All Sciences Proceedings
http://as-proceeding.com/

1st International Conference on Trends in Advanced Research

March 4-7, 2023 : Konya, Turkey



ICTAR

# Sign Language Alphabets Classification by Convolutional Neural Networks

Halit Çetiner*

[1]*Vocational School of Technical Sciences, Isparta University of Applied Sciences, Türkiye*

*(halitcetiner@isparta.edu.tr) *Email of the corresponding author*

***Abstract*** – Sign language, which is not a universal language, contains differences in terms of language and communities. Recognizing and translating the symbols of sign languages is a very important method in the communication of citizens with disabilities. Sign language numerals consist of numbers from 0 to 9, while alphabetic letters cover the alphabet ranging from the letter A to the letter Z. An artificial intelligence supported automatic classification system has been developed by using a 24-letter data set in the meaning of Sign language (SL) from the symbols representing the specified alphabetic signs. Recently, Convolutional Neural Network (CNN) based models have performed quite well in computer vision problems. Detailed feature maps were created using the unique automatic distinctive feature extraction structure of CNN methods. In this context, different numbers of hidden layers are defined in the proposed model to capture detail features. The results were evaluated in terms of F1 score, accuracy, recall and precision performance metrics obtained by Adam and Adamax optimization methods with a new CNN model consisting of a total of 15 layers. The proposed CNN model provided 0.99 performance metrics in terms of accuracy, precision, recall and F1 score with Adam and Adamax optimization method. It has been observed that training and test performance measures are close to each other and give satisfactory results in terms of performance metrics. With these results, more effective results can be obtained with CNN models, Capsule Networks structures that keep the spatial relations of the features in further studies.

*Keywords –Classification, CNN, Sign Language, Artificial Intelligence, Deep learning*

## I. INTRODUCTION

Thousands of data are shared every day from social networks where data is increasing rapidly. Powerful artificial intelligence libraries are needed to extract meaningful information from this growing data. Powerful artificial intelligence libraries such as Pytorch and tensorflow can be run on graphics processors with powerful computational capabilities such as Graphics Processing Unit (GPU) as well as Central Processing Unit (CPU) support. A new research area has been opened in the fields of image classification with the methods called Convolutional Neural Network (CNN) instead of classical machine learning techniques by using the specified libraries. [1]. Mathematical and statistical analyzes are needed in many areas from wheat yield estimation to analysis of fake news data. [1], [2]. CNN methods can be used in these analyzes as well.

The MNIST dataset of handwritten digits has been expanded for different problems, from solving optimization problems to classifying fashion images. Using the MNIST dataset, the researchers produced new approaches for use in real-world problems from different perspectives. The MNIST dataset has a format adapted for fashion called Fashion MNIST. It is a dataset developed to perform automatic classification of different clothing types in this format [3]. It is reported that there are 70,000 fashion images from this data set. In the dataset with 10 different categories, each image has 28 width and height values. Based on the Fashion MNIST dataset given to the detail specifications, a new dataset was created for sign languages. The Sign Language MNIST dataset is a multi-class problem with hand gestures representing a total of 24 letters.

Sign languages created by facial and hand gestures should be defined. The language used by disabled people to express their feelings and thoughts to ordinary people is called sign language [4]. It is stated that sign language is used by approximately 2,500,000 people in the world [4]. Different approaches have been developed for people with disabilities to communicate with each other and ordinary people. The Sign Language (SL) used in this article is a heavily used language for its alphabet recognition [5]. SL is a sign language in which people in the disabled population communicate their thoughts using fingers, hand, facial expressions and gestures [4]. If these sign languages are not used, it is not possible to communicate between disabled citizens. Sign languages are the easiest and most effective way to solve this problem [6].

CNN, which is one of the deep learning-based effective methods, is used in many methods from semantic segmentation to localization and visual classification problems [7]–[9]. It is seen that various studies have been carried out to classify SL alphabets in the literature. Previously, it is seen that the histogram gradient features were classified using Support Vector Machine (SVM) algorithms for real-time hand gesture recognition [10]. In another study, it is seen that the results obtained from machine learning techniques such as Random Forest, SVM, MultiLayer Perceptron and CNN-based methods are compared [11]. With the Deep CNN model, a model has been developed to describe hand movements in the SL dataset [4]. They developed a new architecture to describe sign languages with CapsNet based campus networks [6].

The following sections of the article consist of the dataset used, the performance results obtained and the conclusion of the study.

## II. Materials and Method

The Sign Language MNIST dataset contains 27,455 training samples, 7,172 test images, half the size of the standard MNIST dataset.

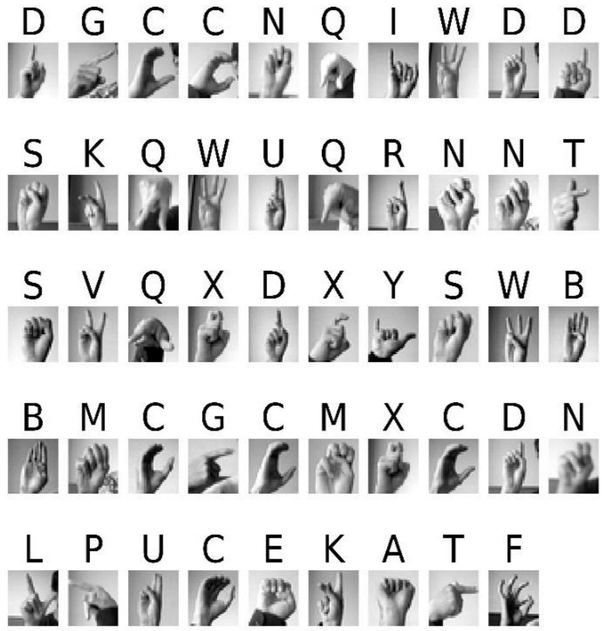

Fig. 1. Example image in dataset

Each image in the dataset has a width and height of 28. This data set has been developed to be useful to people with hearing and hearing difficulties. Some of the alphabets in the data set are shown in Figure 1.

### A. *Proposed Method*

A 16-layer CNN model is proposed for automatic classification of hand signs representing the letters SL. The Proposed model is shown in Figure 2. 28x28x1 input images are taken in Proposed CNN model. The images are taken in one dimension and the operations are performed quickly. In the second layer, a convolution layer was created by applying 125 filters in 3x3 window sizes. ReLU activation function is used in the created layer. By adding the Batch normalization layer in the third layer, the arrangement between the model layers is carried out.

In the fourth layer, a maximum pooling layer of 2x2 dimensions is defined. In the fifth layer, 100 filters were moved over the images with 3x3 windows. In the sixth layer, 0.3 neuronal dropout was achieved. Batch normalization layer was applied in the seventh layer as in the third layer. In the eighth layer, 2x2 maximum pooling was applied. In the ninth layer, 3x3 windows and feature maps consisting of 75 filters were created.

In the tenth layer, the batch normalization layer was applied as in the third and seventh layers. In the eleventh layer, a maximum pooling layer of 2x2 dimensions was applied. In the twelfth layer, the Flatten layer, which converts the obtained feature map to vector, is applied. In the thirteenth layer, the model was strengthened by defining 512 hidden neuron connections. In the fourteenth layer, 0.5 neuron dropout was performed, preventing the model from memorizing.

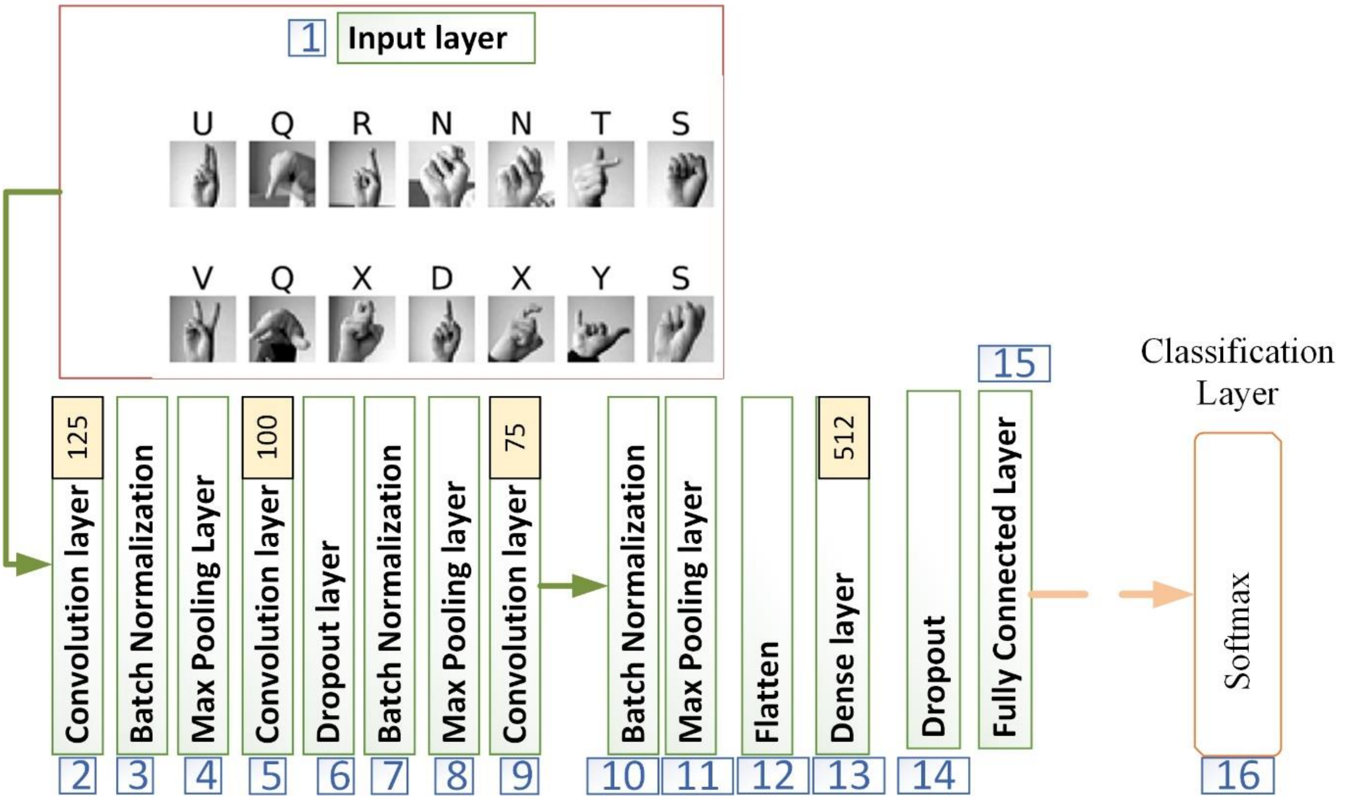


Fig. 2. Proposed model for SL

## III. RESULTS

Performance metrics obtained by Adam and Adamax optimization methods from the proposed model, whose block diagram is given in Figure 2, are shown in Figure 3-7. Adam and Adamax optimization function gave similar results in terms of accuracy in both training and validation. Although there are fluctuations in the validation results, the performance metrics obtained are similar to each other in terms of accuracy.

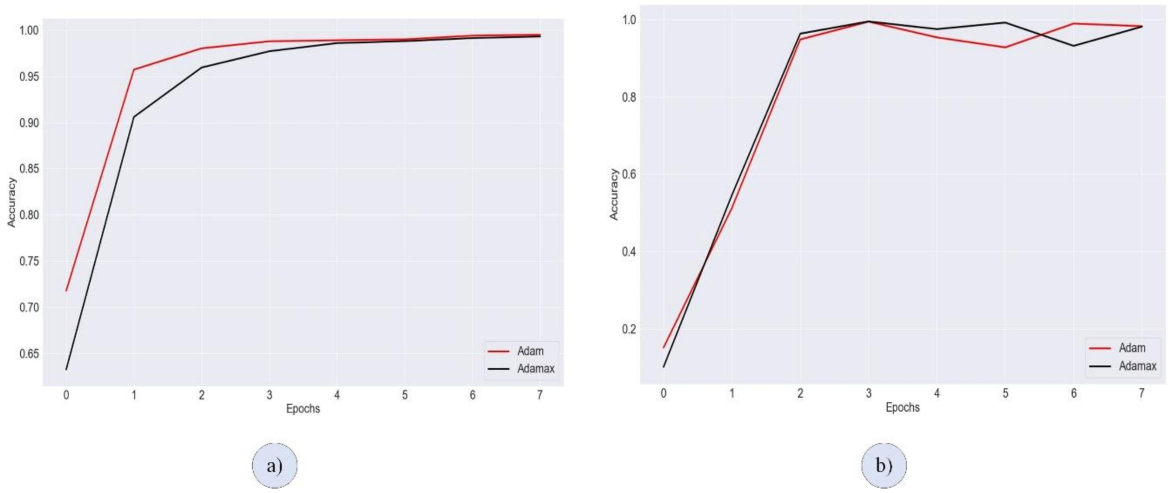


Fig. 3. a) Train accuracy, b) Validation accuracy

While Adam optimization training loss function decreased to 0.016, Adamax optimization method decreased to 0.023 loss value. The Adam function provided better results than the Adamax function. The Adam and Adamax optimization method decreased the validation loss to 0.053 as the loss value.

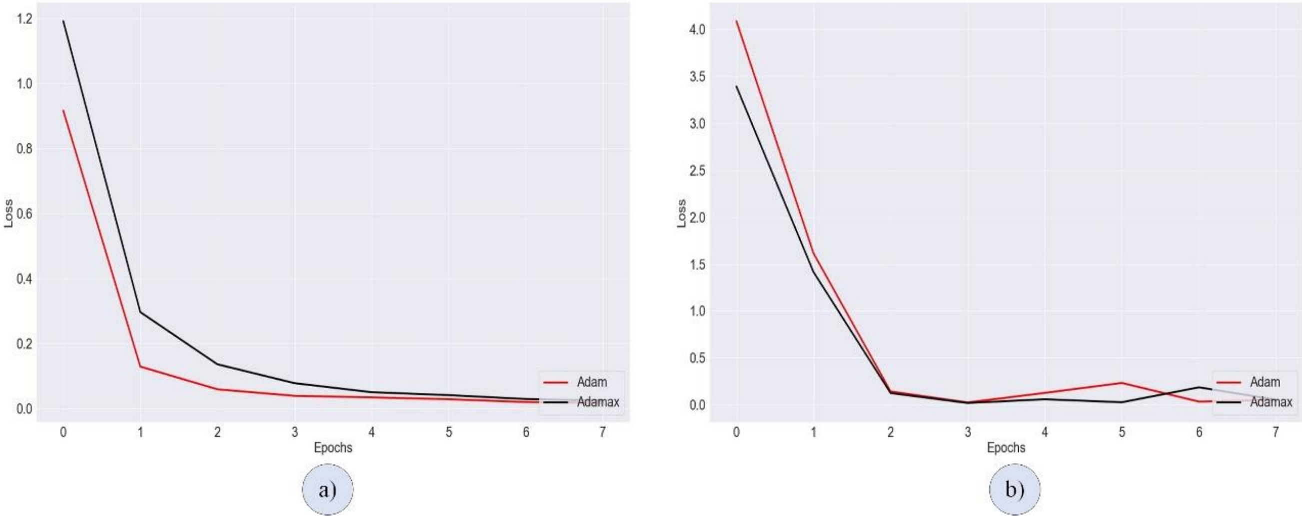


Fig. 4. a) Train loss, b) Test loss

Adam and Adamax optimization function gave similar results in terms of F1 score in both training and validation. Although there are fluctuations in the validation results, the performance metrics obtained are similar to each other in terms of the F1 score.

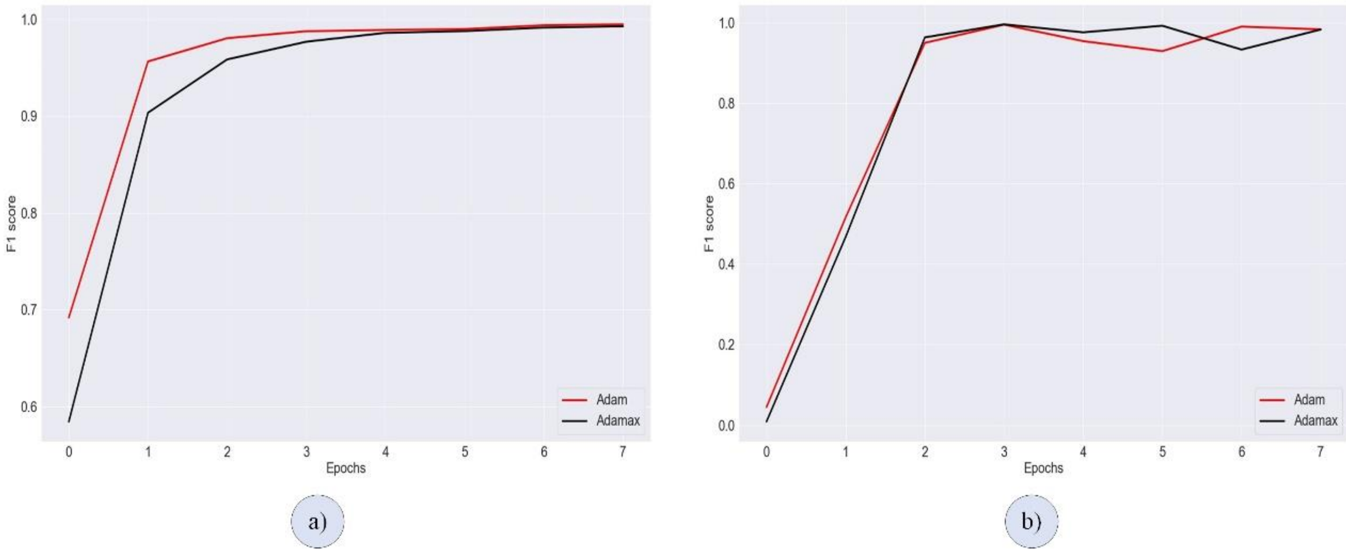


Fig. 5. a) Train F1 score, b) Validation F1 score

Adam and Adamax optimization function yielded similar results in terms of precision in both training and validation. Although there is fluctuation in the validation results, the performance metrics obtained are similar to each other in terms of precision.

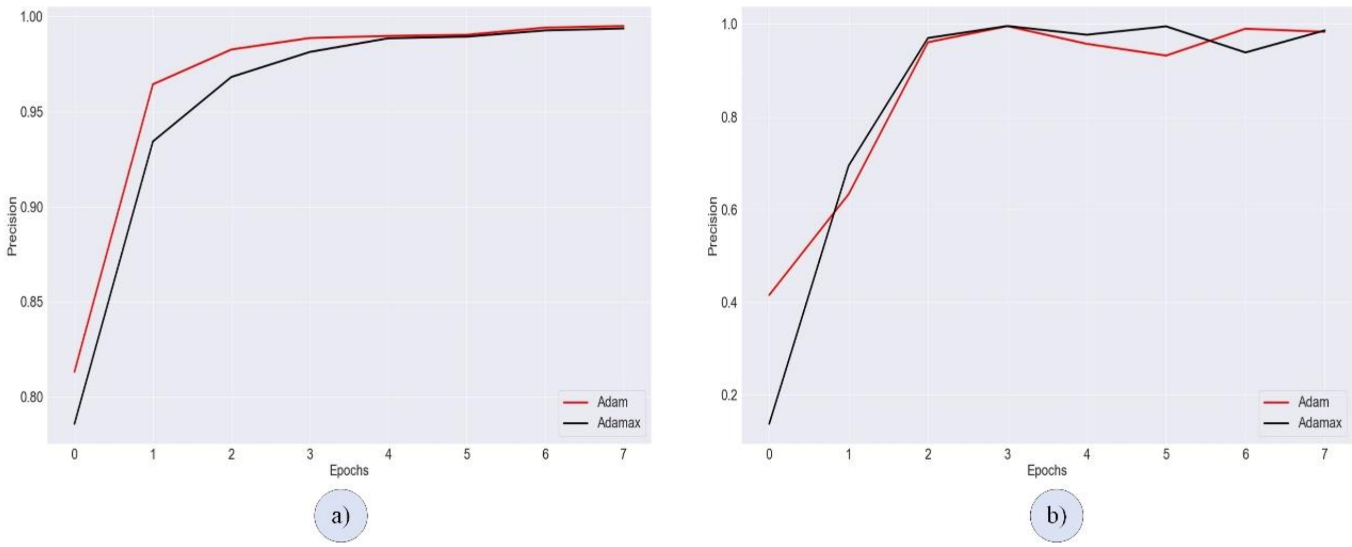


Fig. 6. a) Train precision, b) Validation precision

Both training and validation gave similar results in terms of Adam and Adamax optimization function recall. Although there are fluctuations in the validation results, the performance metrics obtained are similar to each other in terms of recall.

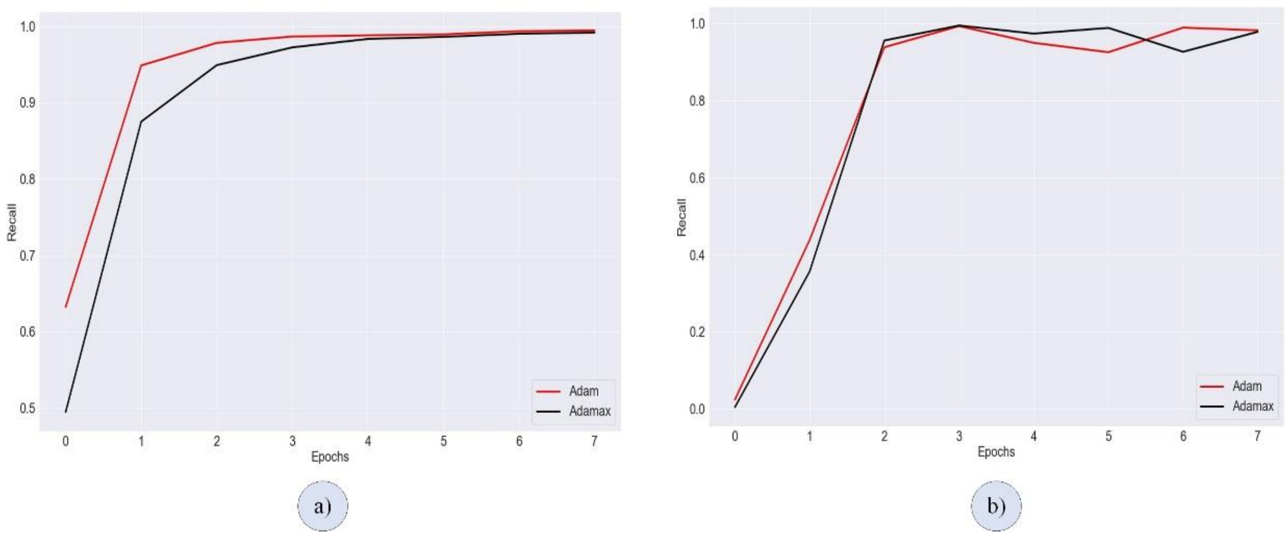


Fig. 7. a) Train recall, b) Validation recall

## IV. CONCLUSION

In the real world, SL recognition is an area of research because of its practical use. There is an increase in research in this field day by day. The proposed CNN method gave good performance in SL recognition. CNN architectures are very successful in extracting distinctive features as well as being scalable and generalizable. The proposed CNN architecture consists of convolution-based architectures using different filters and window sizes. In further studies, in addition to CNN methods, different studies on SL recognition and classification can be developed with the recently popular capsule network-based architectures.

## REFERENCES

[1] H. Çetiner and B. Kara, “Recurrent Neural Network Based Model Development for Wheat Yield Forecasting,” *J. Eng. Sci. Adiyaman Univ.*, vol. 9, no. 16, pp. 204–218, 2022, doi: 10.54365/adyumbd.1075265.

[2] J. A. Nasir, O. S. Khan, and I. Varlamis, “Fake news detection: A hybrid CNN-RNN based deep learning approach,” *Int. J. Inf. Manag. Data Insights*, vol. 1, no. 1, p. 100007, 2021, doi: https://doi.org/10.1016/j.jjimei.2020.100007.

[3] H. Xiao, K. Rasul, and R. Vollgraf, “Fashion-mnist: a

novel image dataset for benchmarking machine learning algorithms," *arXiv Prepr. arXiv1708.07747*, 2017.

[4] A. Mannan, A. Abbasi, A. R. Javed, A. Ahsan, T. R. Gadekallu, and Q. Xin, "Hypertuned Deep Convolutional Neural Network for Sign Language Recognition," *Comput. Intell. Neurosci.*, vol. 2022, p. 1450822, 2022, doi: 10.1155/2022/1450822.

[5] S. Y. Kim, H. G. Han, J. W. Kim, S. Lee, and T. W. Kim, "A hand gesture recognition sensor using reflected impulses," *IEEE Sens. J.*, vol. 17, no. 10, pp. 2975–2976, 2017.

[6] H. Xiao *et al.*, "Sign language digits and alphabets recognition by capsule networks," *J. Ambient Intell. Humaniz. Comput.*, vol. 13, no. 4, pp. 2131–2141, 2022, doi: 10.1007/s12652-021-02974-8.

[7] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Commun. ACM*, vol. 60, no. 6, pp. 84–90, 2017.

[8] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 3431–3440.

[9] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask r-cnn," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2961–2969.

[10] E. Ohn-Bar and M. M. Trivedi, "Hand gesture recognition in real time for automotive interfaces: A multimodal vision-based approach and evaluations," *IEEE Trans. Intell. Transp. Syst.*, vol. 15, no. 6, pp. 2368–2377, 2014.

[11] D. Chakraborty, D. Garg, A. Ghosh, and J. H. Chan, "Trigger detection system for american sign language using deep convolutional neural networks," in *Proceedings of the 10th International Conference on Advances in Information Technology*, 2018, pp. 1–6.