

# Application of neural networks ensemble method for the Kazakh sign language recognition

Yedilkhan Amirgaliyev<sup>1</sup>, Aisulyu Ataniyazova<sup>1,2</sup>, Zholdas Buribayev<sup>2</sup>, Mukhtar Zhassuzak<sup>1,2</sup>,  
Baydaulet Urmashhev<sup>2</sup>, Lyailya Cherikbayeva<sup>1,2</sup>

<sup>1</sup>Department of Artificial Intelligence and Robotics, Institute of Information and Computational Technologies, Almaty, Kazakhstan

<sup>2</sup>Department of Computer Science, Faculty of Information Technologies, Al-Farabi Kazakh National University, Almaty, Kazakhstan

## Article Info

### Article history:

Received Aug 21, 2023

Revised Mar 3, 2024

Accepted Mar 20, 2024

### Keywords:

Classification

Dactyl alphabet

Deep learning

Ensemble

Gesture recognition

## ABSTRACT

Sign languages are an extremely important means of communication in many cases, especially for deaf and hard of hearing people. But the same gesture can convey different meanings in different countries, so many different sign languages have been developed all over the world. In this study, a convolutional neural network (CNN) model was developed based on an ensemble method containing the ResNet-50 and VGG-19 architectures, which will be able to classify the Kazakh sign language (KSL) consisting of 42 Kazakh alphabet signs (classes). A proprietary data set of 57,708 images for 42 signs of the KSL has been formed. The ensemble model was compared with ResNet-50 and VGG-19 by evaluation metrics such as accuracy, precision, recall, f1-measure, and loss function. The recognition accuracy of the ensemble method reached 95.7%, exceeding the performance of ResNet-50 and VGG-19. The developed method was also tested on test data, where 35 out of 42 gestures were recognized completely correctly. The reliability of the proposed approach and the classification results obtained by using preprocessing methods and data augmentation techniques to expand the data set was confirmed by a computational experiment.

*This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.*



## Corresponding Author:

Lyailya Cherikbayeva

Department of Artificial Intelligence and Robotics

Institute of Information and Computational Technologies

Almaty 050010, Kazakhstan

Email: cherikbayeva.lyailya@gmail.com

## 1. INTRODUCTION

According to the data presented by the World Health Organization (WHO) [1], more than 430 million people suffer from hearing problems, which makes it difficult for them to communicate with other people. For the hearing impaired and people with severe hearing impairment, sign language is the main communication tool [2]. But sign speech is not capable of conveying all the words of ordinary language, therefore, for words that do not have their own analogue of a gesture, their meaning is shown letter by letter using a dactylem. This led to the development of a sign language recognition system capable of identifying dactyl language gestures through efficient preprocessing and accurate character classification.

Sign language serves as a communication tool for people with conductive and sensorineural speech perception disorders in order to build a society without any restrictions, therefore, the relevance of this research work is explained by its social significance, which solves the problems of transmitting and understanding information through gesture without linguistic means. In this regard, in recent decades, special attention has been paid to the implementation of gesture recognition systems using the capabilities of modern

computer technologies [3], [4]. The results of automatic sign language recognition are used not only to bridge the communication gap between ordinary and deaf people, but will also lead to improved digitalization of human-computer interaction. However, the majority of research efforts have focused on widely spoken sign languages, leaving lesser-known sign languages, such as Kazakh sign language (KSL), underrepresented. The Kazakh language is distinguished by the presence of a complex grammar, many endings, the presence of several types of past and future tenses, which makes it very difficult for native speakers to understand sign language. This paper proposes the use of ensemble neural networks as a promising approach to solve the inherent problems associated with KSL recognition.

Existing gesture recognition approaches can be divided into two categories: based on wearable devices and based on computer vision. Methodologies were explored aimed at obtaining a representation space that allows one to identify the dynamics of hand movements using Microsoft Kinect [5], [6]. The preprocessing steps for subtracting the region of interest from the original image for the sensor-based recognition method were presented in [7], [8]. But the use of sensors in the form of special gloves has not become widespread due to the inconvenience of use, so many recent works in the field of gesture recognition are focused on methods based on machine learning [9]–[13].

A systematic review conducted on sign language recognition technology found that the commonly used methods for data collection are the Microsoft Kinect device and the camera. The advantage of Kinect is that it provides depth data for the video stream, which makes it easy to distinguish between background and signer. Microsoft Kinect was used to develop an applied gesture recognition system, but it was noticed that the device is expensive and must be connected to a computer [5]–[8], [14]. It can also be affected by lighting conditions, segmentation of the hands and face, complex background and noise [15]. The main advantage of using the camera is that it does not require the wearing of other external devices and users only need to use their hands within the range of the camera. To create sign recognition system [11], [16] used a high-performance webcam, removing the need for sensors in sensory gloves. But computer vision-based research will require several image processing techniques that can affect recognition accuracy.

Classification methods also vary among researchers. They tend to develop their own concept based on known methods to give a better result in sign language recognition. In works [17]–[19], classical machine learning algorithms were applied, but recently a convolutional neural network (CNN) is a method that is gaining great popularity in studies of sign language recognition [20]–[22]. There are many recorded research systems where the importance of a computer vision-based methods for effective gesture recognition has been accentuated due to a comprehensive review [21], [23], [24]. Sharma *et al.* in [25] CNN-based model using a camera for processing continuous images is proposed. The advantage of this study is to reduce the susceptibility of channels to the influence of noise. Also, the work has a number of drawbacks, such as the effect of camera quality on good performance and the speed of sign coordination with content.

Previous studies of KSL were carried out on a pre-trained model of the Russian language, since the Kazakh alphabet contains all the letters of the Cyrillic alphabet [26], [27]. 78.5% similarity of the Kazakh and Russian alphabets circumscribed the recognition of the Kazakh dactyl language by the scientific community as a individual sign language with its own specific vocabulary. The studies devoted to the development of the KSL recognition application in [19], [28] were not sufficient to obtain results worthy of global use. The ensembling of CNN models contributes to the development of a gesture recognition system with the highest possible accuracy. By combining multiple models, the ensemble can gain a more complete representation of the underlying patterns in the data, making it more effective at recognizing gestures in different contexts and environments.

Based in our work, we proposed a system for recognizing static hand gestures using the ensemble of CNNs such as ResNet-50 and VGG-19 in order to reduce the variance of predictions. The relationship between the ensemble and its constituent neural networks in the context of classification is also presented, which shows that the ensemble method significantly improves the quality of recognition by combining the functional advantages of different deep learning algorithms and reduces the time spent on training. To achieve this goal, a number of specific tasks were set that require urgent solutions:

- Creating own data set for 42 gestures based on the dactyl alphabet of the Kazakh language, with the addition of images of right and left hands on various backgrounds;
- Training of ResNet-50 and VGG-19 CNNs for the classification of static gestures;
- Development of an ensemble-based meta model that enhances the quality of recognition of individual neural network models.

## 2. METHOD

Developing an ensemble method for gesture recognition based on the integration of ResNet-50 and VGG-19 involves combining the strengths of both models to improve the overall accuracy and performance

of the gesture recognition system. The development of an automatic sign recognition system required the creation of a database for the KSL. The initial step in this direction was the creation of a database consisting of the dactyl alphabet (Figure 1) of forty-two gestures presented at [28].



Figure 1. Dactylic alphabet of the KSL [28]

The data set used was formed on the basis of the dactyl alphabet of the KSL. The dataset is balanced and consists of 57,708 images for 42 signs of the KSL (42 classes). Figure 2 shows a sample of each gesture corresponding to each letter of the Kazakh alphabet. During the training process, augmentation was used, which allows increasing the initial data set by changing the original image (using mirror reflection and image rotation). Therefore, it is recommended to use the left hand for fingerprinting and any "one-handed gestures".



Figure 2. Kazakh dactyl alphabet

The signs "Б" and "Б" are used only when writing words of foreign language origin (Figure 3). The main function of these signs is to soften the consonant in front of it: "медаль-medal". There is no phonemic distinction between these consonant sounds in English.

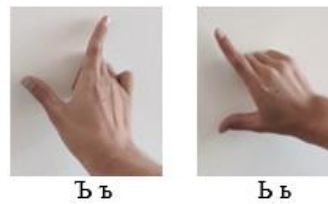


Figure 3. Distinctive signs of Kazakh alphabet

The photographs were obtained using video taken with different backgrounds, from different angles in different lighting conditions. Signs for left-handed people were obtained using the mirroring technique (Figure 4). Short video clips were cut into frames, and the resulting images were reduced to the desired square input image of  $224 \times 224$ . The training data set was expanded by augmenting the samples included in it, that is, by reflecting each photograph about the vertical axis.

ResNet-50 and VGG-19 were selected as the base ensemble models. ResNet-50 is known for its deep architecture and skip connections, which help in capturing fine-grained details. VGG-19, on the other hand, has a simpler architecture but is effective in capturing general features.



Figure 4. Images of the same gesture on different backgrounds

## 2.1. ResNet-50

The ResNet-50 model consists of 5 stages, each of which has a convolution and identification block. Each convolution block has 3 convolution layers, and each identity block also has 3 convolution layers [29]. Before submitting the image, its size was reduced to  $224 \times 224$ . Images are convolved  $7 \times 7$  with 64 different kernels with a stride of 2, then max pooling is used with a stride of 2 (Figure 5). After the resulting matrix, convolution is used 3 times using the kernels  $1 \times 1$ -64  $3 \times 3$ -64,  $1 \times 1$ -256. Further, the convolution  $1 \times 1$ -128,  $3 \times 3$ -128,  $1 \times 1$ -512 is used 4 times. Then  $1 \times 1$ -256,  $3 \times 3$ -256,  $1 \times 1$ -1024 convolution was performed 6 times.

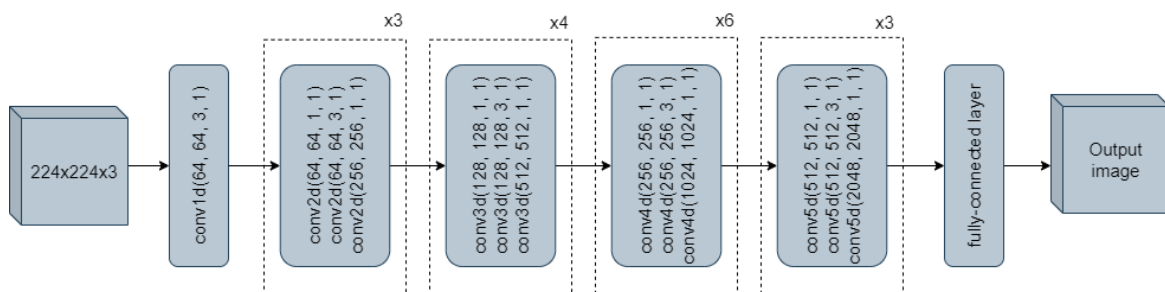


Figure 5. ResNet-50 architecture

The last convolution is applied 3 times using kernels  $1 \times 1$ -512,  $3 \times 3$ -512 and  $1 \times 1$ -2048. At the end, average pooling is used to average the result and a fully connected layer with 42 nodes equal to the number

of classes. Images are given to the neural network in groups of 8 at a time, when all the images pass one epoch ends, as a result of experiments, it was concluded that 20 epochs would be enough. The cross-entropy loss function, the Adam optimizer and the rectified linear unit (ReLU) activation function were applied.

## 2.2. VGG-19

The VGG-19 architecture is built from 16 convolutional layers which are used for feature extraction and the next 3 layers work for classification [30]. The layers used for feature extraction are divided into 5 groups, where each group is followed by the layer with the maximum pooling. A  $224 \times 224$  image is input to this model, and the model outputs an object label on the image (Figure 6).

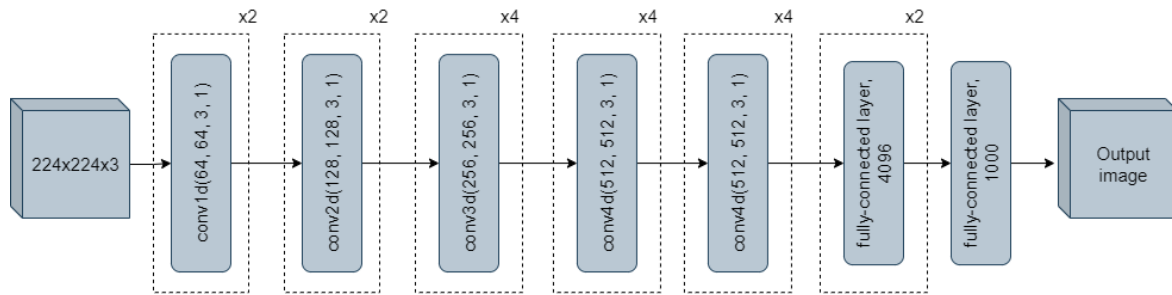


Figure 6. VGG-19 architecture

At all stages of convolution, a kernel of size  $[3 \times 3]$  was used. One of the main differences between VGG and previous architectures (previously used tangential and sigmoid functions) is the use of ReLU activation functions to improve model classification and reduce computation time. After completing the last stage of convolution, three fully connected layers were implemented, of which the first two had a size of 4096, and then a layer with 1000 channels for classification.

## 2.3. Proposed ensemble method

The main task of neural networks with deep learning is to solve optimization problems. There were works where there were claims that the method of ensemble training of several neural networks is not quite effective due to the distribution of losses [31]. However, experimental results were shown in [32], where ensemble methods showed good performance compared to the basic models. Thus, in addition to the two deep learning architectures, it was decided to use an ensemble method based on the same architectures to classify the Kazakh dactyl alphabet. Usually, this approach requires big data. The classification is calculated based on the number of votes of the models.

For a mathematical description of the ensemble method of CNNs, some evaluation parameters were first given:  $B = \{x_i, y_i\}^n$  – data set,  $\{h_1, h_2, \dots, h_m\}$  – base estimators,  $o_i^m$  – output from the baseline  $h_m$  for the set  $x_i$ , and  $L(o_i, y_i)$  is the cross-entropy error function for calculating  $o_i$  the output at the target value  $y_i$  of class  $i$ . Since we use VGG-19 and ResNet50, in our case the value for  $m=2$ , and the value of  $n=42$ , because the number of letters in the alphabet is 42. The basis of the work of the ensemble method [33] is the averaged approach of merging the estimates of all algorithms  $m$ . For output  $o_i$ , this approach will look like this:

$$o_i = \frac{1}{m} \sum_{j=1}^m o_i^j \quad (1)$$

During training, all merge estimation algorithms are trained together using small gradient descent. The loss function will be:

$$\frac{1}{n} \sum_{i=1}^n L(o_i, y_i) \quad (2)$$

The main goal is to minimize the loss function and for this we used gradient boosting, which builds a prediction model in the form of an ensemble of weak predictive models. This approach trains all estimation algorithms sequentially, that is, the value  $h_m$  will directly depend on previously trained estimators  $h_1, h_2, \dots, h_{m-1}$

For output  $o_i$ , the gradient descent compression rate parameter will be multiplied  $\varepsilon$  and will look like:

$$o_i = \frac{1}{m} \sum_{j=1}^m o_i^j \varepsilon \quad (3)$$

where  $0 < \varepsilon \leq 1$ .

In this approach, as already mentioned, the value up to the penultimate  $m - 1$  estimate will accumulate for class  $i$  with a target value it will look like:

$$O_i = \sum_{k=1}^{m-1} o_i^k \varepsilon \quad (4)$$

After that, it will be necessary to determine learning target for class  $i$ :

$$r_i^m = -\frac{\partial L(O_i, Y_i)}{\partial o_i} \quad (5)$$

The calculation of the loss for  $m$ th algorithm will be as (6):

$$l^m = \frac{1}{n} \sum_{i=1}^n \|r_i^m - o_i^m\|_2^2 \quad (6)$$

We will update the output for each next  $m + 1$ :

$$O_{i+1} = O_i + \varepsilon o_i \quad (7)$$

Thus, at each step, this algorithm tries to minimize the errors made in the previous steps. However, it does not change the weights, but will train the next model on the residual errors of the previous estimator.

Since we are doing a classification, cross-entropy will be chosen for this and we will use the Softmax (SM) function for the probability of estimating that the output belongs to all classes:

$$r_i^m = Y_i - \text{SM}(O_i), \quad (8)$$

where  $Y_i$  is the label vector for class  $i$ .

The advantage of this approach is the determination of the importance of each feature. In addition, the algorithm shows a very good recognition result due to the interaction of the model with each other (Figure 7). However, this approach requires high computational complexity and strictly sequential training.

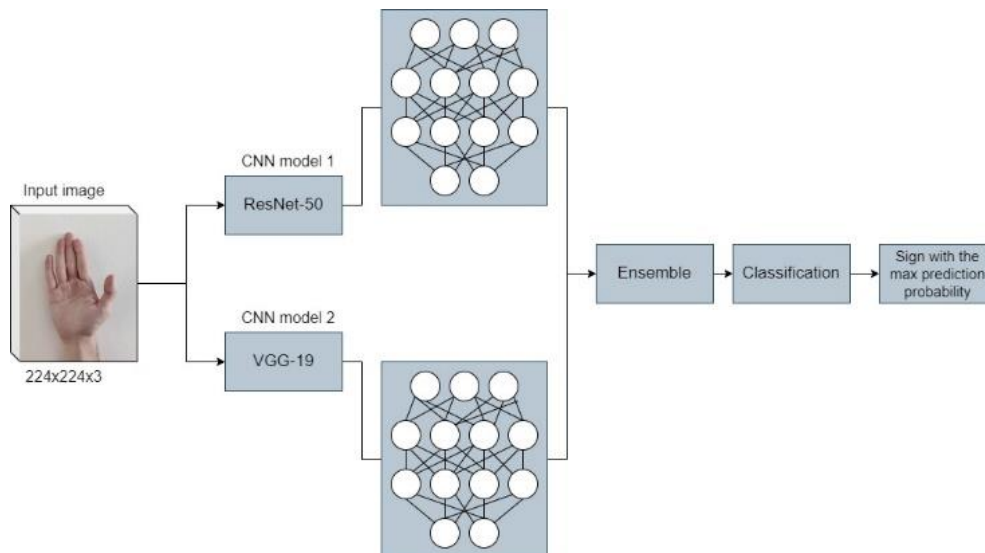


Figure 7. Proposed method architecture

An image of a gesture/hand is taken as input, and the size of each image in the dataset is  $224 \times 224$  pixels. After the image is transferred to two models (ResNet-50 and VGG-19). They also have been pre-trained to take advantage of training instead of starting the model weights with random initializations. In this

way, the functional advantages of each model are combined in an ensemble. The meta model receives the results of separately trained neural models as input and returns the gesture class with the maximum prediction probability as the final result. The proposed ensemble method is described by the pseudocode presented in Algorithm 1.

Algorithm 1. Pseudo code for the proposed method

```

class EnsembleModule(Module)
  procedure init(modelA, modelB)
    init()
    classifier ← linear transformation(42 * 2, 42)
  end procedure
  procedure forward(x):
    x1 ← modelA(x)
    x2 ← modelB(x)
    x ← concatenate((x1, x2), dim=1)
    out ← classifier(x)
    return out
  end procedure
end class
ensemble ← EnsembleModule(resnet50, vgg19)
for param in ensemble model parameters():
  param.requires_grad ← False
end for
for param in ensemble model classifier parameters():
  param.requires_grad ← True
end for
ensemble_training_results=training(ensemble_model, 20epoch)

```

The parameters of ensemble neural networks can vary depending on the specific implementation and requirements of the ensemble model. Table 1 shows the parameters of the proposed method. Adam was chosen as the optimization algorithm, it corrects the weights and biases of the neural network well during training in order to minimize the loss function.

Table 1. Summary of the developed ensemble model

Parameters	Ensemble
Number of layers	50 layers of ResNet+19 layers of VGG
Number of neurons in fully connected layers	42
Activation function	ReLU
Optimizer	Adam
Batch size	8
Loss function	Cross entropy
Number of epochs	20

### 3. RESULTS AND DISCUSSION

Hand gesture recognition accuracy is evaluated using a test set that splits a dataset of 57,708 images with a ratio of 80% of the training set to 20% of the test set, which is 46,166 samples out of 11,542 samples, and our results are based on these 11,542 images. Figure 8(a) shows the recognition accuracy of the ensemble model compared to the accuracy of ResNet-50 and VGG-19, whereas Figure 8(b) shows the result of the loss function of the ensemble method against the result of individual architectures. It is noticeable from the graphs a positive correlation between the recognition accuracy results at 20 epochs.

Graphical interpretation of precision metrics and recall shown in Figures 9(a) and (b), and it provide us that VGG-19 is not quite able to distinguish the given class from all other classes. Also, as a result of the classification of VGG-19, most of the positive examples are lost. ResNet-50 and ensemble demonstrated excellent ability to detect a particular class, having high recall metrics.

Figure 10 presented the metric F1, which is defined as the harmonic mean of the precision and recall metrics. By evaluating this metric, can balance the accuracy for non-uniform class distributions. But in our case, it is unimportant, since the data was evenly distributed into classes, and both metrics showed almost the same result. The numerical indicators of the average estimates of each model are presented in Table 2.

If at the first iteration the accuracy of Res-Net-50 was 80.97%, then at the next iteration the neural network already produces a positive trend in learning, showing a result of over 90% (Figure 11). Comparatively, VGG-19 scores poorly on precision, recall, and F1. But the loss result is much less than that of ResNet-50. The ensemble method for all evaluation metrics showed high results, thereby ensuring the effectiveness of our developed approach.

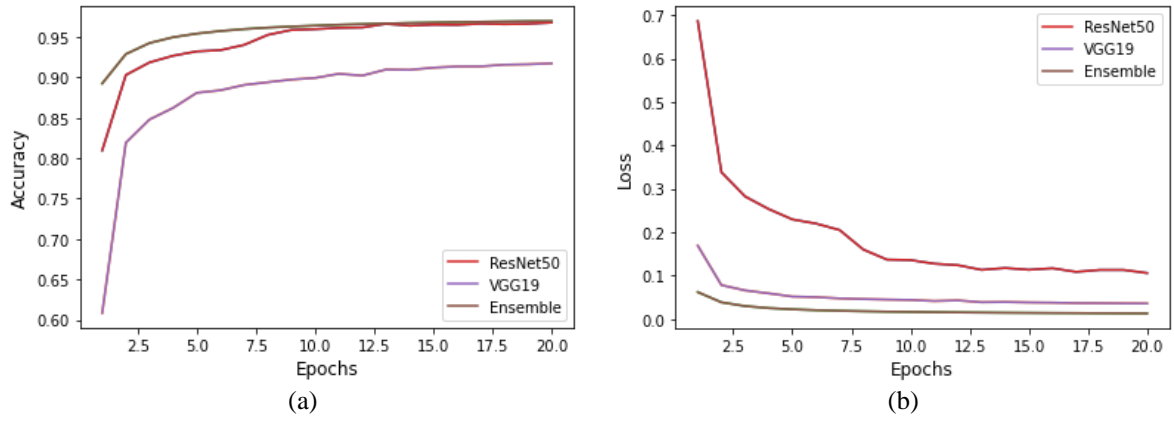


Figure 8. Comparative evaluations of CNN models and ensemble; (a) accuracy and (b) loss function

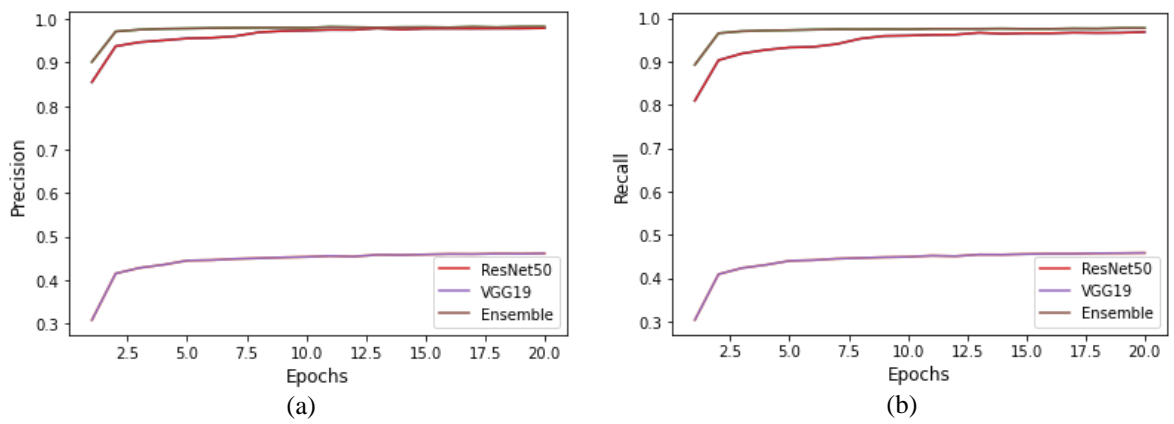


Figure 9. Comparative evaluations of CNN models and ensemble; (a) precision and (b) recall

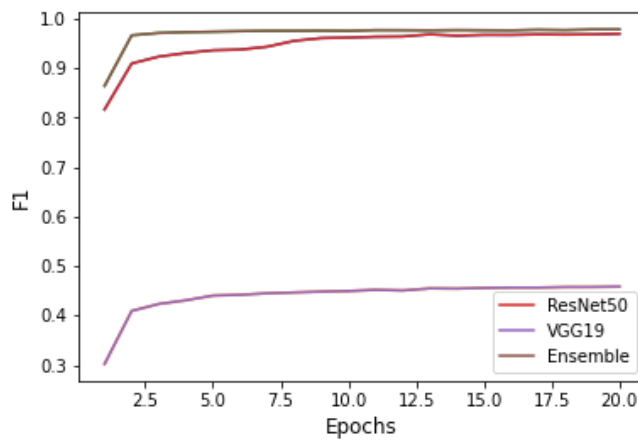


Figure 10. F1-measure

Table 2. Numerical indicators of evaluation metrics

Model	Accuracy (%)	Loss	Precision	Recall	F1
ResNet-50	94.4	0.19	0.96	0.94	0.95
VGG-19	88	0.05	0.44	0.44	0.44
Ensemble	95.7	0.02	0.98	0.97	0.97



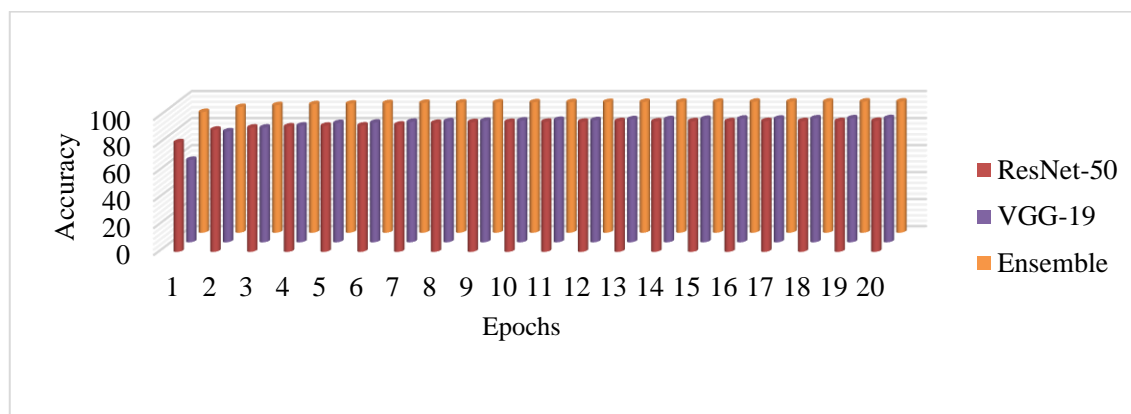


Figure 11. Accuracy comparison in different epochs

A comparison of the recognition accuracy results of the ResNet-50, VGG-19, and ensemble methods is shown in Table 3. VGG-19 has a disadvantage in the form of a tendency to retrain when the training dataset is limited. By combining several models, the ensemble can gain a more complete understanding of the basic patterns in the data, which will make it more effective in recognizing gestures in various contexts and environments.

Table 3. Accuracy comparison

Epochs	ResNet-50	VGG-19	Ensemble
1	80.97	60.84	89.26
2	90.31	81.93	92.9
3	91.87	84.8	94.27
4	92.7	86.24	94.99
5	93.24	88.11	95.44
6	93.39	88.42	95.76
7	94.04	89.08	96
8	95.28	89.42	96.18
9	95.89	89.75	96.32
10	95.99	89.95	96.43
11	96.15	90.46	96.54
12	96.2	90.24	96.62
13	96.65	90.99	96.69
14	96.42	90.95	96.76
15	96.54	91.22	96.81
16	96.52	91.37	96.86
17	96.67	91.38	96.9
18	96.61	91.59	96.94
19	96.65	91.64	96.98
20	96.81	91.75	96.99

Hereupon, the ensemble method was tested on test data. The number of examples in each class were unevenly distributed, averaging over a hundred images (Figure 12). According to the presented numbers in Table 4, gesture recognition on the test data showed a good result. However some letters of the Kazakh alphabet were not recognized due to their similarity with other gestures. For example, the algorithm predicted the letter "З" as the letter "Б". The hard sign (Б) and the soft sign (б) were completely confused with each other, which affected the recognition accuracy. The same is the case with the gestures of the letters "O" and "Ө", since they differ only in the angle of rotation. The letter "И" in sign language was also difficult to recognize, confusing with the letters "K" and "III". The reason for this was a similar look to convey the gesture, the difference is only in the number of fingers involved.

In sign language recognition in other countries, the problem of confusing gestures with each other is not observed, since their alphabet does not contain letters similar to each other. The Kazakh alphabet contains specific letters similar in sound to the letters of the Russian alphabet ("А" and "Ә", "О" and "Ө", "X" and "Һ", "H" and "Ғ", "K" and "Қ", "Y" and "Ү", also "Ы" and "И"). A similar visual transmission of these "paired" letters in the form of a gesture negatively affected the recognition process. In the future, it is planned to eliminate problem using optimization procedures of deep learning methods.

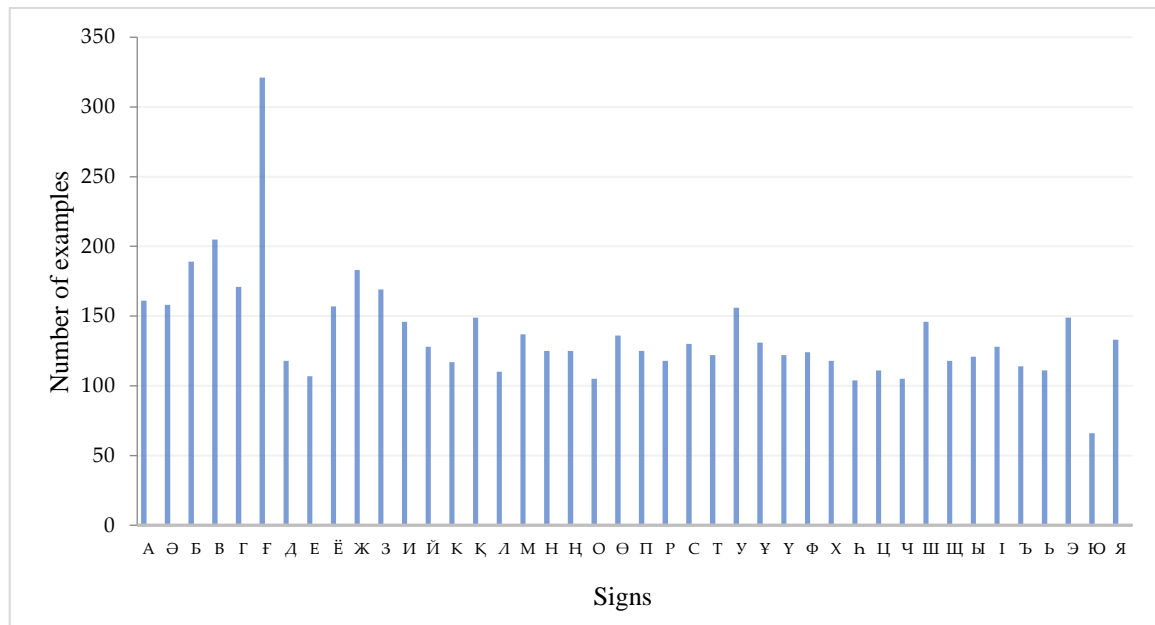


Figure 12. Distribution of classes in the test sample

Table 4. Result of correct predictions on test data

Sign	All predictions	Correct predictions	Sign	All predictions	Correct predictions
A	161	161	П	125	77
Ә	158	158	Р	118	118
Б	189	189	С	130	130
В	205	205	Т	122	104
Г	171	171	У	156	156
Ғ	321	321	Ұ	131	131
Д	118	118	Ү	122	122
Е	107	107	Ф	124	121
Ё	157	157	Х	118	98
Ж	183	183	Һ	104	0
З	169	0	Ц	111	0
И	146	146	Ч	105	105
Й	128	128	Ш	146	146
К	117	117	Щ	118	118
Қ	149	149	Ы	121	119
Л	110	110	І	128	128
М	137	137	Ї	114	0
Н	125	125	Љ	111	0
Ң	125	125	Њ	149	78
О	105	0	Ю	66	66
Ө	136	0	Я	133	11

The main feature of this approach is that the ensemble method was applied to recognize the Kazakh dactyl alphabet. Previously, works have been published on the recognition of the Kazakh gesture, mainly using classical machine learning algorithms or neural networks. In this work, two neural networks were also trained and the resulting models were combined to obtain better results. While testing the ensemble method on the test data with 5,769 examples, the model achieved an accuracy of 80.3%, which is a good result for the test data.

The limitation for this approach is the number of models to supply the ensemble itself. That is, with an increase in the number of models, the risk of overfitting increases. There is no specific requirement for the number of models to create an ensemble, but this question still depends on the condition of the task itself and the data. The disadvantage is that it takes a lot of time and computational resources to train two or more models in an ensemble. However, if we take into account the fact that we will win on the evaluation of the algorithm, it is possible to use specialized computers that have very good technical parameters and calculation speeds. It is also possible to apply methods for organizing optimal parallel computing processes to reduce the computation speed. This study can give rise to the study of the application of new optimization methods for recognizing gestures of the Kazakh alphabet. As already mentioned, classical algorithms were

used for this problem and quite good results were obtained, but this approach for the KSL recognition was used for the first time.

#### 4. CONCLUSION

In this study, a model was developed based on the ensemble method, containing the ResNet-50 and VGG-19 architectures, which will be able to classify the KSL, consisting of 42 character classes of the Kazakh alphabet. A peculiar data set was formed from 57,708 hand images for 42 signs of the KSL. The ensemble model was compared with ResNet-50 and VGG-19 on evaluation metrics such as accuracy, precision, recall, f1-measure, and loss. The identification accuracy of the ResNet-50 model was 94.4% and VGG-19 showed 88% due to the low ability to distinguish examples of a given class from other classes. The recognition accuracy of the ensemble method reached 95.7%, exceeding the performance of ResNet-50 and VGG-19. Findings from this research could pave the way for future development of sign language recognition systems for other lesser-known sign languages, helping to create a more inclusive and accessible world.

#### ACKNOWLEDGEMENTS

This work was supported by research grant BR18574144 of the Ministry of Higher Education and Science of the Republic of Kazakhstan.





#### REFERENCES

- [1] World Health Organization, "Deafness and hearing loss," *World Health Organization (WHO)*. <https://www.who.int> (accessed Oct. 01, 2023).
- [2] Suharjito, R. Anderson, F. Wiryana, M. C. Ariesta, and G. P. Kusuma, "Sign language recognition application systems for deaf-mute people: a review based on input-process-output," *Procedia Computer Science*, vol. 116, pp. 441–448, 2017, doi: 10.1016/j.procs.2017.10.028.
- [3] A. Wadhawan and P. Kumar, "Sign language recognition systems: a decade systematic literature review," *Archives of Computational Methods in Engineering*, vol. 28, no. 3, pp. 785–813, May 2021, doi: 10.1007/s11831-019-09384-2.
- [4] M. S. Altememe and N. K. El Abbadi, "A review for sign language recognition techniques," in *2021 1st Babylon International Conference on Information Technology and Science (BICITS)*, Apr. 2021, pp. 39–44, doi: 10.1109/BICITS51482.2021.9509905.
- [5] D. Ramirez-Giraldo, S. Molina-Giraldo, A. M. Alvarez-Meza, G. Daza-Santacoloma, and G. Castellanos-Dominguez, "Kernel based hand gesture recognition using kinect sensor," in *2012 XVII Symposium of Image, Signal Processing, and Artificial Vision (STSIVA)*, Sep. 2012, pp. 158–161, doi: 10.1109/STSIVA.2012.6340575.
- [6] E. Gani, A. Kika, and B. Goxhi, "A real-time vision based system for recognition of static dactyls of Albanian Alphabet," in *Recent Trends and Applications in Computer Science and Information Technology*, 2016, pp. 17–22.
- [7] S. Lang, M. Block, and R. Rojas, "Sign language recognition using kinect," in *The 10th IEEE International Conference on Automatic Face and Gesture Recognition*, 2012, pp. 394–402, doi: 10.1007/978-3-642-29347-4\_46.
- [8] L. Pigou, S. Dieleman, P.-J. Kindermans, and B. Schrauwen, "Sign language recognition using convolutional neural networks," in *Workshop at the European Conference on Computer Vision*, 2015, pp. 572–578, doi: 10.1007/978-3-319-16178-5\_40.
- [9] V. K. Verma, S. Srivastava, and N. Kumar, "A comprehensive review on automation of Indian sign language," in *2015 International Conference on Advances in Computer Engineering and Applications*, Mar. 2015, pp. 138–142, doi: 10.1109/ICACEA.2015.7164682.
- [10] A. Chaudhary, J. Raheja, K. Das, and S. Raheja, "Intelligent approaches to interact with machines using hand gesture recognition in natural way: a survey," *International Journal of Computer Science & Engineering Survey*, vol. 2, no. 1, pp. 122–133, Feb. 2011, doi: 10.5121/ijcses.2011.2109.
- [11] J. P. Sahoo, S. Ari, and D. K. Ghosh, "Hand gesture recognition using DWT and F -ratio based feature descriptor," *IET Image Processing*, vol. 12, no. 10, pp. 1780–1787, Oct. 2018, doi: 10.1049/iet-ipr.2017.1312.
- [12] E. Buchicchio, F. Santoni, A. De Angelis, A. Moschitta, and P. Carbone, "Gesture recognition of sign language alphabet with a convolutional neural network using a magnetic positioning system," *ACTA IMEKO*, vol. 10, no. 4, p. 97, Dec. 2021, doi: 10.21014/acta\_imeko.v10i4.1185.
- [13] O. Patsadu, C. Nukoolkit, and B. Watanapa, "Human gesture recognition using Kinect camera," in *2012 Ninth International Conference on Computer Science and Software Engineering (JCSSE)*, May 2012, pp. 28–32, doi: 10.1109/JCSSE.2012.6261920.
- [14] L. K. Phadtare, R. S. Kushalnagar, and N. D. Cahill, "Detecting hand-palm orientation and hand shapes for sign language gesture recognition using 3D images," in *2012 Western New York Image Processing Workshop*, Nov. 2012, pp. 29–32, doi: 10.1109/WNYIPW.2012.6466652.
- [15] I. A. Adeyanju, O. O. Bello, and M. A. Adegboye, "Machine learning methods for sign language recognition: a critical review and analysis," *Intelligent Systems with Applications*, vol. 12, p. 200056, Nov. 2021, doi: 10.1016/j.iswa.2021.200056.
- [16] K. Dabre and S. Dholay, "Machine learning model for sign language interpretation using webcam images," in *2014 International Conference on Circuits, Systems, Communication and Information Technology Applications (CSCITA)*, Apr. 2014, pp. 317–321, doi: 10.1109/CSCITA.2014.6839279.
- [17] B. Gupta, P. Shukla, and A. Mittal, "K-nearest correlated neighbor classification for Indian sign language gesture recognition using feature fusion," in *2016 International Conference on Computer Communication and Informatics (ICCCI)*, Jan. 2016, pp. 1–5, doi: 10.1109/ICCCI.2016.7479951.
- [18] S. Nagarajan and T. S. Subashini, "Static hand gesture recognition for sign language alphabets using edge oriented histogram and multi class SVM," *International Journal of Computer Applications*, vol. 82, no. 4, pp. 28–35, Nov. 2013, doi: 10.5120/14106-2145.
- [19] 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.
- [20] B. Kang, S. Tripathi, and T. Q. Nguyen, “Real-time sign language fingerspelling recognition using convolutional neural networks from depth map,” in *2015 3rd IAPR Asian Conference on Pattern Recognition (ACPR)*, Nov. 2015, pp. 136–140, doi: 10.1109/ACPR.2015.7486481.
- [21] W. Tao, M. C. Leu, and Z. Yin, “American sign language alphabet recognition using convolutional neural networks with multiview augmentation and inference fusion,” *Engineering Applications of Artificial Intelligence*, vol. 76, pp. 202–213, Nov. 2018, doi: 10.1016/j.engappai.2018.09.006.
- [22] J. Gangrade and J. Bharti, “Vision-based hand gesture recognition for Indian sign language using convolution neural network,” *IETE Journal of Research*, vol. 69, no. 2, pp. 723–732, Feb. 2023, doi: 10.1080/03772063.2020.1838342.
- [23] G. A. Rao, K. Syamala, P. V. V. Kishore, and A. S. C. S. Sastry, “Deep convolutional neural networks for sign language recognition,” in *2018 Conference on Signal Processing And Communication Engineering Systems (SPACES)*, Jan. 2018, pp. 194–197, doi: 10.1109/SPACES.2018.8316344.
- [24] P. Nakjai and T. Katanyukul, “Hand sign recognition for Thai finger spelling: an application of convolution neural network,” *Journal of Signal Processing Systems*, vol. 91, no. 2, pp. 131–146, Feb. 2019, doi: 10.1007/s11265-018-1375-6.
- [25] R. Sharma, R. Khapra, and N. Dahiya, “Sign language gesture recognition,” *International Journal of Recent Research Aspects*, vol. 7, no. 2, pp. 14–19, 2020.
- [26] A. Beibut, “Development of automatic speech recognition for Kazakh language using transfer learning,” *International Journal of Advanced Trends in Computer Science and Engineering*, vol. 9, no. 4, pp. 5880–5886, Aug. 2020, doi: 10.30534/ijatcse/2020/249942020.
- [27] O. Mamyrbayev *et al.*, “Continuous speech recognition of Kazakh language,” *ITM Web of Conferences*, vol. 24, p. 01012, Feb. 2019, doi: 10.1051/itmconf/20192401012.
- [28] S. A. Kudubayeva, D. A. Ryumin, and M. U. Kalzhanov, “The method of basis vectors for recognition sign language by using sensor KINECT,” *Journal of Mathematics, Mechanics and Computer Science*, vol. 3, no. 91, pp. 86–96, 2018.
- [29] L. Ali, F. Alnajjar, H. Al Jassmi, M. Gocho, W. Khan, and M. A. Serhani, “Performance evaluation of deep CNN-based crack detection and localization techniques for concrete structures,” *Sensors*, vol. 21, no. 5, p. 1688, Mar. 2021, doi: 10.3390/s21051688.
- [30] M. Bansal, M. Kumar, M. Sachdeva, and A. Mittal, “Transfer learning for image classification using VGG19: caltech-101 image data set,” *Journal of Ambient Intelligence and Humanized Computing*, vol. 14, no. 4, pp. 3609–3620, Apr. 2023, doi: 10.1007/s12652-021-03488-z.
- [31] A. Choromanska, M. Henaff, M. Mathieu, G. Ben Arous, and Y. LeCun, “The loss surfaces of multilayer networks,” in *Artificial intelligence and statistics*, 2015, pp. 192–204.
- [32] C. Ju, A. Bibaut, and M. van der Laan, “The relative performance of ensemble methods with deep convolutional neural networks for image classification,” *Journal of Applied Statistics*, vol. 45, no. 15, pp. 2800–2818, Nov. 2018, doi: 10.1080/02664763.2018.1441383.
- [33] G. Huang, Y. Li, G. Pleiss, Z. Liu, J. E. Hopcroft, and K. Q. Weinberger, “Snapshot ensembles: train 1, get m for free,” *arXiv*, 2017, doi: 10.48550/arXiv.1704.00109.

## BIOGRAPHIES OF AUTHORS






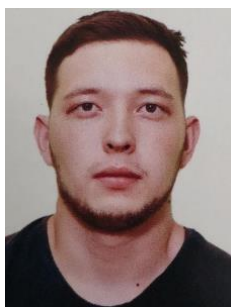
**Yedilkhan Amirgaliyev**     Academician of the National Engineering Academy of the Republic of Kazakhstan, corresponding member of the National Academy of Sciences of the Republic of Kazakhstan, Doctor of Technical Sciences, Professor. Currently, he is the Head of the Laboratory of Artificial Intelligence and Robotics of the Institute of Information and Computational Technologies of CS MSHE RK. His research interests include mathematical methods of pattern recognition, methods of processing speech and graphic signals, image processing, identification and recognition of dynamic objects, and artificial intelligence in robotics. He can be contacted at email: amir\_ed@mail.ru.






**Aisulyu Ataniyazova**     received the B.Sc. and M.Sc. degrees in Computer Engineering from Al-Farabi Kazakh National University. Currently, she is a Ph.D. student of Al-Farabi Kazakh National University. She works as a software engineer of Laboratory of Artificial Intelligence and Robotics of the Institute of Information and Computational Technologies of CS MSHE RK. Her research interests are primarily in the area of machine learning, computer vision, robotics, and remote sensing tasks. She can be contacted at email: aisulu.ataniyazova@gmail.com.






**Zholdas Buribayev**    received the Ph.D. degree from Al-Farabi Kazakh National University, in 2022. Currently, he is the leader of the scientific project No. AP19579370 «Development of an autonomous mobile robot and an object recognition system for patrolling the area based on machine learning algorithms. His research interests include a image processing, artificial intelligence, data science, and bioinformatics. He can be contacted at email: zholdas.buribayev@kaznu.edu.kz.






**Mukhtar Zhassuzak**    Ph.D. student of the Al-Farabi Kazakh National University. He is a junior researcher and software engineer of the Laboratory of Artificial Intelligence and Robotics of the Institute of Information and Computational Technologies of CS MSHE RK. His research interests are mainly in the fields of artificial intelligence, computer vision, digital image processing, heuristic algorithms, and autonomous movement of robots. He can be contacted at email: zhassuzak.mukhtar@gmail.com.



**Baydaulet Urmashev**    received the B.Sc. and M.Sc. degrees in Applied Mathematics from Al-Farabi Kazakh National University. In 2001, he defended his doctoral dissertation on the specialty (Computational Mathematics). Currently he is a Professor of Al-Farabi Kazakh National University. His main research areas include pattern recognition, parallel programming, and computational methods for resource-intensive tasks. He can be contacted at email: baydaulet.urmashev@gmail.com.



**Lyailya Cherikbayeva**    received the Ph.D. degree from Al-Farabi Kazakh National University, in 2020. She works as a senior researcher of Laboratory of Artificial Intelligence and Robotics of the Institute of Information and Computational Technologies of CS MSHE RK. Currently, she is engaged in the research and development of a combination of cluster analysis algorithms and recognition methods with semi-training based on the core function. She can be contacted at email: cherikbayeva.lyailya@gmail.com.