JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue

JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)

An International Scholarly Open Access, Peer-reviewed, Refereed Journal

GESTICULATION RECOGNITION INTO TEXT AND SPEECH FOR NATIVE SIGN LANGUAGE

[1]Sakthivel C, [2]Sneha TM, [3]Shamanth Achar, [4]Rakshith A, [5]Prof. Soumyashree Pattar, [6]Prof. Sneha K

[1234]Final Year Students, [56]Assistant Professors
Department of Information Science and Engineering,
HKBK College of Engineering, Bangalore, India

Abstract : Dumb people are usually deprived of normal communication with other people in the society. It has been observed that they find it really difficult at times to interact with normal people with their gestures, as only a very few of those are recognized by most people. Since people with hearing impairment or deaf people cannot talk like normal people so they have to depend on some sort of visual communication in most of the time. Sign language is the only tool of communication for the person who is not able to speak and hear anything. Sign language is a boon for the physically challenged people to express their thoughts and emotion. In this project, a novel scheme of sign language recognition has been proposed for identifying the alphabets and gestures in sign language. With the help of computer vision and neural networks we can detect the signs and give the respective text and voice output.

Keywords **– Sign language, Classification, K Nearest Neighbor, Hand Gesture Recognition, Text conversion and Speech generation.**

I. INTRODUCTION

The hand gesture is a nonverbal form of communication. It consists of linguistic content that carries a large amount of information in sign language. It also plays a pivotal role in human-computer interaction (HCI) systems. Therefore, automatic hand gesture recognition is in high demand. The importance of automatic hand gesture recognition has increased for the following reasons: (1).The growth of the deaf and hard-of-hearing populations, and (2).The extended use of vision-based and touchless applications and devices such as video games, smart TV control, and virtual reality applications. Robust hand gesture recognition is required as a part of sign language interpretation to help hearing-impaired people. There is a significant communication gap between people who can hear and hearing-impaired people. System between gestural language and verbal language will bridge this communication gap. This translation system will facilitate the lives of hearing-impaired people and help them to integrate with society. CNNs have shown excellent performance in fields such as object and speech recognition, image classification, and edge distribution, and human activity recognition. The existence of large datasets that comprise millions of annotated samples is the main reason behind such excellent performance.

II. LITERATURE SURVEY:

Murat Taskiran et al. [1] have proposed a real-time system for recognition of American Sign Language by using deep learning. Techniques such as classification, convex hull, and convolutional neural network are performed on images in the dataset for better results. Convolutional neural network was trained by using dataset collected in 2011 by Massey University, Institute of Information and Mathematical Sciences, and 100% test accuracy was obtained. The accuracy of the real-time system is 98.05%.

Olivia Kembuan et al. [2] designed an image classification of Indonesia sign language using tensorflow. Techniques such as Convolutional Neural Network (CNN) Architecture and Tensorflow library to build the model image of classification are performed on images in the dataset, which contains 2659 images of Indonesian Sign Language (BISINDO) twenty-six (26) letter categories.The

images are divided into training and validation datasets. The experimental results show that the model has achieved an accuracy of 96.67% on the training dataset, and an accuracy of 100% for the validation dataset.

Dr. S. Pariselvam et al. [3] have propsed an interaction system using speech and gesture based on CNN. This model consists of two main systems. One is voice input is converted into text and hand gestures and second approach is hand gestures conversion to text. These systems are implemented in Python and OpenCV is used to capture images. These systems are free from lighting conditions and background noise by using CNN algorithm.

Dardina Tasmere et al. [4] proposed Hand Gesture Recognition for Bangla Sign Language Using Deep Convolution Neural Network. The hand was detected practicing HSV and YCbCr color space. In total thirty-seven characters (8 vowels and 29 consonants) are recognized by deep convolution neural networks. Test accuracy as gained an accuracy of 99.22%.

Tanuj Bohra et al. [5] proposed a real-time two-way sign language communication system built using image processing, deep learning and computer vision. Techniques such as hand detection, skin color segmentation, median blur and contour detection are performed on images in the dataset for better results. CNN model trained with a large dataset for 40 classes and was able to predict 17600 test images in 14 seconds with an accuracy of 99%.

Salma Hayani et al. [6] proposed an arab sign language recognition system based on CNN, inspired from LeNet-5. Dataset contained 7869 images of arab signs of numbers and letters. Various experiments were conducted by varying the number of training sets from 50% to 80%. 90% accuracy was obtained with 80% training dataset. The author has also compared the results obtained with machine learning algorithms like KNN (k-nearest neighbour) and SVM (support vector machine) to show performance of the system. This model was purely image based and it can be extended to video based recognition.

Kshitij Bantupalli and Ying Xie [7] worked on American Sign Language recognition system which works on video sequences based on CNN, LSTM and RNN. A CNN model named Inception was used to extract spatial features from frames, LSTM for longer time dependencies and RNN to extract temporal features. Various experiments were conducted with varying sample sizes and dataset consists of 100 different signs performed by 5 signers and maximum accuracy of 93% was obtained. Sequence is then fed to a LSTM for longer time dependencies. Outputs of softmax layer and maxpooling layer are fed to RNN architecture to extract temporal features from softmax layer.

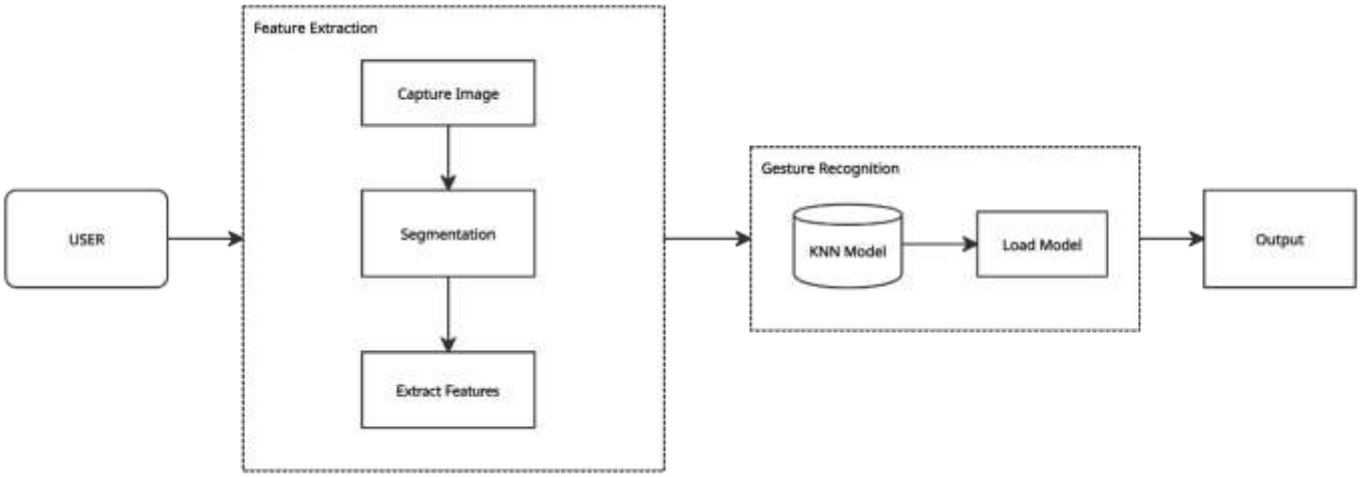
III. EXISTING SYSTEM

The research efforts can be categorized into two approaches:-The contact-based approach and the vision-based approach. In the contact-based approach/ Glove based approaches: In this category requires signers to wear a sensor glove or a colored glove. The task will be simplified during segmentation process by wearing glove. The drawback of this approach is that the signer has to wear the sensor hardware along with the glove during the operation of the system. The signer should be familiar with interfacing devices like motion sensors, data gloves, position trackers, and accelerometers, to collect hand gesture data. This approach disadvantages are the high cost and discomfort to the signer. The studies in the vision-based approach revoked these draw-backs. Various imaging devices such as cameras are used.

IV. PROPOSED SYSTEM

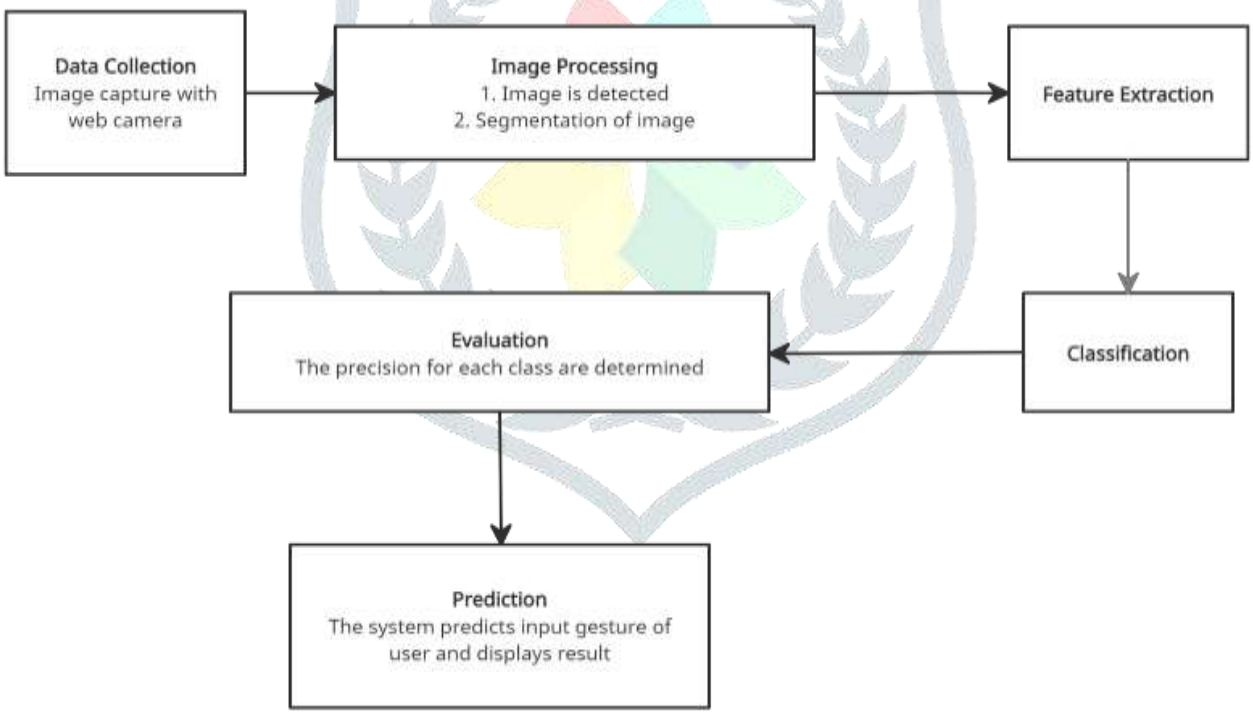
In this study, we utilized a KNN architecture for feature learning using two approaches. In the first approach, KNN was used to extract the features from the entire video sample and used for classification. In the second approach, we aimed to enhance the temporal dependency of the video frames. To achieve this, the same KNN architecture will trained to extract the features from different regions in the video sample. The purpose of the system is to improve the existing system in this area in terms of response time and accuracy with the use of efficient algorithms, high quality data sets and better sensors. Technologies used- Tensorflow, KNN, OpenCv. We then investigated different techniques for feature fusion. The main objective is to develop a system that translates sign language to understandable text and voice which will ease their communication. Acquisition of data's from users. Detection of image and segmentation of the image will be done. Classification of the image takes place. Predicts the input gesture of the user and will display the result. Reduce cost and improve robustness of the proposed system using simple web camera.

4.1 SYSTEM ARCHITECTURE



In Gesticulation Recognition to Text and Speech for Native Sign Language we will collect the data's will be captured through web camera and image will be processed i.e. detection of the image and image segmentation takes place. Then Feature extraction of the image is done then classification will be done to determine the precision of each image. The system will compare the input data with the provided dataset and will predict the output. Later the output is displayed in the form of text and voice. The input images from the camera will be collected and will be forwarded to image processing where image will be detected and segmentation of the image will be done. The feature of the image will be extracted and the image will be classified with the initial datasets and the precision of the image will be determined. Then the system will predict the input gesture and display the result.

4.2 BLOCK DIAGRAM



The captured images from the camera will be collected and will be forwarded for image detection and segmentation of the image. The feature of the image will be extracted and the image will be classified with the initial datasets and the precision of the image will be determined. Then the system will predict the input gesture and display the result.

V. IMPLEMENTATION

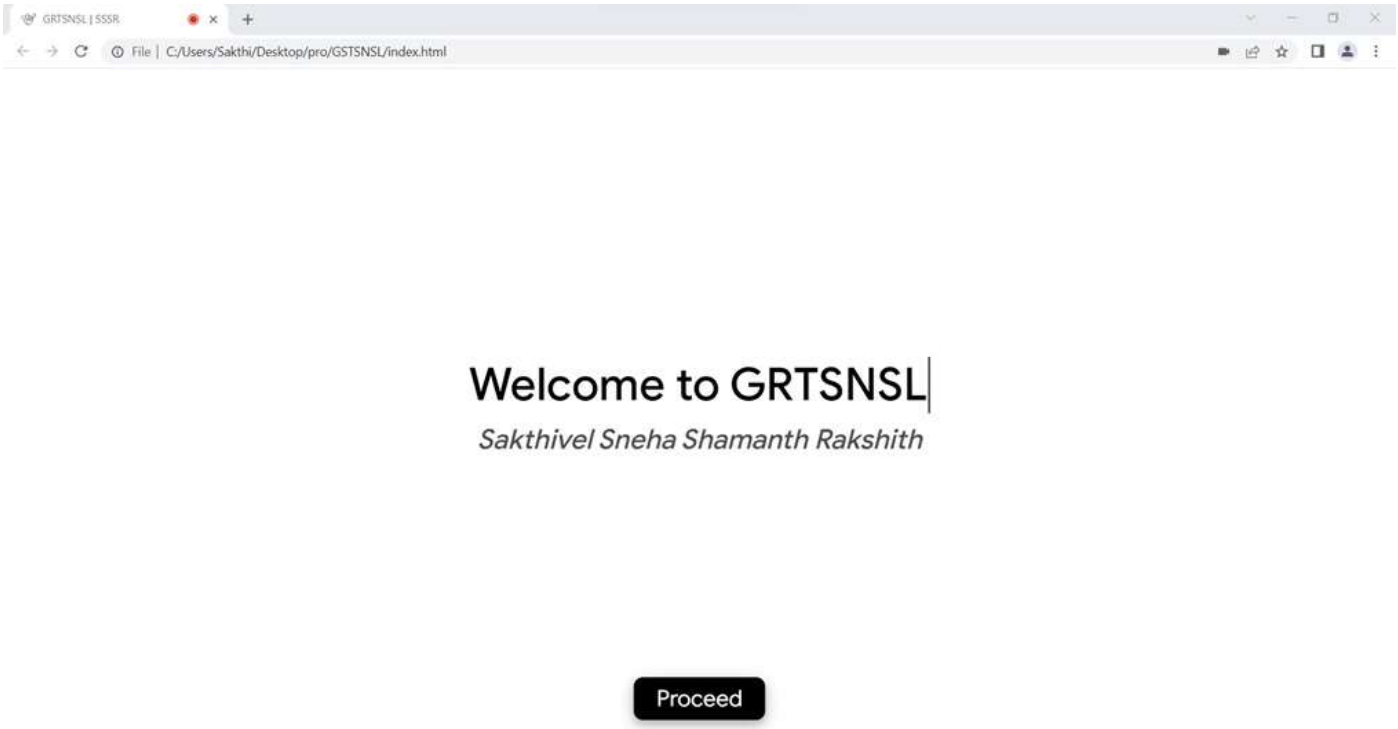


Fig 1:- As the first process of the part of implementation ,the display screen appears as shown.

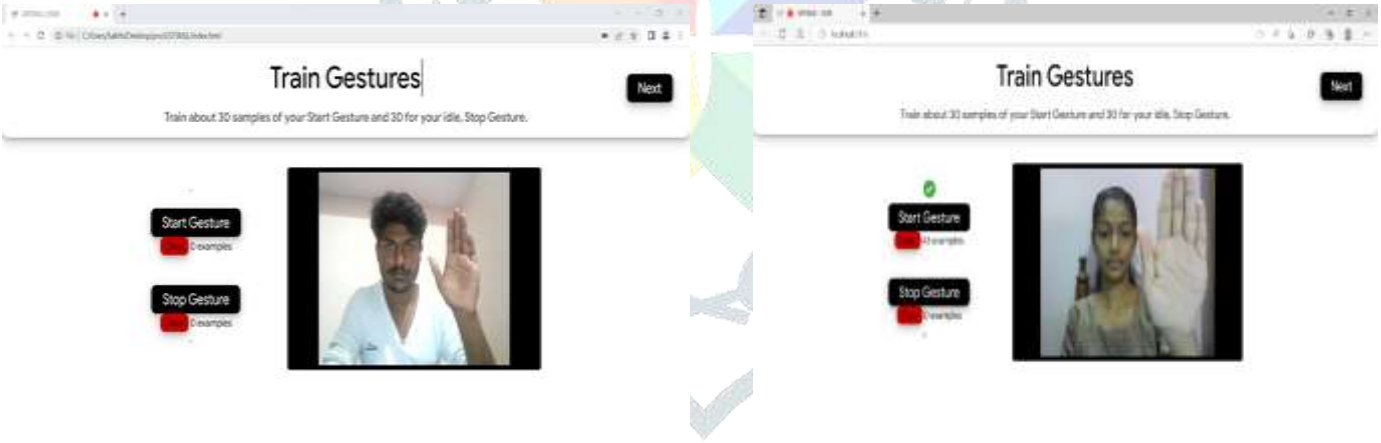


Fig 2:- Next we add Start and Stop gestures as a part of training in order for recognition purpose.

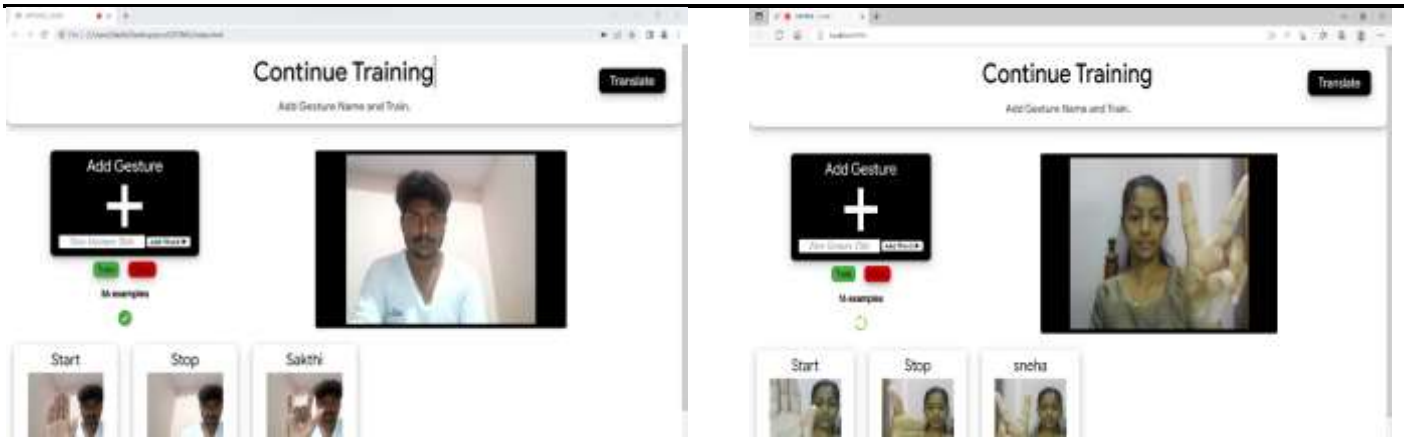


Fig 3:- Next as a part of training, we add new words and capture the gestures simultaneously for translation purpose.

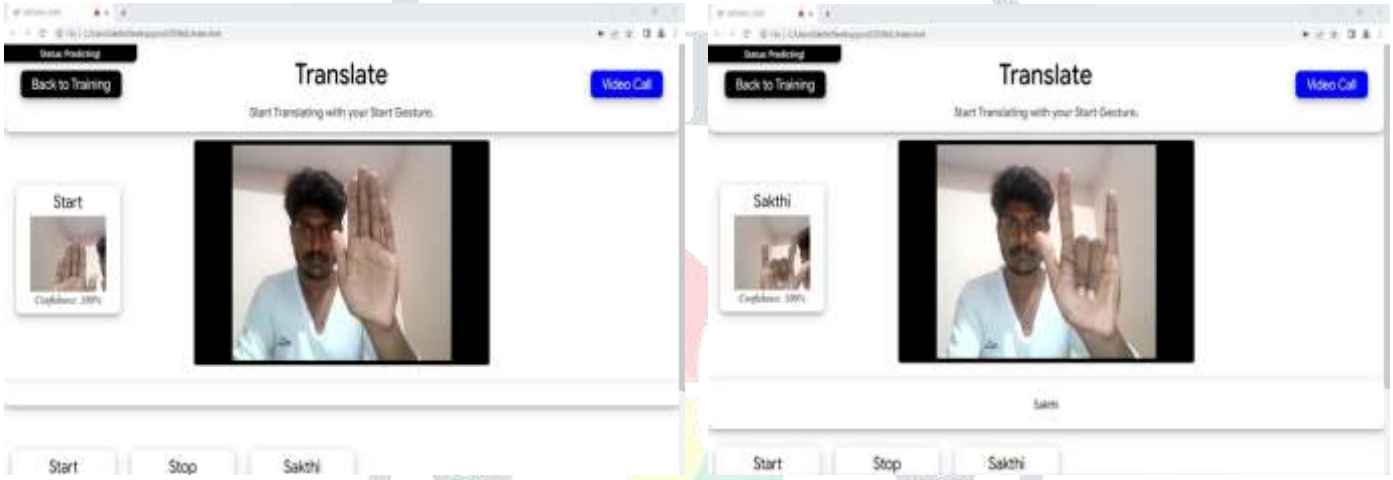


Fig 4:- Lastly we translate the process by showing gestures which recognizes and converts into speech

VI. Conclusion

We presented the concept of gesture-to speech conversion concept, due to which the communication between the vocally impaired people of the society and the common people will be carried out without any obstruction. Interaction is more important for conveying some information. Thus, the proposed system acts as a communication platform for different people. The proposed system can classify the captured images more accurately, and gives relevant output for the given input. This system helps in many to solve the communication problems. Hand gesture detection is fundamental to provide a natural HCI skill. It is now known that in gesture recognition, the most essential aspects are detection, segmentation and tracking. In this project, a system will be created for hand gestures recognition using features extraction and classification in KNN technique. As compared to the other system this concept not only focuses on the gesture to word display but also on the speech synthesis.

VII. Reference

[1] Sakthivel C, Sneha TM, Shamanth Achar, Rakshith A, Prof.Soumyashree Pattar, "Gesticulation Recognition into Text and Speech for Native Sign Language", INTERNATIONAL JOURNAL OF RESEARCH AND ANALYTICAL REVIEWS, Volume:9 Issue:3 Mar ' 2022.

[2] T. Bohra, S. Sompura, K. Parekh and P. Raut, "Real-Time Two Way Communication System for Speech and Hearing Impaired Using Computer Vision and Deep Learning," 2019 International Conference on Smart Systems and Inventive Technology (ICSSIT), Tirunelveli, India, 2019, pp. 734-739, doi: 10.1109/ICSSIT46314.2019.8987908

[3] Singha, Joyeeta & Das, Karen. (2013), "Recognition of Indian Sign Language in Live Video," International Journal of Computer Applications. 70. 10.5120/12174-7306.

[4] H. Muthu Mariappan and V. Gomathi, "Real-Time Recognition of Indian Sign Language," 2019 International Conference on Computational Intelligence in Data Science (ICCIDS), Chennai, India, 2019, pp. 1-6, doi: 10.1109/ICCIDS.2019.8862125.

[5] S. Hayani, M. Benaddy, O. El Meslouhi and M. Kardouchi, "Arab Sign language Recognition with Convolutional Neural Networks," 2019 International Conference of Computer Science and Renewable Energies (ICCSRE), Agadir, Morocco, 2019, pp. 1-4, doi: 10.1109/ICCSRE.2019.8807586.

[6] K. Bantupalli and Y. Xie, "American Sign Language Recognition using Deep Learning and Computer Vision," 2018 IEEE International Conference on Big Data (Big Data), Seattle, WA, USA, 2018, pp. 4896-4899, doi: 10.1109/BigData.2018.8622141.

[7] Kumar Mahesh, "Conversion of Sign Language into Text," International Journal of Applied Engineering Research ISSN 0973-4562 Volume 13, Number 9 (2018) pp. 7154-7161.

[8] Suharjito, Suharjito & Gunawan, Herman & Thiracitta, Narada & Nugroho, Ariadi. (2018). Sign Language Recognition Using Modified Convolutional Neural Network Model. 1-5. 10.1109/INAPR.2018.8627014.

[9] Koller, Oscar & Zargaran, Sepehr & Ney, Hermann & Bowden, Richard. (2016). Deep Sign: Hybrid CNN-HMM for Continuous Sign Language Recognition. 10.5244/C.30.136.

[10] M. Xie and X. Ma, "End-to-End Residual Neural Network with Data Augmentation for Sign Language Recognition," 2019 IEEE 4th Advanced Information Technology, Electronic and Automation Control Conference (IAEAC), Chengdu, China, 2019, pp. 1629-1633, doi: 10.1109/IAEAC47372.2019.8998073

[11] Dynamic Indonesian sign language recognition by using weighted K-Nearest Neighbor https://ieeexplore.ieee.org/document/8265682