

For people with disabilities, sign language is the most important means of communication. Therefore, more and more authors of various papers and scientists around the world are proposing solutions to use intelligent hand gesture recognition systems. Such a system is aimed not only for those who wish to understand a sign language, but also speak using gesture recognition software. In this paper, a new benchmark dataset for Kazakh fingerspelling, able to train deep neural networks, is introduced. The dataset contains more than 10122 gesture samples for 42 alphabets. The alphabet has its own peculiarities as some characters are shown in motion, which may influence sign recognition.

Research and analysis of convolutional neural networks, comparison, testing, results and analysis of LeNet, AlexNet, ResNet and EffectiveNet – EfficientNetB7 methods are described in the paper. EffectiveNet architecture is state-of-the-art (SOTA) and is supposed to be a new one compared to other architectures under consideration. On this dataset, we showed that the LeNet and EffectiveNet networks outperform other competing algorithms. Moreover, EffectiveNet can achieve state-of-the-art performance on nother hand gesture datasets.

The architecture and operation principle of these algorithms reflect the effectiveness of their application in sign language recognition. The evaluation of the CNN model score is conducted by using the accuracy and penalty matrix. During training epochs, LeNet and EffectiveNet showed better results: accuracy and loss function had similar and close trends. The results of EffectiveNet were explained by the tools of the SHapley Additive exPlanations (SHAP) framework. SHAP explored the model to detect complex relationships between features in the images. Focusing on the SHAP tool may help to further improve the accuracy of the model

Keywords: hand gesture recognition, sign language recognition, convolutional neural network (CNN), deep learning

UDC 004.93

DOI: 10.15587/1729-4061.2021.241535

A COMPARISON OF CONVOLUTIONAL NEURAL NETWORKS FOR KAZAKH SIGN LANGUAGE RECOGNITION

Chingiz Kenshimov

PhD, Leading Researcher*

Samat Mukhanov

Corresponding author

Doctoral Student, Vice-Dean of Faculty, Senior Lecturer

Faculty of Computer Technologies and Cybersecurity

Department of Computer Engineering

and Information Security

Laboratory of Artificial Intelligence and Robotics

International Information Technology University

Manas 34/1, Almaty, Republic of Kazakhstan, 050010

Timur Merembayev

Postgraduate Student, Software-Engineer*

Didar Yedilkhan

PhD, Software-Engineer*

*Laboratory of Artificial Intelligence and Robotics

Institute of Information and Computational Technologies

Shevchenko str., 28, Almaty,

Republic of Kazakhstan, 050010

Received date 12.08.2021

Accepted date 11.10.2021

Published date 29.10.2021

How to Cite: Kenshimov, C., Mukhanov, S., Merembayev, T., Yedilkhan, D. (2021). A comparison of convolutional neural networks for Kazakh sign language recognition. *Eastern-European Journal of Enterprise Technologies*, 5 (2 (113)), 44–54.

doi: <https://doi.org/10.15587/1729-4061.2021.241535>

1. Introduction

The topic of hand gesture recognition has a special place and active application in computer vision science. Machine learning algorithms and artificial neural networks are improving at optimizing the operation of pattern recognition systems and the analysis of digital text images. Machine learning libraries are widely used in the research and analysis of the gesture recognition algorithm – a sign language in this case.

A machine learning-based gesture recognition system allows people who don't know the alphabet of a sign language to understand its meaning. Vision-based hand gesture recognition involves several tasks that need to be solved in order to get a good gesture recognition result, for example, hand zooming, reducing noise, changing lighting and view

angle, differentiating between similar gestures and working with a complex background. Recognizing hand gestures in images can be very challenging due to the varying conditions of the image captured. The task of recognizing a sign language is complicated owing to the dynamism of gestures, where one symbol can be displayed in dynamics. In recent years, many works have been devoted to gesture recognition, but we have not found any works related to Kazakh Sign Language gesture recognition using modern deep learning algorithms.

In modern conditions, gesture recognition of sign language will allow a person to communicate and interact with machines naturally, without any mechanical intermediaries. Gesture recognition will practically allow computers to be more accessible to people with disabilities, as well as to use interaction in virtual or augmented reality.

2. Literature review and problem statement

AR/VR technologies have prompted the development of gesture recognition. Gestures are used in various applied tasks, for example, virtual reality control, video games, etc. There are several studies carried out by research organizations and companies to solve these problems.

GoogleAI offers an approach of accurate tracking of hands and fingers using machine learning on the MediaPipe cross-platform framework, which in turn builds data processing pipelines (video, audio and time series) [1]. The authors use such models as a palm detector (blazePalm), a model for determining key points on the hand and an algorithm for gesture recognition. All of these models form a single base for the abovementioned framework. Each of these models is unique and defines the key points of identifying special elements in gesture recognition.

The authors of the papers [2, 3] propose a solution to improve the recognition of surgical hand gestures using a capsule network for a contactless interface in the operating room. In [2], the authors investigated the issue of contactless recognition of sign language as a way to avoid the risk of infection and, in general, the improvement of the sign language recognition system. It is shown that the methods of interfaces using CapsNet can achieve the greatest efficiency in comparison with other methods. The method is implemented using a capsule network (CapsNet) and Leap Motion. This method involves the extraction and preprocessing of infrared images at 60 frames per second via Leap Motion, as well as training various types of networks and assessing gesture recognition in the operating room. Application of the CapsNet method demonstrates a classification accuracy of 86.46 %, which is quite high compared to 73.67 % of CNN (convolutional neural network). This means that the capsular network approach is more efficient than the convolutional neural network one. In their research [3], the authors dealt with the issue of the American sign language recognition system. Various models of hand gesture recognition were used as a practical application of theoretical data.

A deep convolutional neural network approach for static hand gesture recognition [4] of convolutional neural networks shows how a training network in a 3D model for gesture recognition can be built. They propose a hand gesture recognition methodology, which is a core component of a sign language vocabulary, based on an efficient deep transformation neural network architecture. This method has been tested on several publicly available datasets, i.e. the NUS hand position dataset and the American fingerprint dataset A, which showed a very high recognition accuracy equal to 87 %.

Leap Motion is a database of hand gesture recognition, which consists of a set of images in the near-infrared range obtained by the Leap Motion sensor [5]. The content of the database consists of 10 different hand gestures performed by 10 different subjects (5 men and 5 women). The database is structured in different folders, in addition to gesture recognition, it also performs the task of recognizing sign languages. The task has a social focus and will help people with hearing problems to communicate with minimal restrictions. Modern technologies can solve this problem.

The authors of the paper [6] demonstrated a new system for communicating with people with visual impairments or radio communication. An experimental method was used

as a study. The authors note that the algorithm device can extract icons from video sequences on a minimally cluttered and dynamic background using the CapsNet technologies. In their work, they used HSV Segmentation and Finger-tip detection algorithms. The authors also classified using Support Vector Machines in sign language recognition. As a result, they achieved 93 % accuracy Static Gesture recognition using HSV Segmentation and Finger-tip detection methods. And finally classified popular methods making comparison and demonstrated results in the table.

The authors of another research paper [7] touched upon the creation of a full-fledged recognition system for Turkish Sign Language. The paper explored the practical side of this system. Based on the recognition of Turkish sign language, hand movement recognition was used. The authors divided this recognition into two components: one-handed and two-handed signs. Repetitive dominant hand gestures that differ from the original version of a given gesture have also been studied. The authors of the paper showed that the Turkish Sign Language recognition system is also not perfect, as the Turkish Sign Language has not been fully studied yet. Their research has shown that it will take a long time of computation and experimentation to create an ideal sign language recognition system. Humanity is still at an early stage in the creation of a sign language recognition system. The authors have only experimented with hand gesture recognition. The process of recognizing sign language through facial expressions is not yet fully understood and has its own specific difficulties, which the authors [7, 8] hope to overcome in the near future through trial and error. The authors showed in research that they used Hidden Markov Models in the sign language recognizer. Turkish Sign Language uses more than 2000 gestures manually. They also researched feature extraction and motion extraction. As experimental results, they tested one-handed, two-handed and final test results. Finally, they concluded that the learning phase, which takes too much time, makes it difficult to design the signer independent system. The authors guess that they should concentrate on continuous sign language recognition after obtaining a large vocabulary size.

The authors of the paper [8] suggest a methodology for recognizing Turkish Sign Language using Kinect, which takes into account the skeletal feature of each pose. The classification of postures is not yet fully understood, but the authors believe that in the near future they will achieve the expected results in experiments. Each idea or hypothesis in the paper has been researched by a variety of mathematical calculations and experimental studies. The work demonstrated many mathematical formulas and models, especially Finite State Automata for Gestures Classification. They highlighted the Performance of Posture Labeling, and Performance of Gesture Classification and high results of the accuracy are more than 97 % in the posture labeling scheme, it lowers to 93 % on average in gesture classification. The research topic was “CNN” and “kNN” machine learning algorithms in Turkish sign language.

The authors of the paper [9] studied the recognition of Kazakh Sign Language. The Kinect sensor used for gesture recognition, the coordinates of the hand skeleton and key characteristics stored in XML files and calculation provided by MatLab. L. S. Dimskis' sign notation selected for the Kazakh sign language, the features of the representation of the Kazakh sign language using L. S. Dimskis' sign notation

in the course of compiling a dictionary of frequently used gestures were described in the paper. The authors used the support vector machine, hidden Markov models, fuzzy models, artificial neuron model, image difference method. They also developed a database for the Kazakh sign language, consisting of the dactyl alphabet of 42 gestures, which is the initial step in creating a system for automatic recognition of individual hand gestures.

The authors of the paper [10] investigated various algorithms and methods of image recognition. This work showed how algorithms such as the active contour method, canny boundary detector, local processing, path tracking, and graph analysis and convexity defects were used and tested. All the methods differ in their own way. As a result, it was shown that using the OpenCV technology, the convexity defects algorithm was applied, since this pattern recognition method is closest to recognizing the hand sign language (gesture recognition). The authors focused on analyzing and using canny boundary detectors, path tracking, clustering, local processing, canny border detector, graph analysis and other algorithms for pattern recognition. In the experimental part, they demonstrated the contour analysis method to recognize pictures. The effectiveness of this algorithm is visualized in the picture.

This is a base for all areas with image visualization. Starting with the basic algorithms for recognizing the contours of an object (contour analysis) and detecting angles (Harris corner detection), including mathematical models that are used as the basis of the equations and formulas for which the function and algorithm of the program are built. It was researched how, using programming technologies in the C++ language and the OpenCV library with computer vision tools, to recognize data that are uploaded in an image format. Contour recognition is still under improvement, however, it can identify the majority of object contours in digital pictures.

3. The aim and objectives of the study

The aim of the study is to compare the SOTA Efficient-NetB7 with classical deep learning architectures on our dataset of Kazakh Sign language and provide a detailed analysis of the explanation of the CNN models. This will make it possible to develop new methods of analysis and synthesis of visual communication information with possible application to solving artificial intelligence problems, building new man-machine interfaces, new devices, etc.

To achieve this aim, the following objectives are accomplished:

- collecting a sufficient number of images for the Kazakh sign language, which are necessary for training and validation of CNN architectures;
- performing a comparative analysis of CNN architectures for assessing accuracy on the collected dataset of Kazakh Sign Language;
- analyzing the obtained results of a separate model.

4. Materials and methods

The first version of the Kazakh alphabet consists of 42 letters. The development of a database for the Kazakh

Sign Language, consisting of a dactyl alphabet of 42 gestures, is the initial step in creating a system for automatic recognition of individual hand gestures. The Dactyl Alphabet for the first Kazakh sign language is shown in Fig. 1.

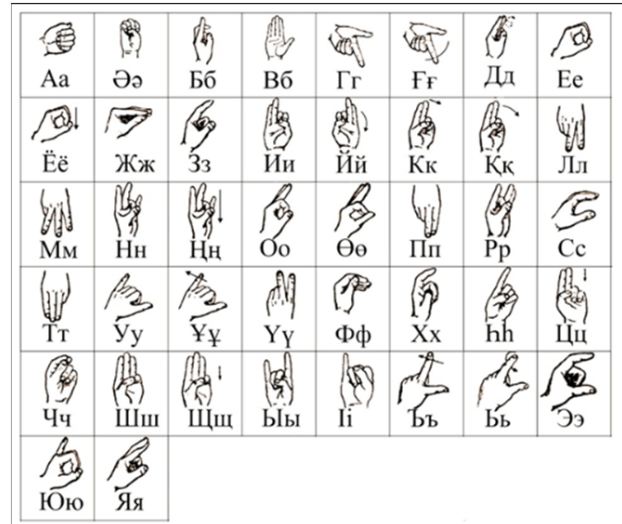


Fig. 1. Kazakh Dactyl Alphabet [9]

We considered the hand gestures of the Kazakh alphabet, the dataset was collected by a web camera in room conditions. It consists of 2100 images for 42 classes of gestures, 50 images for each class. Fig. 2 shows the sample of 8 classes in 1 band.

CNN architecture. Advances in artificial intelligence and deep learning have facilitated rapid evolution in computer vision and image analysis. This became possible thanks to the emergence and development of convolutional neural networks (CNN).

A convolutional neural network is a deep learning algorithm that can recognize and classify features of images for computer vision.

LeNet. The LeNet architecture is one of the original convolutional neural network algorithms that was developed and presented in the late 90s in the research paper called “Gradient Learning Used for Document Recognition” in the field of deep learning [11].

This architecture is a seven-tier artificial convolutional neural network. It was designed for low-resolution black and white object recognition. Input data consisted of 32×32 images of 32 bits in size, which were then divided into six channels of 28×28 pixels, which were afterwards reduced to an average merging of 14×14.

AlexNet. This method is a kind of convolutional neural network. AlexNet influenced greatly the development of machine learning, in particular computer vision. The AlexNet architecture is similar to that of Yann LeCum’s LeNet. However, AlexNet has more filters per layer and nested convolutional layers. The network includes convolutions, maximum pooling, dropout, data augmentation, ReLU activation functions, and stochastic gradient descent [12].

The AlexNet architecture uses the Relu activation function instead of the arctangent to add nonlinearity to the model. Due to this, with the same accuracy of the method, the speed becomes 6 times higher.

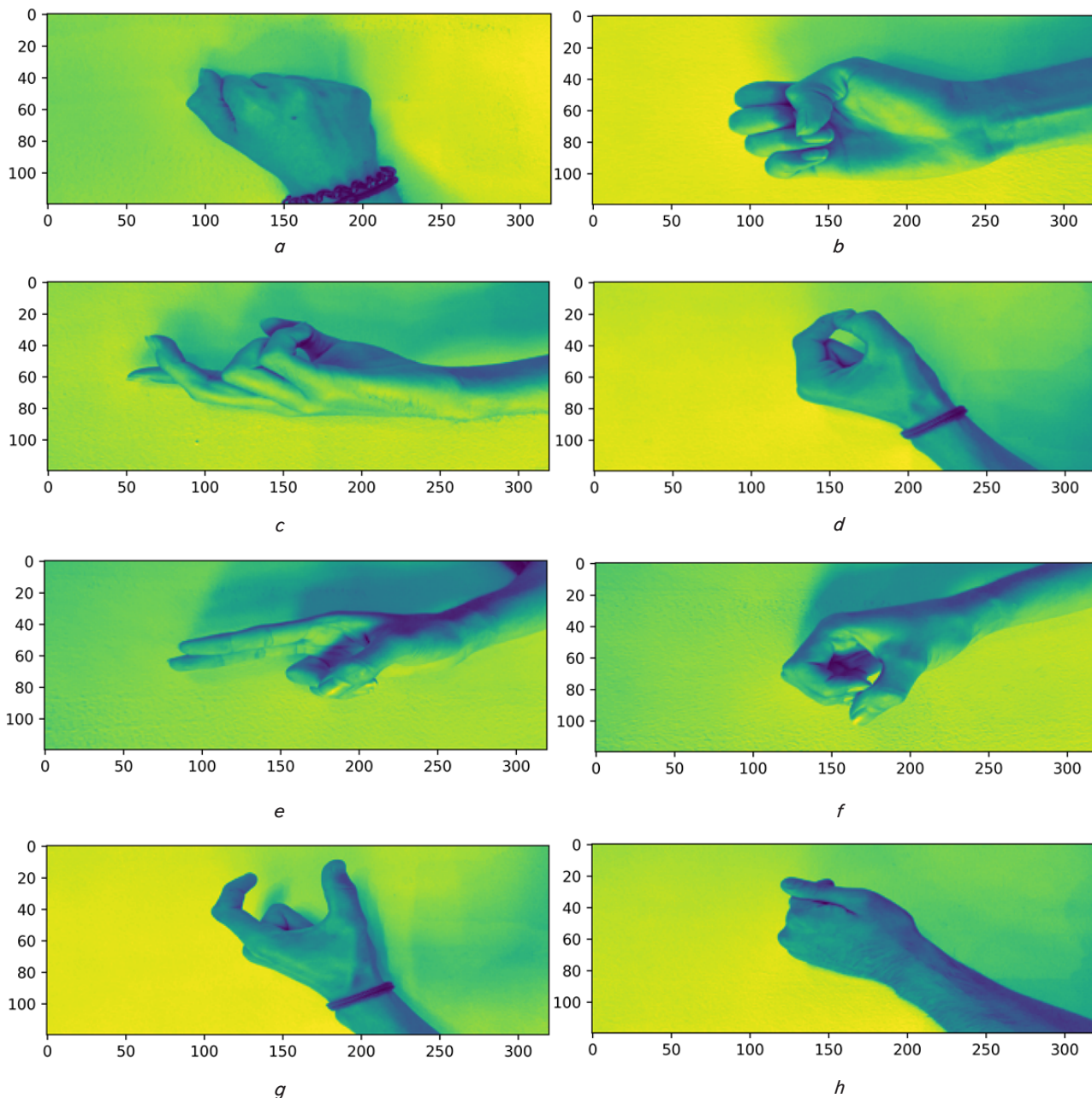


Fig. 2. Sample images from the Kazakh Sign Language dataset: a – “А” – Kazakh Sign from the dataset; b – “Ә” – Kazakh Sign from the dataset; c – “Б” – Kazakh Sign from the dataset; d – “Ё” – Kazakh Sign from the dataset; e – “Д” – Kazakh Sign from the dataset; f – “Е” – Kazakh Sign from the dataset; g – “Ә” – Kazakh Sign from the dataset; h – “Ф” – Kazakh Sign from the dataset

ResNet50. ResNet comes from the generic name for the Residual Network [13]. Deep networks extract low, medium, and high-level features in an end-to-end multilayer mode, thus increasing the number of stacked layers can enrich feature “levels”. At that moment, if the deeper network begins to collapse, a problem arises. It means that with increasing network depth, the accuracy first increases and then quickly deteriorates. The decrease in training accuracy shows that not all networks are easy to optimize. To overcome this problem, Microsoft introduced a deep “residual” learning structure. Instead of relying on the idea that every several stacked layers directly correspond to the desired base view, they allow those layers to correspond to the “residual” one. The $F(x)+x$ formulation can be implemented using neural networks with quick access connections.

EffectiveNet – EfficientNetB7. This architecture presents a machine learning method in a deep neural network. EffectiveNet is a composite scaling method, which consistently improves model accuracy and efficiency for scaling existing models such as MobileNet (+1.4 % image fidelity) and ResNet (+0.7 %) over traditional scaling methods [14].

Comparing EfficientNets to other existing CNNs on ImageNet, it was found that generally, EfficientNet models provide higher accuracy and better efficiency than existing CNNs, significantly reducing the parameter size and quantity of FLOPS. For example, in high fidelity mode, EfficientNet-B7 achieves 84.4 % top-1/97.1 % top-5 accuracy on ImageNet, while being 8.4 times smaller and 6.1 times faster in CPU output.

Two statistical quality indicators evaluate the prediction performance of the algorithms: accuracy and penalty metrics. This allows Kazakh Sign Language predictions to be scored by a degree of “wrongness”.

Accuracy is an important and commonly used metric but does not capture the variability of performance in an unbalanced dataset. In this case, we prepared a balanced dataset, each class containing 50 images.

$$\text{Accuracy} = \frac{(TP + TN)}{(TP + TN + FP + FN)}, \tag{1}$$

where *TP* is true positive predictions, *TN* is true negative, *FP* is false positive and *FN* is false negative as shown in Fig. 3.

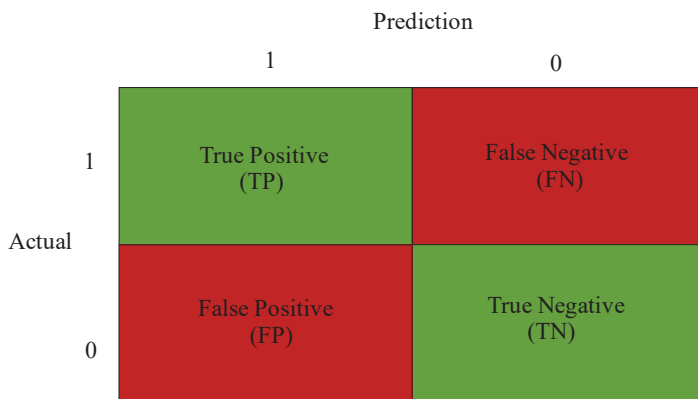


Fig. 3. Confusion matrix of TP, TN, FP and FN

Instead of penalizing each wrong prediction of the Kazakh Sign Language, we additionally used a custom penalty matrix. It was derived from the averaged input of a representative sample of Kazakh Sign Language. This

allows for unreasonable predictions to be scored by a degree of “wrongness”. This metric proved itself in the classification of geoscience problems [15]. The scoring matrix of penalty is defined as follows:

$$S = -\frac{1}{N} \sum_{i=0}^N A_{\hat{y}_i y_i}, \tag{2}$$

where *N* is the number of classes, \hat{y}_i is the true hand gesture, and y_i is the predicted hand gesture.

The value of the matrix *A* in row *i* and column *j* is the penalty given by guessing hand gesture class *i* when the correct label is hand gesture class *j*. Note that the diagonal consists of zeros, no penalty is given for correct predictions.

The penalty matrix is shown in Fig. 4. The penalty matrix is taken into account in a model selection or in the optimization routine itself.

To answer the question: why do CNN models determine accuracy for multi-label classification of Kazakh sign language? we use the SHAP package and explain this row of data [16]. SHAP is a good tool for the explanation of various supervised learning models and calculating an importance value for each input variable for a specific prediction. Also, SHAP can find new classes of additive feature importance measures.

Computations are performed on a desktop machine (3.2 GHz Intel Core i7 8700 processor, NVIDIA RTX 2080 8Gb) with 32 GB RAM. Tuning hyperparameters and cross-validation operations are time-consuming, therefore they are computed in parallel mode using eight cores.

Firstly, we used hardware for our first task. It was necessary to use high technical characteristics to test our data (datasets). It was the best solution to save time and get the desired results quickly.

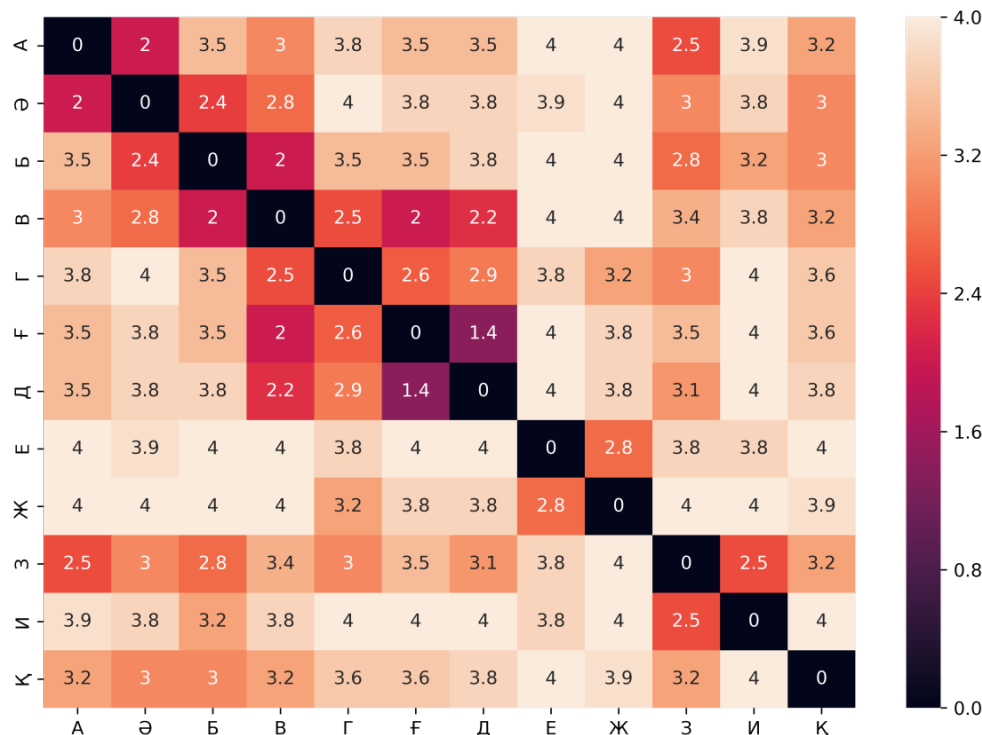


Fig. 4. Example of penalty matrix values

5. Results of comparative analysis of CNN architectures for the collected dataset

5.1. Dataset

The next task was to collect the necessary data for training. Since we used several machine learning models to train the data and we needed to collect many images.

The dataset consists of 2,100 images for 42 classes of gestures, 50 images for each class. Fig. 2 shows an example of collected data for 8 gestures. For training and validation of CNN models, this volume of images is sufficient; when training models using LeNet and EffectiveNet, we did not notice overtraining. This number of images will allow us to avoid the intermediate stage of work on dataset augmentation.

Supervised machine-learning models generally train on a subset of an entire dataset, and then perform it on a test (different subset). The dataset is randomly split into train, validation and test datasets where 80 % of the data was used to train a model (1,680 images), 10 % was used for the validation of the model (210 images) and 10 % was used to check the accuracy of the model (210 images).

5.2. Comparative analysis of CNN architectures

After training these machine learning models, we made a benchmarking analysis. The results of these models were

tested on a test sample to assess the accuracy of the models and the following metrics were used: Accuracy, Loss and Penalty.

For comparison, we trained LeNet, AlexNet, ResNet50 and EffectiveNet. The batch normalization, root mean square propagation (rmsprop) optimizer, loss function, categorical cross entropy are applied for all the networks. We summarized our experimental results in Table 1 for comparison. The training process chose 20 epochs to converge to an optimal CNN model to recognize the hand sign. Performance comparison in terms of accuracy, loss and penalty matrix is shown in Table 1.

Table 1

Performance comparison in terms of accuracy, loss and penalty matrix

Model	Total params (M)	Accuracy	Loss	Penalty
LeNet	1.07	0.990	0.0453	3.017
AlexNet	35.5	0.990	0.1343	3.021
ResNet50	23.7	0.642	1.8544	3.130
EffectiveNetB7	64.2	0.967	0.1355	3.017

The results of training CNN models are given in Fig. 5–8.

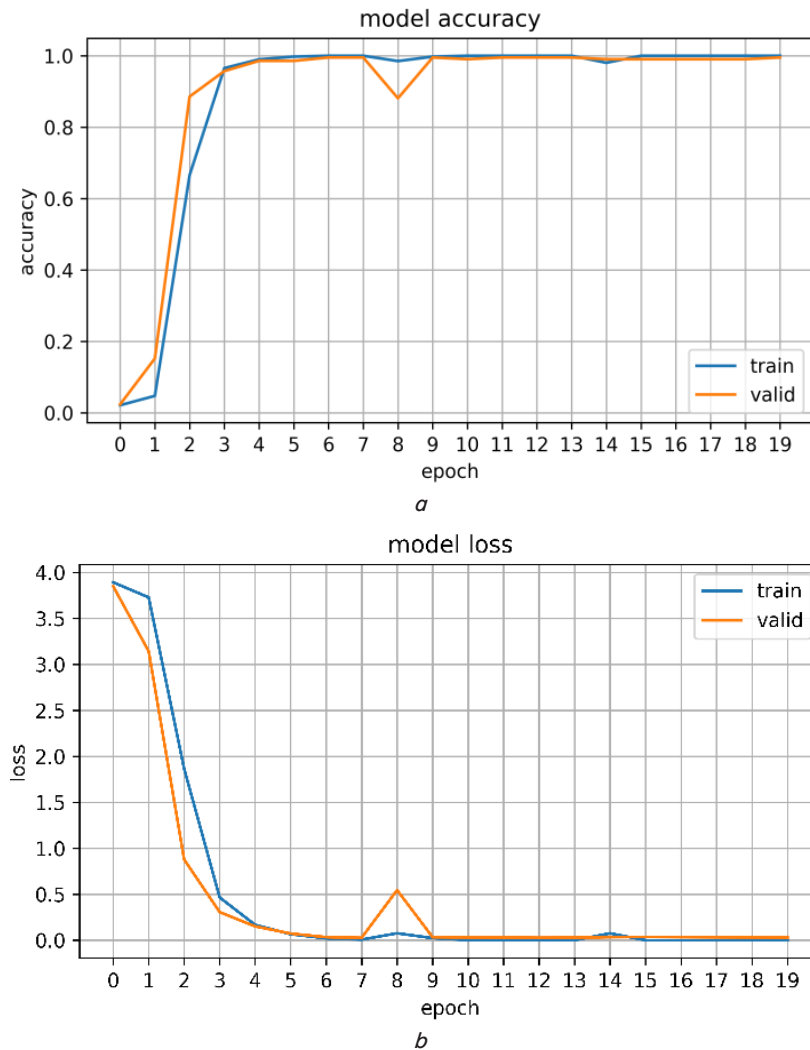


Fig. 5. Accuracy and loss for the LeNet model: *a* – LeNet model accuracy; *b* – LeNet model loss

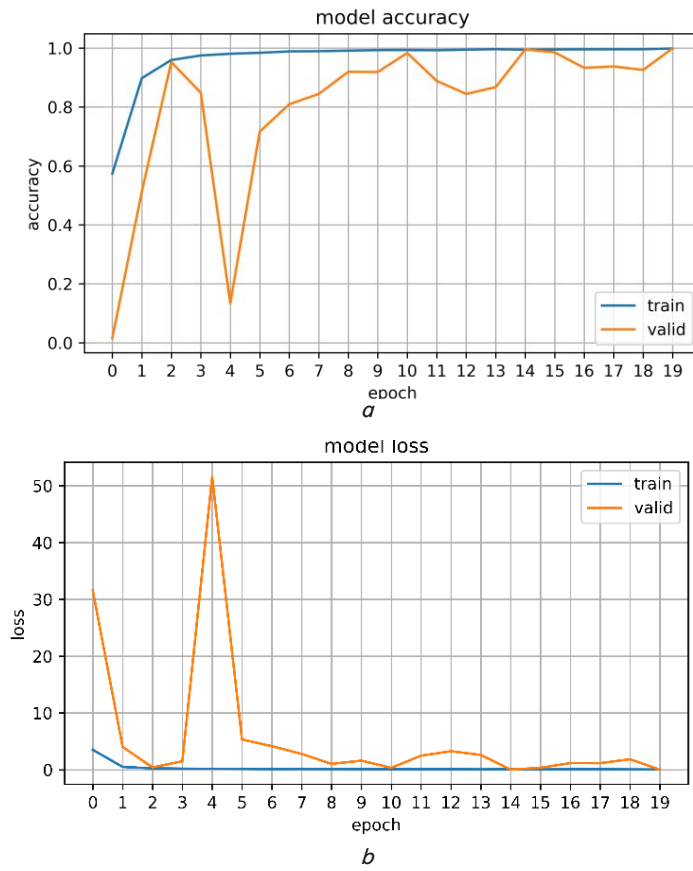


Fig. 6. Accuracy and loss for the AlexNet model: *a* – AlexNet model accuracy; *b* – AlexNet model loss

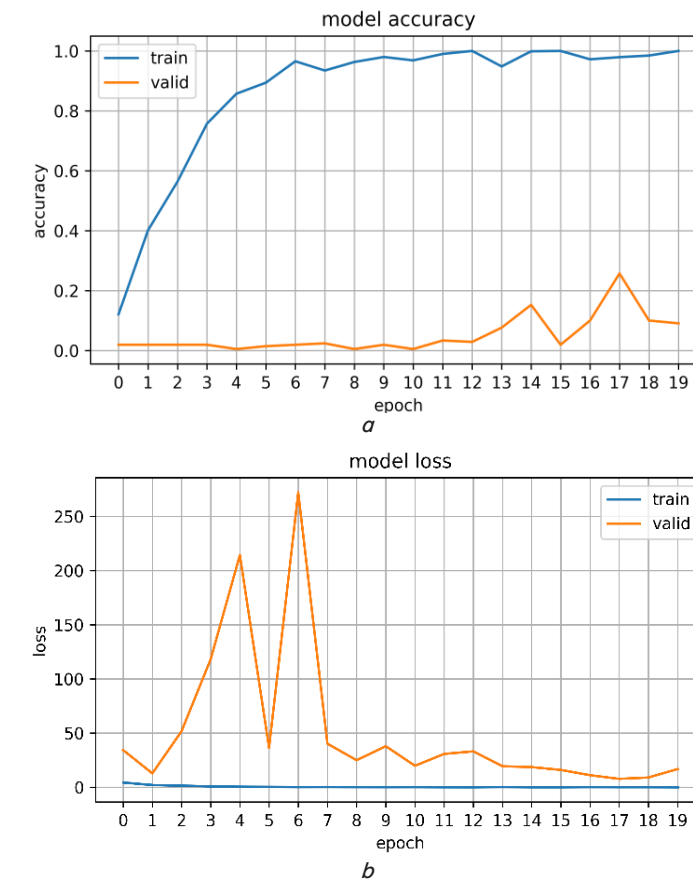


Fig. 7. Accuracy and loss for the ResNet50 model: *a* – ResNet50model accuracy; *b* – ResNet50model loss

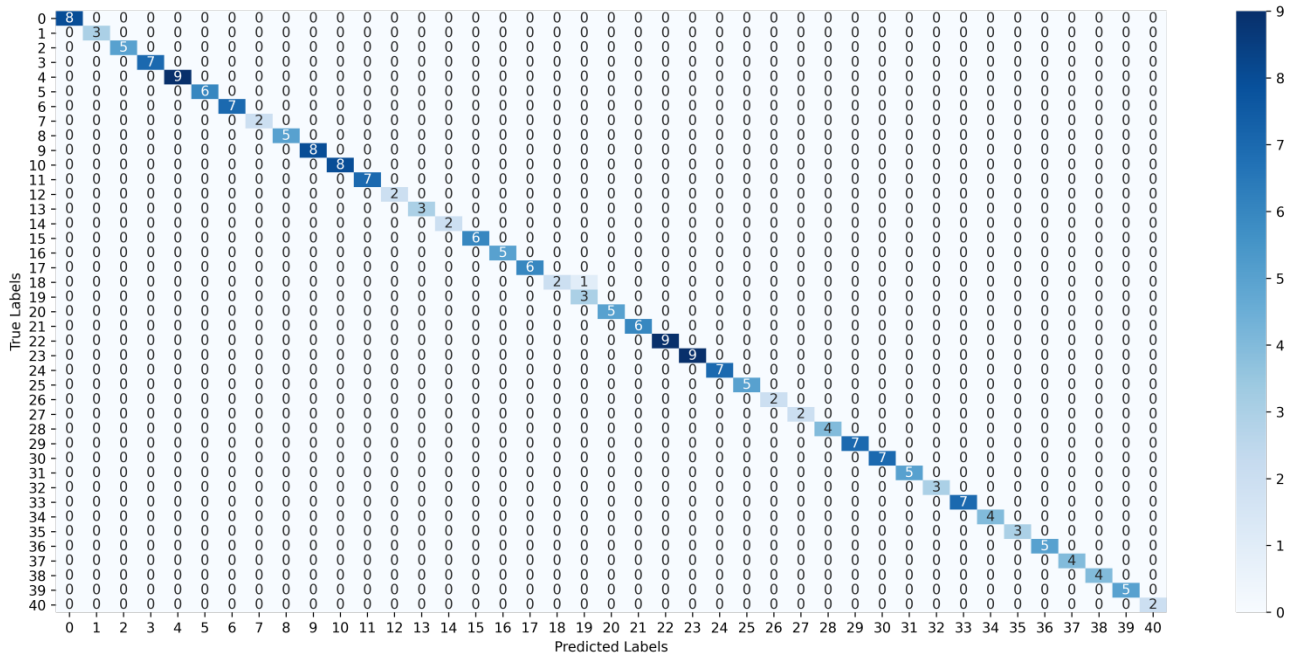


Fig. 10. Confusion matrix for the EffectiveNet model

5.3. Interpreting EffectiveNet Model Results Using SHAP

For this task, we interpreted the results obtained from the EffectiveNet model using SHAP. This method provides a visual explanation of which specific features in the image were used to improve the accuracy of the model.

Fig. 11 shows the explanations for class on ten predictions (images of sign language “C”, “3”, “Э”, “У”, “F”, “Ө”, “Ж”, “Г”, “Л”, “Д”). The figure represents ten classes selected randomly from 42 classes.

In Fig. 11, the left row is the input images, the right images are transparent grayscale backings behind each of the explanations. The explanations are ordered for classes 0–4 from left to right along the rows, starting with the original image. The red pixels increase the model’s prediction. The blue pixels decrease the model’s prediction. The sum of the SHAP values equals the difference between the expected model prediction and the current model prediction.



Fig. 11. Explanation for 10 classes by SHAP

6. Discussion of the results of Kazakh sign language recognition by CNN

According to the experimental results presented in Table 1, the neural networks get comparable and similar results. First, due to the sufficient amount of diverse images available for the networks to train on. Second, due to the effectiveness of CNN for this type of task. Now the CNN architecture is huge and contains millions of parameters to solve difficult image recognition tasks.

Fig. 5–9 illustrate the results of training LeNet, AlexNet, ResNet50 and EffectiveNet for 20 epochs with a comparison of accuracy and loss. LeNet and EffectiveNet showed better results: training speed, accuracy and loss had similar and close trends during epochs. LeNet is more “light” (fewer parameters), EffectiveNet is “heavy” but it has a smoother accuracy trend. AlexNet and ResNet50 were overfitting, as accuracy and loss trends for the validation dataset showed small values.

Based on the confusion matrix, it was noted that CNN provided some errors for complicated signs like “Ө” and “Б”, due to the specifics of showing these signs, as they are shown with motion up. This motion is a complicated task for image recognition, but it can be solved by video capture or series of language sign images. CNN solved the issue with some improvement.

In Fig. 11, the prediction score of the “Y”, “F”, “Ө”, “Г”, “Л”, “Д” images was low. The results were explained in the right images in gray color. If pink dots defined objects of the image that affect the model prediction. At the same time, we noticed that at least the model pays attention mainly to the shape of a hand. However, the model did not care about the background. Focusing on these areas may help to further improve the accuracy of the model. In the rest of the images, there are pink dots, which means that those pixels contributed to the class prediction.

One of the directions that this research can develop is the possibility of integrating a trained neural network into mobile devices. Mobile devices are the primary tool where sign language recognition can have practical applications. But there are large architectures with millions of parameters. This causes the size and resource issue on mobile

devices. One of the possible solutions is to use compression techniques for the trained neural network. Another approach is to use lightweight architectures such as LeNet.

We emphasized that the research results helped us understand that they are widely used in applied problems, for example, in mobile applications, you can use a program for recognizing gestures in pictures or video images. It can also be used in applied problems or fields, for example, in training programs for those who want to learn sign language, as well as for people with disabilities.

7. Conclusions

1. A marked up dataset of gesture images is assembled, which consists of 2100 images for 42 classes of gestures, 50 images for each class. The validation of CNN models has shown that this volume of images is sufficient. Also, when training models using LeNet and EffectiveNet, we did not notice overtraining. This number of images will allow us to avoid the intermediate stage of work on dataset augmentation.

2. The findings suggest that LeNet and EffectiveNet show the best results among the considered architectures. LeNet has an accuracy of 0.990, penalty matrix score of 3.017, EffectiveNet has an accuracy of 0.967, penalty matrix score of 3.017. Architecture selection depends on the production task, LeNet is more “light”, EffectiveNet is “heavier”.

3. We used SHAP to explore the model to define the feature importance and detect complex relationships between features in the images. Focusing on these areas may help to further improve the accuracy of the model. Using SHAP images helps to define that the model needs to be improved in some classes, especially about “F”, “H”, “Қ”, “Y” classes.

Acknowledgments

This research is funded by the Science Committee of the Ministry of Education and Science of the Republic of Kazakhstan (AP08053034).

References

1. Bazarevsky, V., Fan, Zh. (2019). On-device, real-time hand tracking with mediapipe. Google AI Blog. Available at: <https://ai.googleblog.com/2019/08/on-device-real-time-hand-tracking-with.html>
2. Lee, A., Cho, Y., Jin, S., Kim, N. (2020). Enhancement of surgical hand gesture recognition using a capsule network for a contactless interface in the operating room. *Computer Methods and Programs in Biomedicine*, 190, 105385. doi: <https://doi.org/10.1016/j.cmpb.2020.105385>
3. Bilgin, M., Mutludogan, K. (2019). American Sign Language Character Recognition with Capsule Networks. 2019 3rd International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT). doi: <https://doi.org/10.1109/ismsit.2019.8932829>
4. Adithya, V., Rajesh, R. (2020). A Deep Convolutional Neural Network Approach for Static Hand Gesture Recognition. *Procedia Computer Science*, 171, 2353–2361. doi: <https://doi.org/10.1016/j.procs.2020.04.255>
5. Mantecón, T., del-Blanco, C. R., Jaureguizar, F., García, N. (2016). Hand Gesture Recognition Using Infrared Imagery Provided by Leap Motion Controller. *Lecture Notes in Computer Science*, 47–57. doi: https://doi.org/10.1007/978-3-319-48680-2_5
6. Kumar, A., Thankachan, K., Dominic, M. M. (2016). Sign language recognition. 2016 3rd International Conference on Recent Advances in Information Technology (RAIT). doi: <https://doi.org/10.1109/rait.2016.7507939>
7. Haberdar, H., Albayrak, S. (2005). Real Time Isolated Turkish Sign Language Recognition from Video Using Hidden Markov Models with Global Features. *Lecture Notes in Computer Science*, 677–687. doi: https://doi.org/10.1007/11569596_70
8. Saykol, E., Türe, H. T., Sirvanci, A. M., Turan, M. (2016). Posture labeling based gesture classification for Turkish sign language using depth values. *Kybernetes*, 45 (4), 604–621. doi: <https://doi.org/10.1108/k-04-2015-0107>

9. Kudubayeva, S. A., Ryumin, D. A., Kalzhanov, M. U. (2016). The method of basis vectors for recognition sign language by using sensor KINECT. *Journal of Mathematics, Mechanics and Computer Science*, 91 (3), 86–96. Available at: <https://bm.kaznu.kz/index.php/kaznu/article/view/541>
10. Uskenbayeva, R. K., Mukhanov, S. B. (2020). Contour analysis of external images. *Proceedings of the 6th International Conference on Engineering & MIS 2020*. doi: <https://doi.org/10.1145/3410352.3410811>
11. Lecun, Y., Bottou, L., Bengio, Y., Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86 (11), 2278–2324. doi: <https://doi.org/10.1109/5.726791>
12. Krizhevsky, A., Sutskever, I., Hinton, G. E. (2017). ImageNet classification with deep convolutional neural networks. *Communications of the ACM*, 60 (6), 84–90. doi: <https://doi.org/10.1145/3065386>
13. He, K., Zhang, X., Ren, S., Sun, J. (2016). Deep Residual Learning for Image Recognition. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). doi: <https://doi.org/10.1109/cvpr.2016.90>
14. Tan, M., Le, Q. V. (2019). EfficientNet: Rethinking model scaling for convolutional neural networks. *Proceedings of the 36th International Conference on Machine Learning*. Available at: <https://proceedings.mlr.press/v97/tan19a/tan19a.pdf>
15. Merembayev, T., Kurmangaliyev, D., Bekbauov, B., Amanbek, Y. (2021). A Comparison of Machine Learning Algorithms in Predicting Lithofacies: Case Studies from Norway and Kazakhstan. *Energies*, 14 (7), 1896. doi: <https://doi.org/10.3390/en14071896>
16. Lundberg, S. M., Lee, S.-I. (2017). A unified approach to interpreting model predictions. 31st Conference on Neural Information Processing Systems (NIPS 2017). Available at: <https://arxiv.org/pdf/1705.07874.pdf>