Jilin Daxue Xuebao (Gongxueban)/Journal of Jilin University (Engineering and Technology Edition)
ISSN：1671-5497
E-Publication Online Open Access
Vol: 41 Issue: 11-2022
DOI 10.17605/OSF.IO/NKHJT

DESIGN AND IMPLEMENTATION OF CNN FOR SIGN LANGUAGE RECOGNITION

IMRAN KHAN
Department of Telecommunication Engineering, Dawood University of Engineering & Technology, Karachi, Pakistan.

SOHAIL RANA
Department of Electronic Engineering, Dawood University of Engineering & Technology, Karachi, Pakistan.

NADIA MUSTAQIM ANSARI
Department of Electronic Engineering, Dawood University of Engineering & Technology, Karachi, Pakistan.

RIZWAN IQBAL
Department of Telecommunication Engineering, Dawood University of Engineering & Technology, Karachi, Pakistan.

TALHA TARIQ
Department of Electronic Engineering, Dawood University of Engineering & Technology, Karachi, Pakistan.

MAQSOOD UR REHMAN AWAN
Department of Electronic Engineering, Dawood University of Engineering & Technology, Karachi, Pakistan.

ADNAN WAQAR
Department of Electronic Engineering, Dawood University of Engineering & Technology, Karachi, Pakistan.

Abstract

Every day, we witness numerous deaf, mute, and blind people. They have a hard time interacting with others. Sign Language is a language in which we use hand movements and gestures to communicate with people who are mainly deaf and hard of hearing. This paper proposes a system to recognize hand gestures using different libraries of Python and Deep Learning Algorithms, to process the image and predict the gestures. The Web Camera captures images of various gestures used as input, and this project shows the sign language recognition of 1-10 digits hand gestures, including OK and Salaam. Modules for preprocessing and feature extraction, model training, testing, and sign-to-text translation are included in the proposed system. Various CNN architectures and preprocessing techniques, such as greyscale and thresholding, were built and evaluated using our dataset to improve recognition accuracy. Our proposed system of convolutional neural networks (CNNs) obtains an average accuracy of 98.76 percent for real-time hand gesture identification on a dataset with nine hand movements and 500 photos for each gesture.

Keywords*:* Sign Language Recognition, MLP, Deep Learning, Feature Extraction, Neural Network

Jilin Daxue Xuebao (Gongxueban)/Journal of Jilin University (Engineering and Technology Edition)
ISSN：1671-5497
E-Publication Online Open Access
Vol: 41 Issue: 11-2022
DOI 10.17605/OSF.IO/NKHJT

I. INTRODUCTION

People often use sign-language gestures as a method of non-verbal communication to prompt their thoughts and feelings[1]. However, non-signatories find it tremendously tough to recognize, and society tends to ignore and isolate these physically challenged humans[2]. People with excellent hearing no longer recognize sign language and often lack speaking with deaf people, so SLR is added to bridge the space between normal humans and the hearing-impaired, so one must know sign language[3]. Skilled sign language interpreters are desired in clinical and criminal appointments and instructional and educational classes[4]. There has been an increasing demand for interpreting the offerings[5]. So they will offer an easy-to-implement sign language interpreter service that can be used, although it has the most important problems such as access to the Internet and the right equipment.

With a strong objective of providing artificially intelligent personal assistive technology to the specially-abled community to overcome their challenges, we aim at deep learning using a sign language Hand gesture recognition system is to be developed[6]. This paper aims to convert human sign language into text with human gesture understanding and motion capture". Two opposing yet complementary goals drive the research. The first is that a sign language system could help people communicate more effectively between a member of the deaf community and the hearing community, and the second is that such a system would be developed using neural networks.

The idea of using technology to identify sign language is often disregarded. It has the potential to help a large social group. Image classification and deep learning can help computers recognize sign language, which can be interpreted by others [7]. This work employs a Deep Neural Network to recognize sign language gestures. Static sign language gestures were taken on the webcam and utilized as the photograph dataset. The photos were processed first, and then the cleaned inputs were used. This sign language gesture dataset is retrieved and tested on a Deep Neural Network model using OpenCV and Tensor Flow.

II.PROBLEM DESCRIPTION

Communication with those who do not understand Sign Language becomes a major difficulty for deaf and hard of hearing people. Understanding the exact context of the symbolic expression of deaf and dumb people is quite challenging for normal people in real life until it is properly specified. A sign language interpreter is required in certain scenarios to ensure communication and understanding.

The issue statement focuses on the concept of a camera-based sign language recognition system or sign language interpreter that would be used by the deaf to transform their sign language gestures into text, enabling hearing individuals to comprehend the precise meaning of each sign. Our objective is to design a solution that is intuitive and simple. Communication for most people is not difficult. It should be the same way for the deaf.

III.RELATED WORK

There were many reforms in the era, and much study has benefited the deaf and dumb people.

In [8], one of the earliest works in Indian Sign Language (ISL) largely considers the identification of different hand signs and traits and therefore selects some traits often from the ISL for identification. This paper offers robust modeling of stationary signals and traits in the context of signal language recognition using Deep Learning-Based Convolutional Neural Networks (CNNs). In this study, a total of 35,000 sign photos of one hundred static signals have been collected from exclusive users. The performance of the proposed device is evaluated on about 50 CNN models. The effects are also evaluated based on different optimizers, and it is observed that the proposed method has the highest training accuracy on color and grayscale images, respectively.

Similarly, in [9], the user should be able to capture hand gesture pictures using a webcam, and the system will predict and display the name of the captured picture. The HSV color set of rules is used to find the hand gesture and set the background to black. Images undergo a series of processing steps, including conversion to grayscale, dispersion and mask operation, and various PC Vision techniques. Moreover, the region of interest, which in our case is a hand gesture, is fragmented. The extracted features are the binary pixels of the images. Convolutional neural networks (CNNs) are used for training and classifying images. 10 American Sign Gesture characters are recognized with high accuracy. This model has achieved remarkable accuracy of above 90%"

In [10], They focused on translation challenges and the characteristics of Russian sign language. The paper emphasizes that, rather than serving as an inclusive SLR translator, the major goal of the SLR system is to provide an operational interpreter with the demands of a matching deaf population. The neural network and input prediction are combined in a new way in the suggested system. It provides the user with an additional option to adjust the result. Review the most recent data collecting techniques as well. By anticipating the next letter, the approach uses the inherent qualities of fingerspelling to improve recognition performance and accuracy.

This [11] study uses an IISL2020 dataset of various hand motions to demonstrate the identification of Indian Sign Language using LSTM and GRU. The suggested model performs better on frequently used terms like "Good Morning," "Thank you," "work," etc. than any other ISL model currently in use. Additionally, adding more layers to the LSTM

Jilin Daxue Xuebao (Gongxueban)/Journal of Jilin University (Engineering and Technology Edition)
ISSN：1671-5497
E-Publication Online Open Access
Vol: 41 Issue: 11-2022
DOI 10.17605/OSF.IO/NKHJT

and GRU and applying LSTM before GRU helps the model recognize ISL more accurately. Over 11 distinct indications, they have an accuracy rate of almost 97%.

In this paper, [12] CNN is employed for performance optimization) This study has attained a high success rate to a new language, Pashto language. Their method will convert the signals and gestures of persons with disabilities into the Pashto alphabet. A CNN model is used to 2500 dataset photos to detect the hand motions. The overall accuracy of this system is 90%, while some gestures have a 98% accuracy rate.

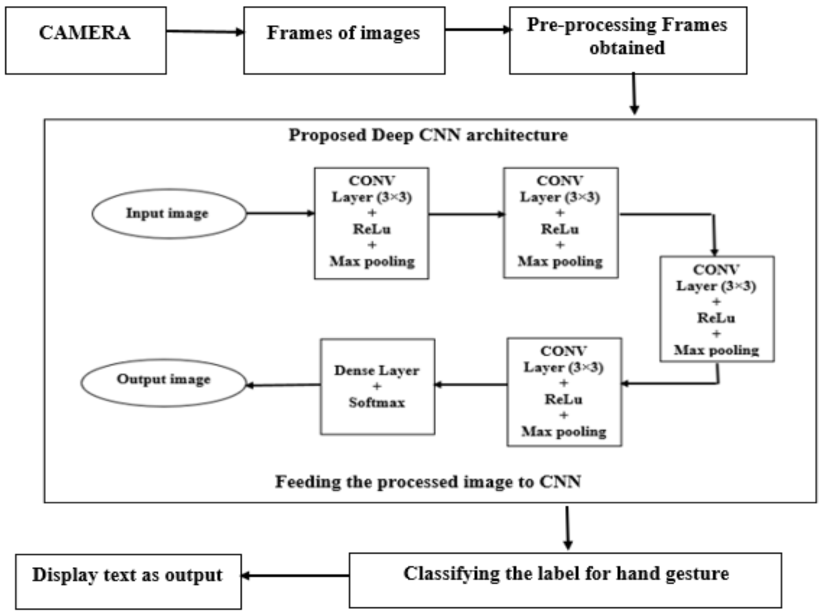
The previous work in this area bridges the communication gap between the hearing impaired and the public. Various researchers did their work using different techniques and algorithms to achieve this objective and introduced their own sign language rearrangement systems. Based on this research, the work is an approach to enhance human well-being using technology such as deep learning.

IV.METHODOLOGY

In this paper, we developed a convolutional neural network to classify finger spell images using image intensity and depth data. The developed convolution network is evaluated by applying it to finger gesture recognition. Image recognition is called the act of directly feeding a picture into a neural network and generating some form of caption for that image. Figure 1 describes the flow of our work. The first Camera takes the input image or sign and creates frames of images. The label with which the network's output will be associated with a predefined class. There could be multiple classes identified as a picture or only one. When there is just one class, the phrase "popularity" is frequently used. However, when there are multiple classes, the term "classification" is frequently used.

Jilin Daxue Xuebao (Gongxueban)/Journal of Jilin University (Engineering and Technology Edition)
ISSN：1671-5497
E-Publication Online Open Access
Vol: 41 Issue: 11-2022
DOI 10.17605/OSF.IO/NKHJT

Figure 1: System Block Diagram



a) Creating the Dataset

To create a dataset although we may be able to get the dataset from the Internet, we build the dataset ourselves for this research. The frame identifies a hand in the ROI (region of interest) created and will be saved in a directory called "gesture directory," as shown in figure 2 that has two directories, "train" and "test". Each directory has ten files, as shown in figure 3. The testing folder directory is similar to the train directory. It contains pictures that were gathered using the creating gesture data.py function.

Figure 2 : Directory Structure

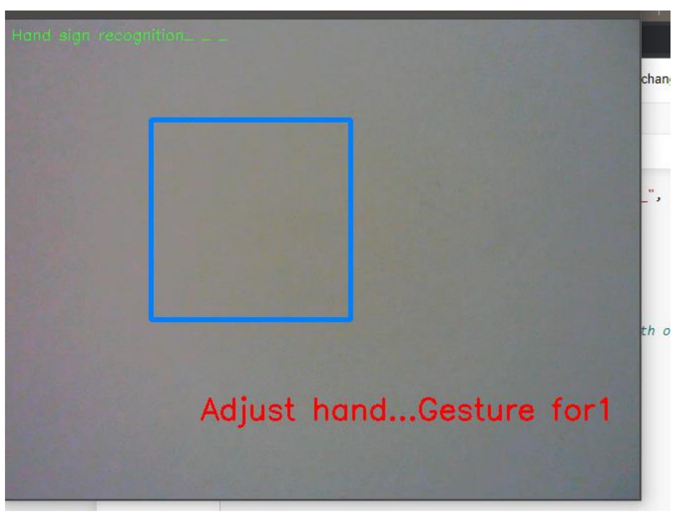
Jilin Daxue Xuebao (Gongxueban)/Journal of Jilin University (Engineering and Technology Edition)
ISSN：1671-5497
E-Publication Online Open Access
Vol: 41 Issue: 11-2022
DOI 10.17605/OSF.IO/NKHJT

Figure 3: Train Directory

b) ROI (Region of Interest)

To create the dataset and record the live camera stream, we used OpenCV. Using ROI, the portion of the frame identifies the hand for the motions. As shown in figure 4, the box represents the ROI window that allows access to the webcam's live feed. To distinguish between the backgrounds, we calculate the background's cumulative weighted average and then deduct it from the frames with a different foreground object in front of the background. To find any object that fills the background, we deduct the cumulative background average from each frame retrieved after 60 frames.

Figure 4: Region of Interest

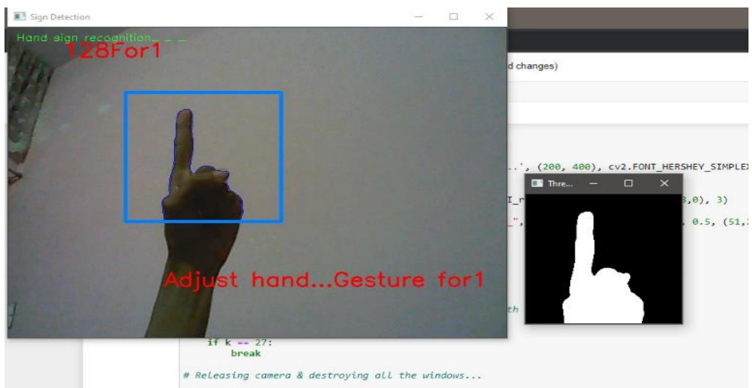


Jilin Daxue Xuebao (Gongxueban)/Journal of Jilin University (Engineering and Technology Edition)
ISSN：1671-5497
E-Publication Online Open Access
Vol: 41 Issue: 11-2022
DOI 10.17605/OSF.IO/NKHJT

c) Determine the Threshold Value

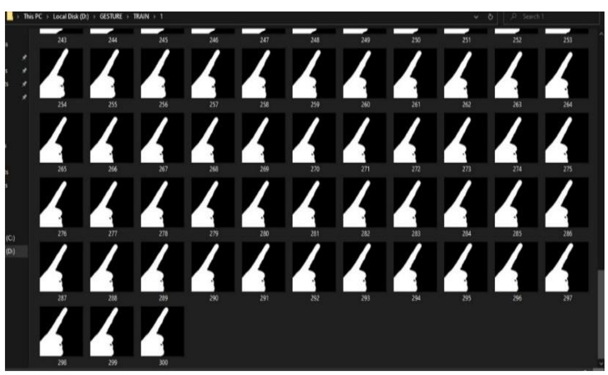
We now find Contours to analyze each frame's threshold value and identify the contours. The segment function allows us to return the maximum number of contours (the object's outermost contours). This helps us determine if there is any foreground item detected in the ROI. In other words, we can tell if a hand is present in the ROI. If contours are identified (or a hand is present in the ROI), we begin storing images of the ROI in the train and test sets for further identification of letters or numbers, as shown in figure 5.

Figure 5 : Gesture of 1



We have 300 photos of each number in the training dataset as represented in figure 6 and 40 photos of each number in the testing dataset.

Figure 6 : Multiple Images of Each Gesture



d) Training CNN

The created data set is now used to train a CNN. We first load the data using the Keras Image Data Generator, which enables us to import the train and test set data using the movement from directory function, with the names of each of the number folders

Jilin Daxue Xuebao (Gongxueban)/Journal of Jilin University (Engineering and Technology Edition)
ISSN：1671-5497
E-Publication Online Open Access
Vol: 41 Issue: 11-2022
DOI 10.17605/OSF.IO/NKHJT

matching the class names for the loaded photographs. The data contains the following sign as displayed in figure 7 and 8

Figure 7 : Signs representing Numbers from 1-10

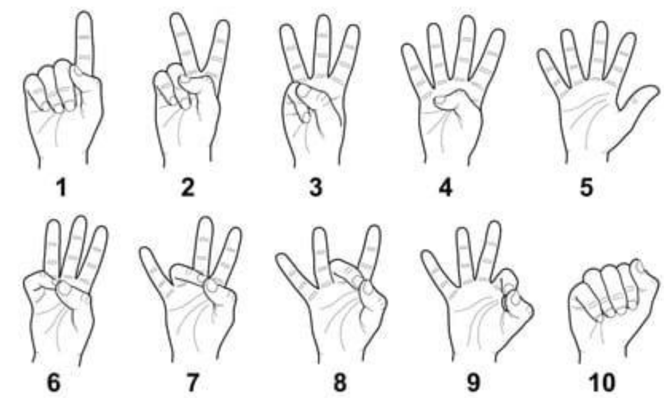


Figure 8 : Signs representing phrases Salam and OK

The model must fit and be stored to utilize in the final module. SGD (stochastic gradient descent, which updates the weights after each training instance) and Adam (a mixture of Adagrad and RMSProp) are the two optimization techniques that are applied. We found that the SGD model had better accuracy. As is obvious, we trained the model with a 100% accuracy rate and an 81 percent validation accuracy. We're modelling and doing a fast test on the model while detecting on the live video stream to make sure everything is operating as it should.

We include "Ten" after "One" in the dictionary because the generator checks the folders inside the test and train folders based on their directory names, such as "1" and "10," when using the Image Data Generator to load the dataset. Therefore, 10 follows 1 in alphabetical order.

Jilin Daxue Xuebao (Gongxueban)/Journal of Jilin University (Engineering and Technology Edition)
ISSN：1671-5497
E-Publication Online Open Access
Vol: 41 Issue: 11-2022
DOI 10.17605/OSF.IO/NKHJT

V.RESULT

We constructed a bounding box while creating the dataset to locate the ROI and determine the cumulative average. This procedure is used to recognize any foreground objects. Now that we have determined the largest contour, we utilize the ROI's threshold as a test picture to see if a hand has been detected.

The previously saved model is loaded using Keras models, and the threshold picture of the ROI containing the hand is sent as model input for prediction.

After acquiring the required imports for the model, we load the model we had previously built and set some of the variables we need to initialize the background variable and specify the ROI dimensions. Determine the background cumulative weighted average, as we did while creating the dataset. The hand is segmented by getting the hand's maximum contours and threshold image. On the live video stream, it will detect the hand gesture.

Figure 9: Representation of a Sign Interpreting 1 in Text

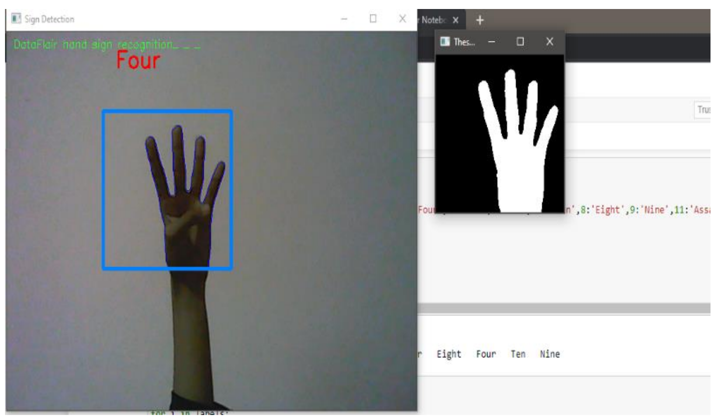
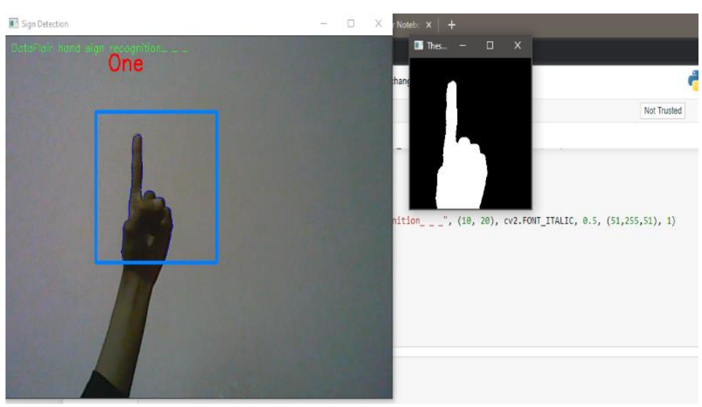


Figure 10 : Representation of a Sign Interpreting 4 in Text



Jilin Daxue Xuebao (Gongxueban)/Journal of Jilin University (Engineering and Technology Edition)
ISSN：1671-5497
E-Publication Online Open Access
Vol: 41 Issue: 11-2022
DOI 10.17605/OSF.IO/NKHJT

Figure 11: Representation of a Sign Interpreting OK in Text

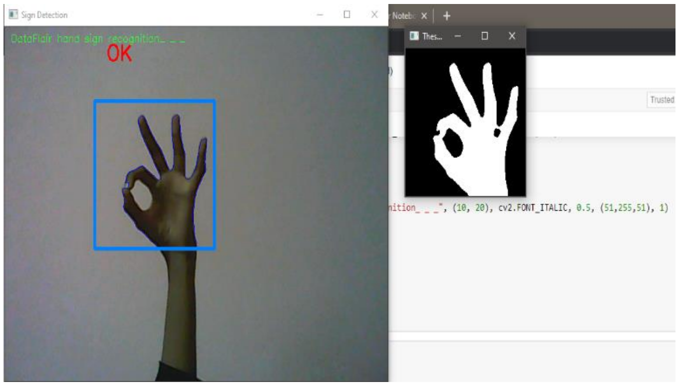
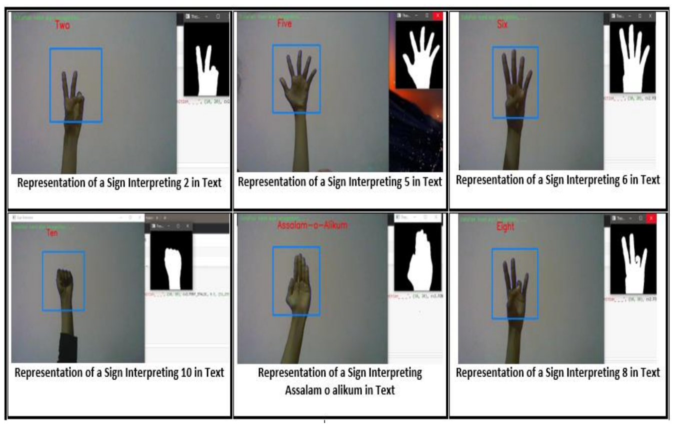


Figure 12: Representation of a Sign Interpreting different signs in Text



Representation of a Sign Interpreting 2 in Text
Representation of a Sign Interpreting 5 in Text
Representation of a Sign Interpreting 6 in Text
Representation of a Sign Interpreting 10 in Text
Representation of a Sign Interpreting Assalam o alikum in Text
Representation of a Sign Interpreting 8 in Text

VI.CONCLUSION

The system is a method for making it easier for people with speech impairments to communicate. As seen by the results above, the model that was constructed pre-processed the picture to the necessary state for it to be input into the model. With the suggested CNN design, decreased training and validation time was seen. Using this system, the collected signs may be inspected, processed, and converted to text for display on the screen. We have discovered that CNN can pick up on and anticipate text. This Sign Language Recognition System has progressed from merely identifying fixed signs and alphanumeric characters to a system that can recognize dynamic motions in continuous series of pictures.

VII.FUTURE RECOMMENDATIONS

Researchers are currently focusing on developing a comprehensive library for sign language recognition systems. Many researchers employ a small vocabulary and a self-

Jilin Daxue Xuebao (Gongxueban)/Journal of Jilin University (Engineering and Technology Edition)
ISSN：1671-5497
E-Publication Online Open Access
Vol: 41 Issue: 11-2022
DOI 10.17605/OSF.IO/NKHJT

made database to construct their Sign Language Recognition System. A dataset built for the system is currently unavailable for some countries engaged in creating Hand Gesture Recognition systems. Especially the Kinect-based data, which provides video in color and depth streams. Researchers differ in their categorization methods for detecting sign language. Comparing one approach to another is still subjective when they use their own concepts and limits for the Sign Language Recognition System. Because of the differences in sign language in different nations and the limitations established by each study, a fair and straight comparison of methodologies is restricted. Most of the country's sign language variations depend on their grammar and how they portray each word, such as displaying the language by word or phrase.

Reference

1. M. Al-Qurishi, T. Khalid, and R. Souissi, "Deep Learning for Sign Language Recognition: Current Techniques, Benchmarks, and Open Issues," *IEEE Access*, vol. 9, pp. 126917–126951, 2021, doi: 10.1109/ACCESS.2021.3110912.
2. I. A. Adeyanju, O. O. Bello, and M. A. Adegboye, "Machine learning methods for sign language recognition: A critical review and analysis," *Intell. Syst. with Appl.*, vol. 12, 2021, doi: 10.1016/j.iswa.2021.200056.
3. N. Adaloglou *et al.*, "A Comprehensive Study on Deep Learning-Based Methods for Sign Language Recognition," *IEEE Trans. Multimed.*, vol. 24, pp. 1750–1762, 2022, doi: 10.1109/TMM.2021.3070438.
4. F. Shah, M. S. Shah, W. Akram, A. Manzoor, R. O. Mahmoud, and D. S. Abdelminaam, "Sign Language Recognition Using Multiple Kernel Learning: A Case Study of Pakistan Sign Language," *IEEE Access*, vol. 9, pp. 67548–67558, 2021, doi: 10.1109/ACCESS.2021.3077386.
5. T. Li, Y. Yan, and W. Du, "Sign Language Recognition Based on Computer Vision," *2022 IEEE Int. Conf. Artif. Intell. Comput. Appl. ICAICA 2022*, pp. 927–931, 2022, doi: 10.1109/ICAICA54878.2022.9844497.
6. Z. Wang *et al.*, "Hear Sign Language: A Real-Time End-to-End Sign Language Recognition System," *IEEE Trans. Mob. Comput.*, vol. 21, no. 7, pp. 2398–2410, 2022, doi: 10.1109/TMC.2020.3038303.
7. I. A. Adeyanju, O. O. Bello, and M. A. Adegboye, "Machine learning methods for sign language recognition: A critical review and analysis," *Intell. Syst. with Appl.*, vol. 12, p. 200056, 2021, doi: 10.1016/j.iswa.2021.200056.
8. A. Wadhawan and P. Kumar, "Deep learning-based sign language recognition system for static signs," *Neural Comput. Appl.*, vol. 32, no. 12, pp. 7957–7968, 2020, doi: 10.1007/s00521-019-04691-y.
9. M. E. Walizad and M. Hurroo, "Sign Language Recognition System using Convolutional Neural Network and Computer Vision," *Int. J. Eng. Res. Technol.*, vol. 9, no. 12, pp. 59–64, 2020, [Online]. Available: www.ijert.org.
10. D. G. Enikeev and S. A. Mustafina, "Sign language recognition through Leap Motion controller and input prediction algorithm," *J. Phys. Conf. Ser.*, vol. 1715, no. 1, 2021, doi: 10.1088/1742-6596/1715/1/012008.

Jilin Daxue Xuebao (Gongxueban)/Journal of Jilin University (Engineering and Technology Edition)
ISSN：1671-5497
E-Publication Online Open Access
Vol: 41 Issue: 11-2022
DOI 10.17605/OSF.IO/NKHJT

11. D. Kothadiya, C. Bhatt, K. Sapariya, K. Patel, A. B. Gil-González, and J. M. Corchado, "Deepsign: Sign Language Detection and Recognition Using Deep Learning," *Electron.*, vol. 11, no. 11, pp. 1–12, 2022, doi: 10.3390/electronics11111780.
12. A. F. Shokoori, M. Shinwari, J. A. Popal, and J. Meena, "Sign Language Recognition and Translation into Pashto Language Alphabets," *Proc. - 6th Int. Conf. Comput. Methodol. Commun. ICCMC 2022*, no. Iccmc, pp. 1401–1405, 2022, doi: 10.1109/ICCMC53470.2022.9753959.