the developed model is able to recognize words directly.

As part of the work, a dataset was created using labelImg software, and gesture identification is carried out both through image uploading and live streaming using OpenCV. The YOLOv3 model is trained on hand images and uses a graphics processor to improve performance.

The system includes a graphical user interface (GUI) created using the PAGE tool and is designed to recognize 10 frequently used words. The results showed that the proposed model provides more reliable and effective gesture recognition compared to existing methods, which greatly facilitates interaction with deaf people in everyday life.

The authors of the article [11] proposed an optimized YOLOv4-CSP model for recognizing hand gestures in Turkish sign language in real time. The model is based on convolutional neural networks and includes the addition of a CSPNet network to YOLOv4 to improve performance. As part of the work, the Mish activation function, the CIoU loss function and the transformer unit are integrated, which contributes to faster and more accurate gesture processing.

To evaluate the effectiveness of the proposed method, a comparison was made with previous versions of YOLO, including YOLOv3 and YOLOv4, as well as with other gesture recognition methods. The results showed high levels of accuracy, precision, memorization and F1 evaluation: 98.95%, 98.15%, 98.55% and 99.49%, respectively, with a processing time of 9.8 ms. The model demonstrates excellent results compared to alternative methods, providing reliable gesture recognition in various conditions.

The authors [12] developed a YOLO-based gesture recognition system for translating American Sign language (ASL) into letters, numbers and words, facilitating communication between speakers and deaf-mutes. The model trained on a set of 8000 images achieved 98.01% accuracy for images and 28.9 frames per second for video. The average losses were 1.3, the return and F1 indicators were 0.96. The YOLOv4-based system showed 98% accuracy for images and similar results for videos.

These authors [13] focused on developing a real-time hand gesture recognition system for Turkish sign language using the optimized YOLOv4-CSP model. The work uses the modern YOLOv4-CSP algorithm, created by adding the CSPNet network to the basic YOLOv4 to improve performance. The model includes a Mish activation function, a CIoU loss function and a transformer unit, which contributes to faster learning and improved accuracy.

To evaluate the performance and speed of detection, the YOLOv4-CSP model was compared with previous versions of YOLO (YOLOv3 and YOLOv3-SPP). Using a labeled dataset with numbers in Turkish sign language, the new model achieved outstanding results: 98.95% accuracy, 98.15% memorization, 98.55% F1 score and 99.49% display results, with a processing time of 9.8 ms.

The proposed system effectively recognizes static hand gestures and classifies them without the influence of background difficulties, which eliminates the need for preprocessing images.

In [14], a sign language recognition system based on two models was proposed: YOLOv4 and the support vector machine (SVM) using MediaPipe. Both models are designed to classify 80 gestures from a sign language based on a dataset consisting of 676 images of static signs. YOLOv4 and SVM with MediaPipe were evaluated for their accuracy, which was 98.8% and 98.62%, respectively. YOLOv4 has demonstrated higher accuracy, but requires significant computing resources, while SVM with MediaPipe provides fast processing, but may be subject to interference due to dependence on key hand points allocated by MediaPipe.

To improve accuracy, an expert system is proposed that combines both models, which allows you to achieve even better results in real time. The system was compared with modern methods, and the results showed that the proposed approach provides higher accuracy compared to existing models.

In addition, a comprehensive system for recognizing Bengali signs and generating text was developed in [15]. The system includes two key stages. At the first stage, the quantization method is used for the YOLOv4-Tiny model, which detects 49 different characters, including characters of the Bengali alphabet, numbers and special characters. YOLOv4-Tiny localizes the signs and predicts the corresponding symbols. In the second stage, the long-term short-term memory (LSTM) model is used to generate meaningful text from the detected characters.

The system was trained on the BSL 49 dataset containing about 14,745 images of 49 different classes, which ensures high accuracy. The proposed YOLOv4-Tiny quantized model demonstrates an accuracy of 99.7%, and the language model reaches an accuracy of 99.12%. The study also analyzed the performance of the YOLOv4, YOLOv4-Tiny and YOLOv7 models.

Also in this paper [16], a gesture recognition model based on YOLOv5 was proposed, which is designed to work in difficult conditions and demonstrates high recognition accuracy. The model was tested on a labeled Roboflow dataset and showed 88.4% accuracy, with 76.6% precision and 81.2% responsiveness. To improve accuracy, several images were added to the training and test sets. As part of the study, the performance of the YOLOv5 model was compared with the results of the convolutional neural network (CNN), where the latter demonstrated significantly lower accuracy – 52.98%. The YOLOv5 model was also tested for real-time detection, where it showed high efficiency and accuracy in gesture recognition. The experimental results confirmed that the proposed model successfully recognizes all alphabets of the sign language and surpasses alternative methods in accuracy and performance.

Researchers [17] present a real-time Arabic gesture recognition system based on YOLOv5. To develop the system, a dataset of 28 Arabic characters was collected, containing about 15,000 images made in various lighting conditions and on different backgrounds. Based on this dataset, various YOLOv5 variants were trained and tested, including YOLOv5s, YOLOv5m and YOLOv5l. Experiments have shown that the adapted YOLOv5 provides higher efficiency and accuracy compared to the faster R-CNN detector. The results showed satisfactory performance both in terms of output time and display accuracy. The study also revealed the advantages of YOLOv5 over R-CNN and highlighted the need for further experiments to improve the performance of YOLOv5s, as well as the prospect of comparison with YOLOX-Tiny. The system successfully recognizes all the characters of the Arabic alphabet, which confirms its effectiveness for real-time use.

In the course of the study [18], a new approach to gesture recognition in Telugu sign language (TSL) using the YOLOv5 platform was presented. The main goal was to create an accurate and effective method of gesture identification to improve communication in the deaf community. The research began by creating an extensive dataset of TSL gestures, which was carefully recorded. Then, a Deep Learning model based on the YOLOv5 architecture was developed, adapted to TSL gestures using transfer learning methods. The YOLOv5-medium model demonstrated outstanding results with 90.5% accuracy, 90.2% memorization, 90.9% F1 score and 98.1% average accuracy (mAP). These results indicate the exceptional performance of the model in gesture recognition tasks, providing a balance between computational complexity and learning time. Thorough testing and validation have confirmed the effectiveness of the YOLOv5-medium model for real-world conditions, providing an advanced sign language recognition solution.

In this context, [19] a version of the YOLOv5 model for gesture recognition was proposed, which demonstrates high efficiency in difficult conditions. The model achieved 88.4% accuracy, with 86.6% accuracy and 87.2% recall rates. A labeled Roboflow dataset was used to test the model, and additional training images were added to improve accuracy. A comparative assessment was also carried out using a convolutional neural network (CNN), where an accuracy of 91.98% was achieved. The study also conducted a targeted analysis of the popularity of Indian Sign language (ISL) in real time, which showed an accuracy of 95.7%. Prediction has been improved through the use of two levels of algorithms, which has allowed for a higher degree of similarity between characters. The results of the study confirm the high efficiency of the proposed model in gesture recognition tasks and emphasize its applicability in real conditions.

In this study [20], a new Deep Learning model was developed aimed at improving existing approaches to sign language recognition. Three advanced models based on YOLOv5x and attention methods have been proposed, including SE and CBAM modules. These models were tested on the MU HandImages ASL and OkkhorNama: BdSM datasets, demonstrating high accuracy: 98.9% and 97.6%, respectively. The models also showed outstanding results on the F1 scale, reaching 98%. The integration of attention modules has improved performance by minimizing the impact of irrelevant data and increasing recognition accuracy. The results of the study confirm the superiority of the proposed models compared to competitors in the literature, emphasizing their effectiveness and convenience for real-time deployment on modern platforms.

Current research [21] focuses on the development of a Deep Learning algorithm for real-time sign language recognition, with the aim of improving communication between deaf people and the general public. The paper presents two approaches: one for static signs using YOLOv6, and the other for continuous signs based on the LSTM and MediaPipe holistic landmarks model. The YOLOv6 model showed 96% accuracy for static signs, while the hybrid model combining LSTM and MediaPipe achieved an accuracy of about 92% for continuous signs. The hybrid approach provides reliable recognition of both static and continuous gestures. The study also revealed that YOLO is the most accurate for static signs, and LSTM is effective for continuous signs, although with limitations when increasing the number of signs.

The development of new approaches to automatic recognition of Bangla sign language (BSL) [22] includes the use of Deep Learning methods and Jetson Nano edge devices. In this study, Deep

Learning models were developed using Detectron2, EfficientDet-D0 and YOLOv7, trained on the Okkhornama database and an additional custom dataset containing 3,760 images. The images were preprocessed and resized to 416×416 pixels using the Roboflow framework. The Detectron2 model showed the best results with accuracy mAP@.5 94,915 and AP 54,814, while YOLOv7 reached mAP@.5 from 85 to 97 percent and mAP@.5-.95 from 41 to 53 percent. In order to minimize training time and ensure a high frame rate, YOLOv7 Tiny technology was selected for deployment on the Jetson Nano edge device for real time.

As a result of the conducted research [23], a system for identifying gestures in the Malayalam language was developed using advanced methods of Deep Learning and computer vision. The focus was on creating a labeled dataset for Malayalam letters and applying Deep Learning algorithms such as YOLOv8 to efficiently recognize static gestures. Experimental results have shown that the accuracy of the system identification is comparable to other existing solutions in the field of gesture recognition.

Special attention in this [24] study is paid to the development of a model for decoding sign language in real time, which is especially important for people with hearing impairments. For this purpose, the YOLOv8 model created by Ultralytics was used, which allows you to effectively translate alphabet gestures (A-Z) into text. The dataset for training and testing the model was downloaded from the Roboflow website. The model performs a preliminary extraction of key gesture components from real-time video and then transmits this data for classification to YOLOv8. The obtained results are compared with the characteristics contained in the neural network and classified according to the corresponding characteristics based on comparison with the initial data. As a result, a system has been created that generates subtitles for videos based on sign language, improving the accessibility of information and multimedia content for deaf and hard of hearing users.

The conducted research [25] indicates that using YOLOv8 for sign language recognition and translation can significantly improve communication for deaf people who face communication barriers due to lack of training and employment. As part of the study, a complete set of gesture image data with appropriate words was collected to train the YOLOv8 model, which is available in five variants with different parameters, which allows you to optimize speed and efficiency depending on GPU resources. The model demonstrated accuracy of up to 98% for converting text to sign and 89% for converting characters to text. To achieve high performance, the system includes two modules: one for translating text into sign language and the other for translating sign language into text, using advanced natural language processing technologies and YOLOv8 for real-time video analysis.

The project [26] evaluated the effectiveness of speech-to-text and text-to-speech technologies to improve communication between the deaf and the hearing. A mixed approach was used with the participation of representatives of both communities. The developed Activescan-SL system using ResNet50 achieved 99.98% accuracy in training and 100% in validation, and YOLOv8 showed 97.8% accuracy in the ASL dataset. These technologies significantly improve the accessibility of communication, but accuracy problems remain due to background noise and accents.

The study [27] analyzes CNN-based object recognition algorithms, with an emphasis on Volovan and Nolovess models trained on 50 and 100 epochs. Using the Oxford Hand and Godhand datasets, the models were evaluated by GFLOPS, average accuracy (mAP) and detection time. YOLOv8n, trained on 100 epochs, showed the best results: 86.7% accuracy on Oxford Hand and 98.9% on Gohans, surpassing previous studies. Increasing the number of epochs improves model performance, but increases processing time.

The study [28] compares machine learning models for gesture recognition of the Kazakh alphabet, such as CNN, LSTM and SVM. CNN recognizes static gestures well, but fails to deal with temporal aspects, while LSTM has shown the best results for dynamic gestures due to sequence processing. The YOLO algorithm, starting from version YOLOv3 and up to YOLOv8 and YOLO NAS, has proven its effectiveness for real-time gesture recognition due to its high speed and accuracy. The latest versions include attention modules and transformers, improving recognition in difficult environments such as different backgrounds and lighting.

## Materials and Methods

The presented YOLO-NAS architecture is an evolution of previous versions of YOLO and effectively uses the technology of neural architectural NAS search. The key components of the architecture (Figure 2) are: backbone, which extracts features; neck, which combines features; head, responsible for detecting objects; and detect head, optimized for detection at different levels. The principle of operation: the input image passes through the backbone, then the features are combined into a neck, and finally, the detect head determines the objects. Potential improvements: modification of the neck block, optimization of the head block, application of more advanced NAS methods, research of new backbone architectures, improvement of the learning process. Comparison with other versions of YOLO shows the advantages of YOLO-NAS: automatic architecture optimization, high accuracy and speed, flexibility. Further research will allow us to create even more effective models.

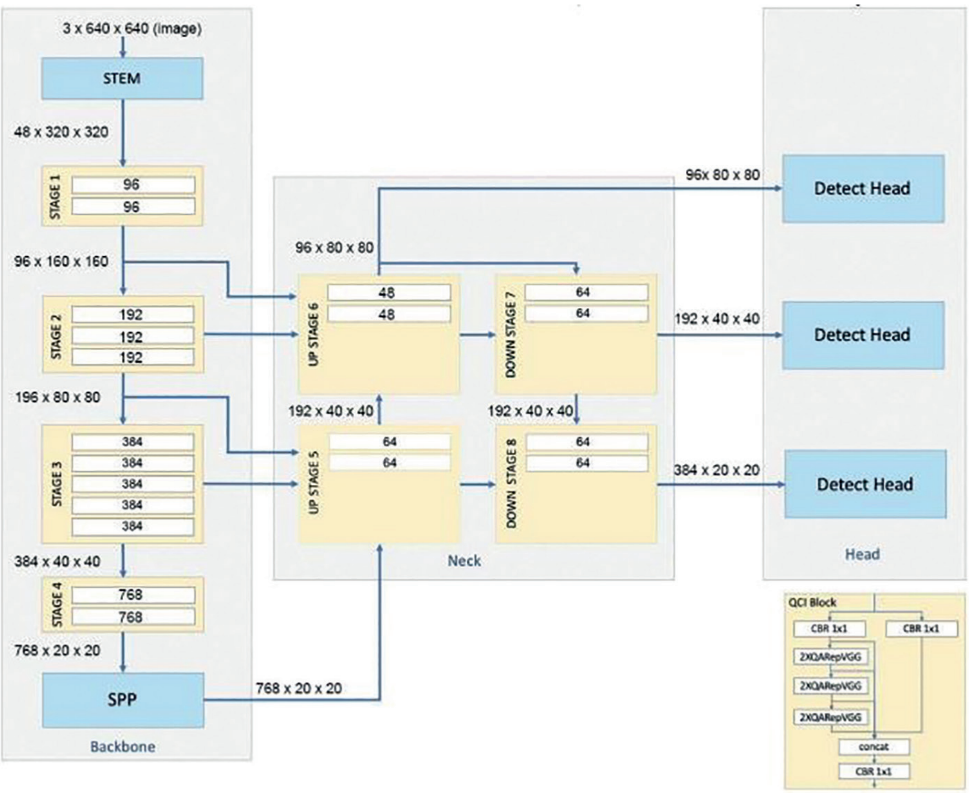


Figure 2 – YOLO-NASs architecture [29]

The basics of the YOLO algorithm. The YOLO (You Only Look Once) algorithm is a single-level neural network that allows you to perform object detection in real time due to its unique architecture. Unlike two-tier systems such as CNN, which analyze individual regions of an image, YOLO analyzes the entire image in one pass through a convolutional neural network, which significantly speeds up the process of object detection.

The use of YOLO in gesture recognition. YOLO, thanks to its ability to process images in real time, finds application in gesture recognition, which is especially important for communicating with people with hearing impairments. This technology allows systems to recognize gestures in real time, which is a significant advantage in creating interactive and accessible user interfaces.

**Results and Discussions**

The data for this study was carefully collected from videos of people demonstrating gestures in Kazakh sign language. Each frame showing an obvious hand gesture was extracted, resulting in a dataset of 5,482 images. These images represent various symbols of the Kazakh language, including its unique letters.

To ensure the correct annotation of the dataset, we used Roboflow, a reliable data annotation and augmentation service. The images were annotated manually to accurately identify hand gestures. This annotation process was very important to prepare the data for effective training of the recognition model. Roboflow played a crucial role in our research on the development of the Kazakh sign language recognition model, optimizing the processes of annotating and supplementing data. The intuitive and user-friendly interface of the platform allowed our team to effectively annotate the dataset, accurately and consistently marking each hand gesture in the images. The annotation process was laborious and took about three weeks due to the careful attention to detail required for accurate labeling. Roboflow's collaboration features allowed multiple team members to work at the same time, which significantly accelerated the annotation process. Beyond annotation, Roboflow's robust data augmentation capabilities have played an important role in expanding our dataset.

We applied several augmentation methods, including horizontal flipping, angle rotation from -45° to +45°, brightness adjustment from -15% to +15%, blur up to 2.5px and noise up to 1.49% pixels. These additions have expanded our dataset from 5,482 to 13,158 images, providing a richer and more diverse set of training data. In addition, effective data set management in Roboflow and various export options have made it easy to integrate an extended data set into our training pipeline, ensuring that our model is trained on a complete and diverse data set.

Description of the YOLO NAS model is a modern neural network architecture designed for real-time object detection. It is based on the YOLO (You Only Look Once) family of models, which are known for their speed and accuracy. YOLO NAS uses neural architecture NAS search methods to optimize the structure of the model, which makes it highly effective for detecting objects in images. We chose the YOLO NAS s because of its superior performance in real-time object detection tasks. The ability to balance speed and accuracy makes it an ideal choice for developing a sign language recognition model that can be used in real-world applications where real-time processing is crucial.

The data set was divided into three groups: training, verification and test. The initial dataset of 5,482 images was first divided into training and verification sets. After supplementing the data with various techniques (flip, rotate, brightness, blur, noise) in Roboflow, the expanded dataset increased to 13,158 images (Table 1). These augmented images were used to expand the training set, which allowed for a good generalization of the model for various real-world scenarios.

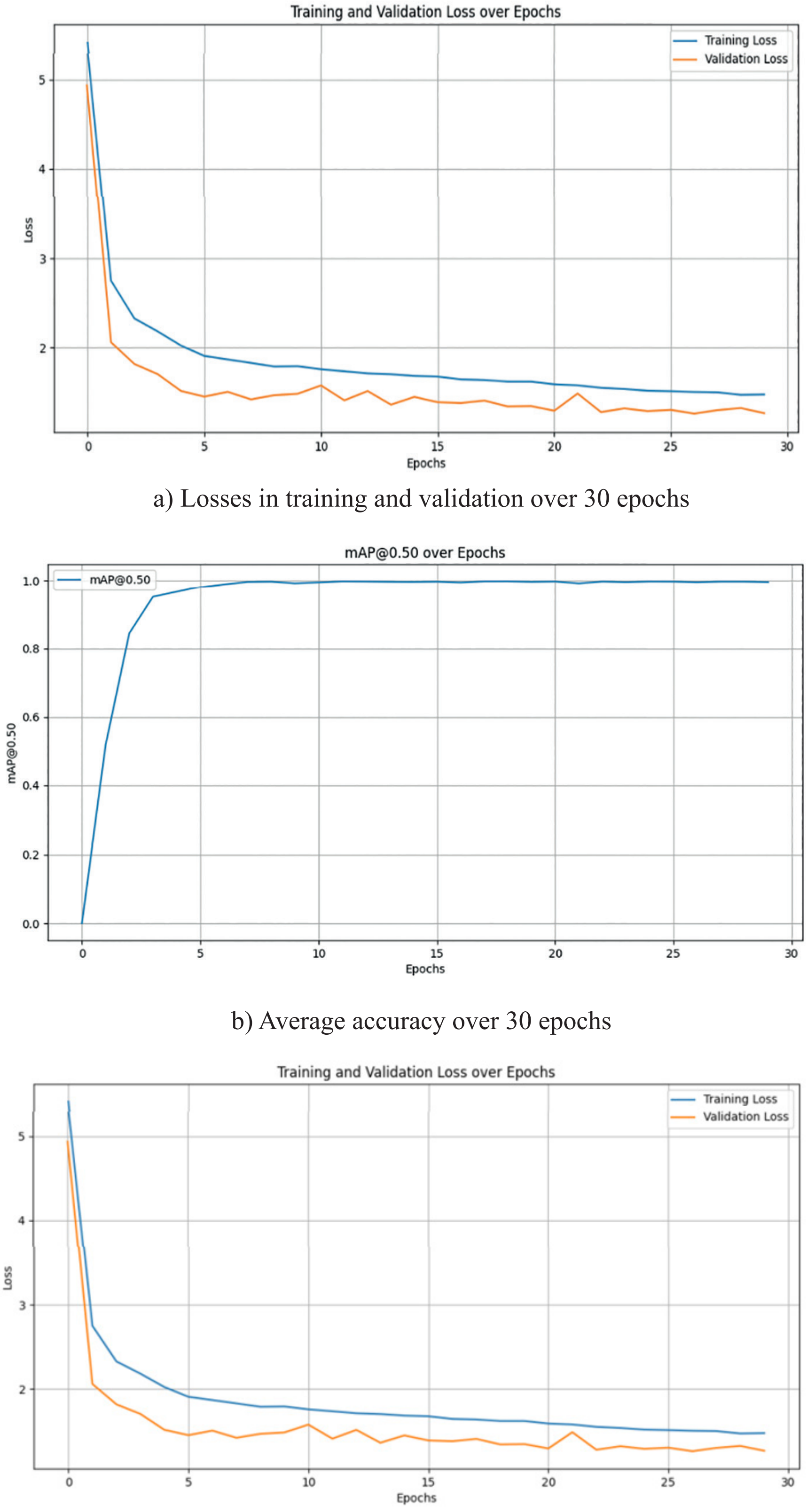
Table 1 – The data set of the model

| Model | The amount of data |
|---|---|
| YOLO NAS s | 5 482 |
| YOLO NAS s + data augmentation | 13 158 |

Parameters: The YOLO-NAS's model was trained on Google Colab to use its computing resources. The following hyperparameters were used for training:

- Learning rate (LR): The initial learning rate is set to 0.001, while the initial learning rate for the warm-up is 1 e-06, and the final ratio of LR to cosine is 0.1.
- Optimizer: Adam optimizer with a weight reduction of 0.0001.
- Package size: 16
- Number of epochs: 30
- Additional parameters: EMA (exponential moving average) with a deviation of 0.9, mixed accuracy training is enabled, and the indicator was also monitored mAP@0.50 to evaluate the model.

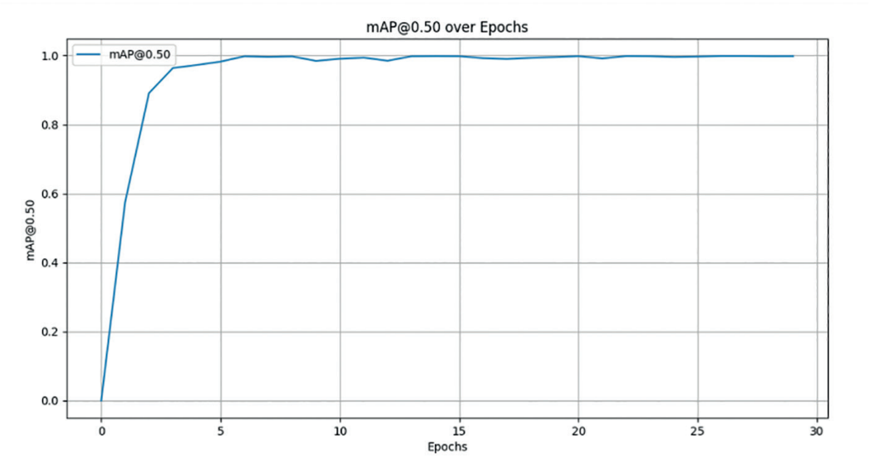
The learning outcomes were visualized using graphs that showed changes in accuracy and loss over epochs (Figure 3). This helped to identify problems with the model and identify ways to improve it.



a) Losses in training and validation over 30 epochs

b) Average accuracy over 30 epochs

c) Loss of learning and validation over 30 epochs for an extended dataset



e) Average accuracy over 30 epochs for an extended dataset

Figure 3. a), b), c), e) – The learning process

Table 2 – Result of traning

| Model | Precision | Recall | mAP |
|---|---|---|---|
| YOLO NAS s | 99.1% | 99.1% | 99.4% |
| YOLO NAS s + data augmentation | 99.6% | 99.8% | 99.7% |

The model developed for recognizing the letters of the Kazakh sign language was evaluated using the accuracy metric (mAP). After training, the YOLO NAS s model achieved 99.4% accuracy. After applying augmentation methods, the YOLO-NAS's + data augmentation model improved its results, showing an accuracy of 99.7% on the evaluation set, and also improved its performance on the test set by almost 0.3% (Table 2).

Table 3 – Comparative analysis of the current study of the Kazakh sign language and other sign languages

| References | Sign language | Metrics | Число эпох | Method |
|---|---|---|---|---|
| [20] | ASL, BdSL | Precision: 98.9% and 97.6% | 100, 200 and 300 | Experimental method: YOLOv5x and attention methods, including SE and CBAM modules |
| [26] | ASL | Precision 97.8% | 50 | Mixed method: quantitative approach (ResNet50 and YOLOv8) and qualitative approach |
| YOLO NAS s | KSL | Precision 99.1%<br>Recall 99.1%<br>mAP 99.4% | 30 | Automatic search and optimization of neural network architecture in the context of object detection |
| YOLO NAS s + data augmentation | KSL | Precision 99.6%<br>Recall 99.8%<br>mAP 99.7% | 30 | Automatic search and optimization of neural network architecture in the context of object detection and data augmentation |

The SOLO-NAS's model and its improved version YOLO-NAS's + data augmentation significantly outperform the methods from sources [20] and [26] in key accuracy metrics. This indicates the high potential of automatic search and optimization of the neural network architecture, as well as the positive impact of data augmentation on model performance (Table 3).

In this study, the YOLO-NAS s model for recognition of the Kazakh sign language (KSL) was proposed, and the results showed its high efficiency, especially after applying data augmentation methods. Training the model on Google Collab using various hyperparameters, such as learning rate, optimizer and package size, gave excellent results. Data augmentation, which includes techniques for flipping, rotating, changing brightness, blurring, and adding noise, has significantly improved model performance. The accuracy (mAP) increased from 99.4% to 99.7%, which confirms the positive effect of expanding the dataset on the generalizing ability of the model.

However, despite the high results, there are several shortcomings and areas for further improvement:

Despite the improvements, the model may be less effective when working with gestures that were not presented in the training dataset. This can be especially noticeable in real-world settings, where gesture variations can be much more diverse.

Although the model shows high metrics, it may encounter problems in real-world conditions where fast real-time processing is required. Optimizing the model for faster and more efficient operation is an important task for its practical application.

The effectiveness of the model largely depends on the quality and variety of data augmentation methods. Suboptimal or insufficient augmentation methods may lead to insufficient improvement in the performance of the model or its retraining.

Training a model on Google Collab requires significant computing resources. When resources are limited or the model needs to be scaled for a larger dataset, difficulties may arise.

Although the results in the control conditions are impressive, additional testing and verification of the model in various real-world scenarios is necessary to ensure its reliability and stability.

These shortcomings indicate the need for further research and development to improve the adaptability of the model, optimize its operation and verify its effectiveness in various conditions.

## Conclusions

In this study, the YOLO-NAS s model for Kazakh sign language recognition (KSL) was developed and evaluated. The results showed that the proposed model demonstrates high efficiency in the task of gesture recognition, especially after applying data augmentation methods. Data augmentation has significantly improved the performance of the model, improving its accuracy (mAP) from 99.4% to 99.7%.

The YOLO-NAS's model and its improved version, YOLO-NAS's + data augmentation, have demonstrated excellent results in classification and gesture recognition tasks. These results confirm that the combination of automatic search and optimization of the neural network architecture with the expansion of the dataset can significantly improve the accuracy and generalization ability of models.

However, despite the progress made, there are several areas for improvement. The model may require additional optimizations to improve real-time processing speed and adapt to new gestures that were not included in the training dataset. It is also worth considering the need to test the model in various real-world scenarios to ensure its reliability and stability.

In the future, it is important to continue research in these areas in order to make the model even more effective and versatile. This will not only improve its performance in real-world applications but also expand its capabilities for gesture recognition in other languages and contexts.

## REFERENCES

1 Daniels S., Suciati N., Fathichah C. Indonesian sign language recognition using YOLO method. IOP Conference Series: Materials Science and Engineering. IOP Publishing, 2021, vol. 1077, no. 1, p. 012029. https://doi.org/10.1088/1757-899x/1077/1/012029.

2 Al Ahmadi S., Mohammad F., Al Dawsari H. Efficient YOLO Based Deep Learning Model for Arabic Sign Language Recognition, 2024. https://doi.org/10.21203/rs.3.rs-4006855/v1.

3 Mesbahi S. C. et al. Hand gesture recognition based on various Deep Learning YOLO models. International Journal of Advanced Computer Science and Applications, 2023, vol. 14, no. 4.

4 Doždor Z. et al. TY-Net: Transforming YOLO for hand gesture recognition. IEEE access., 2023. https://doi.org/10.14569/ijacsa.2023.0140435.

5 Yerraboina S. Real-Time Hand Gesture Recognition System, 2024.

6 Mallikarjuna Swamy N. et al. Indian sign language detection using YOLOv3. High Performance Computing and Networking: Select Proceedings of CHSN 2021. Singapore: Springer Singapore, 2022, pp. 157–168. https://doi.org/10.1007/978-981-16-9885-9_13.

7 Asri M. et al. A real time Malaysian sign language detection algorithm based on YOLOv3. International Journal of Recent Technology and Engineering, 2019, vol. 8, no. 2, pp. 651–656. https://doi.org/10.35940/ijrte.b1102.0982s1119.

8 Khaliluzzaman M., Kobra K., Liaqat S. Comparative analysis on real-time hand gesture and sign language recognition using convexity defects and YOLOv3. Sigma Journal of Engineering and Natural Sciences, 2024, vol. 42, no. 1, pp. 99–115. https://doi.org/10.14744/sigma.2024.00012.

9 Mujahid A. et al. Real-time hand gesture recognition based on Deep Learning YOLOv3 model. Applied Sciences, 2021, vol. 11, no. 9, p. 4164. https://doi.org/10.3390/app11094164.

10 Lawand S. J. et al. Sign Language Hand Gesture Identification Using YOLOv3. Available at SSRN 4385690. http://dx.doi.org/10.2139/ssrn.4385690.

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

12 Al-shaheen A., Çevik M., Alqaraghulı A. American sign language recognition using YOLOv4 method. International Journal of Multidisciplinary Studies and Innovative Technologies, 2022, vol. 6, no. 1, pp. 61–65, https://doi.org/ 10.36287/ijmsit.6.1.61.

13 Alaftekin M., Pacal I. & Cicek K. Real-time sign language recognition based on YOLO algorithm. Neural Comput & Applic, 2024, vol. 36, pp. 7609–7624. https://doi.org/10.1007/s00521-024-09503-6.

14 Sreemathy R. et al. Continuous word level sign language recognition using an expert system based on machine learning //International Journal of Cognitive Computing in Engineering, 2023, vol. 4, pp. 170–178. https://doi.org/10.1016/j.ijcce.2023.04.002.

15 Begum N. et al. Borno-net: a real-time Bengali sign-character detection and sentence generation system using quantized YOLOv4-Tiny and LSTMs //Applied Sciences, 2023, vol. 13, no. 9, p. 5219. https://doi.org/10.3390/app13095219.

16 Bankar S. et al. Real time sign language recognition using Deep Learning. International Research Journal of Engineering and Technology, 2022, vol. 9, no. 4, pp. 955–959. https://doi.org/10.22214/ijraset.2023.55621.

17 Aiouez S. et al. Real-time Arabic Sign Language Recognition based on YOLOv5. IMPROVE, 2022, pp. 17–25. https://doi.org/10.5220/0010979300003209.

18 Reddy P.V. et al. Sign Language Recognition based on YOLOv5 Algorithm for the Telugu Sign Language. arXiv e-prints, 2024. C. arXiv: 2406.10231, https://doi.org/10.48550/arXiv.2406.10231.

19 Venkatraja V.M.C. et al. Sign language to speech converter for indian languages, 2023.

20 Attia N.F., Ahmed M.T.F.S., Alshewimy M.A.M. Efficient Deep Learning models based on tension techniques for sign language recognition. Intelligent systems with applications, 2023, vol. 20, p. 200284. https://doi.org/10.1016/j.iswa.2023.200284.

21 Buttar A.M. et al. Deep Learning in sign language recognition: a hybrid approach for the recognition of static and dynamic signs. Mathematics, 2023, vol. 11, no. 17, p. 3729. https://doi.org/10.3390/math11173729.

22 Siddique S. et al. Deep Learning-based bangla sign language detection with an edge device. Intelligent Systems with Applications, 2023, vol. 18, p. 200224. https://doi.org/10.1016/j.iswa.2023.200224.

23 Nair A. B. et al. Malayalam Sign Language Identification using Finetuned YOLOv8 and Computer Vision Techniques. arXiv preprint arXiv:2405.06702, 2024. https://doi.org/10.48550/arXiv.2405.06702.

24 Kalimuthu S. Video Captioning Based on Sign Language Using YOLOV8 Model, 2023. https://doi.org/10.1007/978-3-031-45878-1_21.

25 Hinge R. et al. Improving Indian Sign Language Interpretation with Deep Learning-Based Translation System. Journal of technical education, p. 44.

26 ZainEldin H. et al. Active convolutional neural networks sign language (ActiveCNN-SL) framework: a paradigm shift in deaf-mute communication. Artificial Intelligence Review, 2024, vol. 57, no. 6, p. 162. https://doi.org/10.1007/s10462-024-10792-5.

27 Purnomo H. et al. Utilizing the YOLOv8 Model for Accurate Hand Gesture Recognition with Complex Background. Available at SSRN 4777516. http://dx.doi.org/10.2139/ssrn.4777516.

28 Mukhanov S. et al. Gesture recognition of machine learning and convolutional neural network methods for kazakh sign language. Scientific Journal of Astana IT University, 2023, pp. 85–100. https://doi.org/10.37943/15lpcu4095.

29 Tang Y., Wang Y., Qian Y. Real-time railroad track components inspection framework based on YOLO-NAS and edge computing. IOP Conference Series: Earth and Environmental Science. – IOP Publishing, 2024, vol. 1337, no. 1, p. 012017. https://doi.org/10.1088/1755-1315/1337/1/012017.

[1]**Отман М.,**
профессор, ORCID ID: 0000-0002-5124-5759,
e-mail: mothmanupm@gmail.com
[2]**Оралбекова Д.,**
сениор-лектор,
e-mail: dinaoral@mail.ru
[2]***Бержанова У.Г.,**
докторант, ORCID ID: 0009-0000-2467-5721
*e-mail: berzhanovaulmekenn@gmail.com

[1]Путра Малайзия университеті, Куала-Лумпур қ., Малайзия
[2]әл-Фараби атындағы Қазақ ұлттық университеті, Алматы қ., Қазақстан

# YOLO NAS НЕГІЗІНДЕ ҚАЗАҚ ЫМ ТІЛІН ТАНУ МОДЕЛІН ӘЗІРЛЕУ

## Аңдатпа

Қазақ ымдау тілін танудың сенімді моделін әзірлеу – есту қабілеті бұзылған адамдар үшін инклюзивті коммуникацияны дамыту мен қолдау көрсетуді жетілдіру жолындағы маңызды қадам. Бұл жұмыс қимыл кескіндерін пайдалану арқылы деректерді жинау және аннотациялау процесін жан-жақты сипаттайды. Деректерді модельмен үйлесімді етіп дайындау және алдын ала өңдеу кезеңдеріне ерекше назар аударылды. Модельді үйрету барысында гиперпараметрлерді оңтайландыру және тану дәлдігін арттыру үшін әртүрлі әдістер қолданылды. Сонымен қатар, модельдің нақты жағдайларда тиімділігін бағалау мақсатында сынақ деректеріне негізделген кешенді өнімділік талдау жүргізілді. Негізгі әзірлеу кезеңімен қатар, дәлдік пен өнімділікті одан әрі жақсарту мүмкіндіктерін зерттеу мақсатында деректер жиынында YOLO NAS үлгісі сыналды. Зерттеу нәтижелері қазақ ымдау тіліне негізделген инклюзивті технологияларды жетілдіруге, сондай-ақ есту қабілеті бұзылған адамдардың қоғамға интеграциялануын қолдауға бағытталған білім беру және коммуникациялық платформаларды дамытуда пайдаланыла алады.

**Тірек сөздер:** You Only Look Once (YOLO), YOLO NAS, тереңдетіп оқыту, конволюционды нейрондық желі (CNN), жасанды интеллект, қазақша ымдау тілі.

[1]**Отман М.,**
профессор, ORCID ID: 0000-0002-5124-5759,
e-mail: mothmanupm@gmail.com
[2]**Оралбекова Д.,**
сениор-лектор,
e-mail: dinaoral@mail.ru
[2]***Бержанова У.Г.,**
докторант, ORCID ID: 0009-0000-2467-5721,
*e-mail: berzhanovaulmekenn@gmail.com

[1]Университет Путра Малайзия, г. Куала-Лумпур, Малайзия
[2]Казахский национальный университет им. аль-Фараби, г. Алматы, Казахстан

# РАЗРАБОТКА МОДЕЛИ РАСПОЗНАВАНИЯ КАЗАХСКОГО ЯЗЫКА ЖЕСТОВ НА ОСНОВЕ YOLO NAS

**Аннотация**

Разработка надежной модели распознавания казахского жестового языка является важным шагом на пути к развитию инклюзивной коммуникации и помощи людям с нарушениями слуха. В данной работе подробно описывается процесс сбора и аннотирования данных, в которых использовались изображения жестов. Особое внимание уделяется подготовке и предварительной обработке данных для обеспечения их совместимости с моделью. Процесс обучения модели включает оптимизацию гиперпараметров и использование различных методов для повышения точности распознавания. Мы также провели комплексную оценку производительности модели на основе тестовых данных, чтобы убедиться в ее эффективности в реальных условиях. Помимо основного этапа разработки мы рассматриваем возможность тестирования модели YOLO-NAS на том же наборе данных для изучения потенциальных улучшений точности и производительности. В заключение следует отметить, что результаты нашего исследования могут быть использованы для дальнейшей разработки технологий, способствующих интеграции людей с нарушениями слуха в общество, а также для создания образовательных и коммуникационных платформ на основе казахского жестового языка.

**Ключевые слова:** You Only Look Once (YOLO), YOLO-NAS, глубокое обучение, сверточная нейронная сеть (CNN), искусственный интеллект, казахский язык жестов.

Article submission date: 07.11.2024