

SIGN LANGUAGE RECOGNITION USING CNN

**Mr. Onkar Janjiral¹, Mr. Harshavardhan Kale², Mr. Anand Dhekare³,
Miss. Snehal Birajdar⁴, Miss. Akshata Kulkarni⁵, Mr. S. S Bhuite⁶**

^{1,2,3,4,5}Bachelor of Technology In Computer Science and Engineering, Brahmadevdada Mane Institute of Technology, India.

⁶Project Guide Bachelor of Technology In Computer Science and Engineering, Brahmadevdada Mane Institute of Technology, India.

DOI: <https://www.doi.org/10.58257/IJPREMS34210>

ABSTRACT

Even though using the most natural way of communication is sign language, deaf and mute persons find it challenging to socialize. A language barrier is erected between regular people and Deaf & Mute (D&M) individuals due to the structure of sign language, which is distinct from text. They converse by using vision-based communication as a result. The gestures can be easily understood by others if there is a standard interface that transforms sign language to visible text. As a result, research and development (R&D) has been done on a vision-based interface system that will allow D&M persons to communicate without understanding one another's languages. In this project, first gathered and acquired data and created a dataset, after which extracting useful data from the images. After verification and trained data and model using the CNN algorithm, TensorFlow, and Keras technology, classified the gestures according to alphabet. Using own dataset, this system achieved an accuracy of around 84.20% in an experimental test. system uses the CNN algorithm to process images and data.

Keywords- Convolutional Neural Network (CNN), TensorFlow, Keras, Sign language, Artificial Neural Network (ANN).

1. INTRODUCTION

Deaf and mute people communicate most naturally through sign language, but it has been noted that they have trouble interacting with hearing people.

Sign language is a mode of communication that conveys meaning through visual means such as expressions, hand gestures, and body movements. The use of sign language by those who have trouble hearing or speaking can be very beneficial. Sign language incorporates existing spoken language into hand gestures and hand expressions. A sign language recognition system is made up of a quick, accurate way to translate sign language into written or spoken language. Deaf and mute persons tend to communicate the most naturally using sign language, although it has been noted that they find it challenging to communicate with hearing people. For communication between regular people and D&M individuals, the structure of sign language, which is different from that of standard text, creates a linguistic barrier. They therefore rely on interaction through vision-based communication. If there is a standard interface that transforms sign language to text, the movements can be easily understood by others. In order to enable D&M persons to communicate without understanding one another's languages, research has been done on a vision-based interface system. Because D&M people's only way of communication is through sign language, due to a communication-related handicap, the only way is through sign language. The act of communicating ideas and messages through a variety of techniques, including speech, signals, behavior, and writing, is known as communication. People who are D&M communicate with others by making various gestures with their hands. Gestures are nonverbal communication tools that are interpreted visually. Sign language refers to deaf and dumb people's nonverbal communication.

2. LITERATURE SURVEY

In this research paper, author had used static, manual signs, alphabets, and numerals. ASL lexicon video dataset, sign stream, Weizmann face database, Yale b frontal, digital still camera, and specially designed input devices such as CyberGlove®, Sensor glove, and Polhemus FASTRACK were used. In SL Detection, they achieved an accuracy of around 82.20% [1].

Authors had used technologies such as face detection, removal, and HSV thresholding, and for SL detection, they used the CNN method, with an accuracy of around 93% in experimental tests [2].

The author had used multilayer perceptron neural network. Also, they used to employ two classifiers for the system and a marcel static hand posture database. Raw Features Classifier's neural network accepted 3072 specializations as input. As input, the Histogram Features Classifier was given 512 specialties. They used backpropagation to train the model. system. With all of these combined they achieved around 70 to 85% positive result [3].

A variety of traditional technologies in conjunction with AI and NLP is used by the author in this research paper. With the process of data acquisition, features extraction, gesture classification and verification they also used CNN and ANN technology for the extraction, cross validation and verification. They didn't specify which system they used for training the module. With the help of all this various methods author achieved an accuracy of 91.1% in their experimental test [4].

In this research paper, author had used technologies such as NLP, CNN methodology, and Sift IN to create models in MATLAB. The author also created their own database. They used double layer of algorithm in which they can verify and forecast which symbols are having close connection with each other. With the help of this technique, they were able to forecast all the symbols with the environment having no noise in the background and adequate lighting. They achieved an accuracy of 92% in their experimental tests [5].

Table. 1 Literature Survey

Sr. No.	Paper Studied (Year)	Methodology	Limitations
1	SIGN LANGUAGE RECOGNITION-STATE OF THE ART -Ashok K Sahoo, Gouri Sankar Mishra and Kiran Kumar Ravulakollu (2014)	Used static, manual signs, alphabets and numerals in their project. They used ASL lexicon video dataset, sign stream, Weizmann face database, yale b frontal, digital still camera and specially designed input devices like CyberGlove, Sensor glove and polhemus FASTRACK. They got the accuracy of around 82.20% in SL Detection.	There is no standard dataset for various countries. System is slower.
2	SIGN LANGUAGE RECOGNITION -Anup Kumar, Karun Thankachan and Mevin M. Dominic(2016)	Used technologies like face detection, removal, HSV thresholding for pre-processing and for SL detection they used CNN method and they got accuracy of around 93% in experimental test.	Can not run under robust and unfavourable environment (lot of clutter and poor lighting).
3	SIGN LANGUAGE RECOGNITION -Tülay Karaylian, Özkan KÖLOĞ (2017)	The author have used ANN(Artificial Neural Network) technology and has achieved an accuracy in the range of around 70% to 85%.	The recognition rate can be increased by improving processing an image step.
4	REVIEW OF THE APPLICATION OF ARTIFICIAL INTELLIGENCE IN SIGN LANGUAGE RECOGNITION SYSTEM - Oyeniran, Oluwashina A, Oyeniyi, Joshua O, Sotonwa, Kehinde A, Ojo, Adeolu O.(2020)	The author have used various traditional technologies while using AI and NLP with them and achieved accuracy of 91.1% in their experimental test.	Using different algorithms increased time, also author didn't specified which system is used for training the module
5	A REVIEW PAPER ON SIGN LANGUAGE RECOGNITION FOR THE DEAF AND DUMB -R. Rumana, Reddygari Sandhya Rani, Mrs. R. Prema (2021)	The author have used technologies like NLP, CNN methodology, Sift IN. they have created models using MATLAB and have achieved accuracy of 92% in their experimental test they have conducted.	Tried to include binary threshold and canny edge detection filter but failed to do.

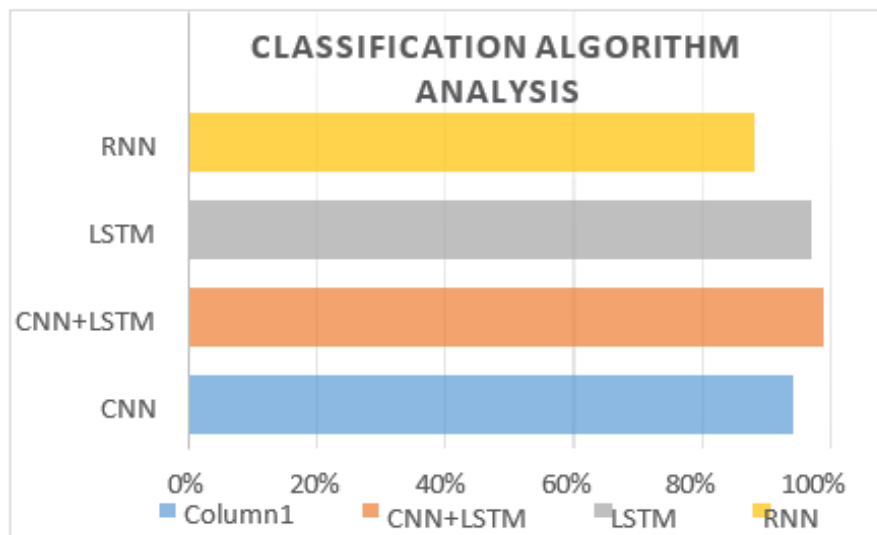


Figure.1 Classification Algorithm analysis

In this project comparing to paper no. [1],[2],[3],[4],[5] we used CNN algorithm for feature extraction and verification and tensor flow, keras for training and detection the language. Achieved the accuracy of around 84.20% with the help of our own created dataset.

Also shown all the research papers algorithm and accuracy in the [Figure. 1]

Proposed system

The important goal of project is to create a model that can recognize alphabet-based hand gestures. Because D&M people's only Due to a communication-related handicap, sign language is the only means of communication available to them. Communication is the act of exchanging thoughts and messages using a variety of behaviors, signs, and written language. People who are D&M interact with others by making various hand signals. Nonverbal cues like gestures can be deciphered by visual perception. Sign language is the nonverbal communication of the deaf and dumb. System architecture consist of 4 stages which is shown in [Figure. 2]

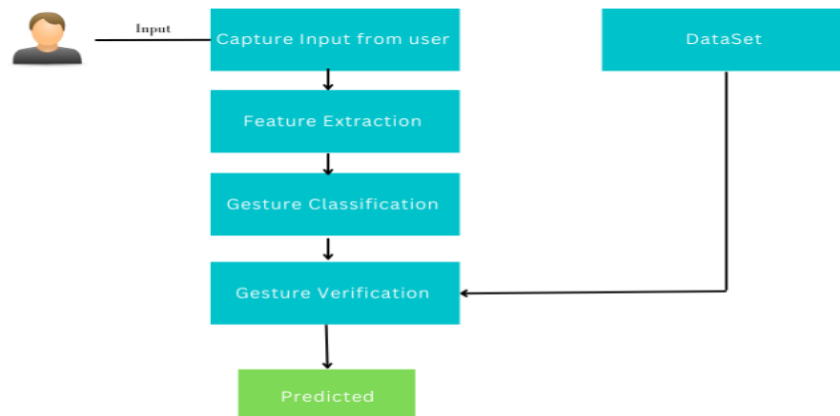


Figure. 2 Sign Language Recognition Architecture

Data Acquisition

During the data acquisition stage, observing hand information using a computer camera as an input device. captured around 4500-5000 pictures for all the static alphabets. Because of this data set our sign recognition accuracy has significantly improved. In this stage, created a dataset of alphabets from A to Z using the Python OpenCV library. [Figure. 3]

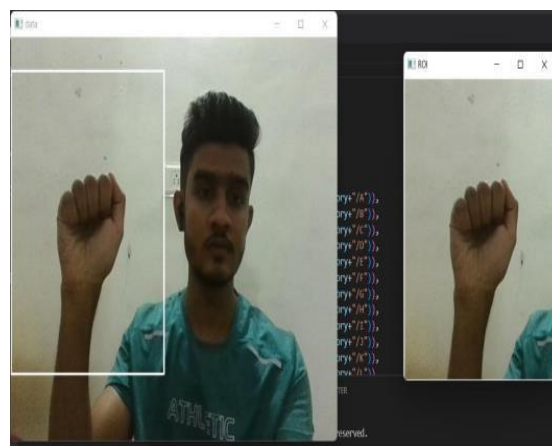


Figure. 3 Data acquisition

Feature Extraction

Following the creation of a dataset, extract the landmark from the hand in order to recognize that alphabet. Used the (CNN) algorithm for extraction.

Following the discovery of a landmark by hand Feature extraction plots, shown in [Figure. 4] the hand and then passes it to (CNN). The (CNN) algorithm is used to train the dataset in Gesture Classification.

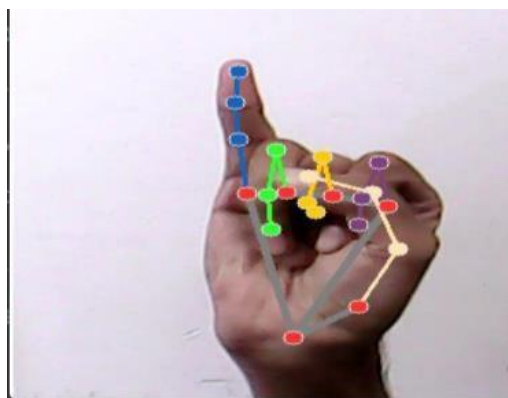


Figure. 4 Feature Extraction

Convolutional neural network (CNN)

Convolutional neural network has three layers in general. And, with the help of a classifier example, we comprehend each layer one by one. It can classify an image of an X and O. So, using the case, we will comprehend all four layers. Shown in [Figure. 5] (CNN) has feedback connections as compared with conventional cascaded neural networks. This type of recurrent neural network (RNN) can process not only but particular data features (such as images), but also entire data flow (such as speech or video).

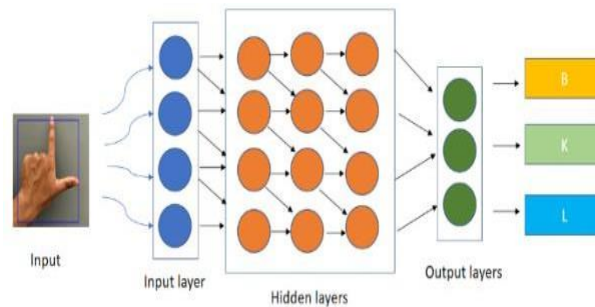


Figure. 5 Gesture Classification

Verification

Using the dataset created, validated the camera input and hand plotting. Print the output and accuracy based on AML after verification (American Sign Language).

To improve the accuracy of model, all of these steps necessitate the use of additional technologies.

The following are some of the technologies used in this project

- **Tensor Flow:**Tensor Flow is a cost free and available for all numerical computation library. After defining the nodes of the computation graph, the 13 actual computations occur within a session. Tensor Flow is a well-known machine-learning framework.
- **Keras:** Keras is a superior level neural network library written in Python that serves as a wrapper for TensorFlow. It is used when need to quickly construct and test a neural network using only a few lines of code.
- **OpenCV (Open-Source Computer Vision):** Real time Machine vision programming functions in an open-source library is called OpenCV (Open Source Computer Vision). It is largely practiced for face and object identification, image processing, and video collection and analysis.
- **Convolutional neural network (CNN)** A Convolutional neural network has three layers in general. And, with the help of a classifier example, we comprehend each layer one by one. It is capable of classifying an image of an X and O. So, using the case, we will comprehend all four layers.
- **Convolutional Neural Networks have the following layers:**
 - Convolutional
 - ReLU Layer
 - Pooling
 - Fully Connected Layer

Convolutional Layer

Convolutional layers are included in the Keras deep learning library

By looking at several worked instances using fictitious data and constructed filters, we can have a better understanding of the convolution procedure.

In this part, we'll look at both a one- dimensional and a two-dimensions convolutional layer example to make the convolution operation explicit and to show a working example of using the Keras layers.

$$Z^1 = h^{1-1} * W^1 \dots \text{Eq. 1}$$

Where,

Z^1 = output layer

h^{1-1} = output of previous layer

W^1 = weight matrix associate with previous layer and current layer

(h = height, w= width)

ReLU Layer

We replace any negative values in the filtered image with zeros in this layer. When the node input surpasses a certain threshold, this function is activated. As a result, the output is zero when the input is less than zero. When the input exceeds a certain threshold, it has a linear relationship with the dependent variable. This means it can accelerate the speed of a training data set in a deep neural network faster than other activation functions while avoiding summing with zero

$$ReLU(Z_i) = \max(0, Z_i) \dots \text{Eq. 2}$$

Where,

i = input to the ReLU function

Pooling Layer

In CNN, pooling layers are often utilized alongside convolutional layers. The spatial dimensions of the feature maps are reduced with each pooling layer, while the convolutional layers capture increasingly complicated properties from the input. The generated feature maps are then passed to a fully connected layer, which performs the final classification or regression task.

In this project, Average pooling layer the average dimensions of our uploaded dataset and consequently, produces a feature map of dataset to push forward in fully connected layer.

Fully Connected Layer

Neural networks are made up of non-linear functions that are linked together. A neuron (maybe a perceptron) is responsible for each function. In fully connected layers, the neuron applies a linear change to the input vector via a weight's matrix. The product is then transformed nonlinearly using a nonlinear activation function f .

$$y_{jk}(x) = f \sum_{i=1}^{n_H} W_{jk} X_i + W_{j0} \dots \text{Eq. 4}$$

$i=1$

Where,

x = input vector

W = weight matrix W_0 = bias term

Every potential layer-to-layer link is present, which means that every input in the input vector has an effect every output of the output vector. However, not all weights affect all outputs in the same way. Examine the lines in the diagram above that connect each node. The orange lines represent the first neuron (maybe perceptron) in the layer. The weights of this neuron only affect output A and have no effect on outputs B, C, or D

SoftMax

The softmax function takes a vector of K real values and turns it to a vector of K real values that add up to one. The softmax converts input values that are positive or negative, zero or greater than one, into values ranging from 0 to 1, permitting them to be comprehended as probabilities. If an input is tiny or negative, the softmax converts it to a tiny probability; if an intake is enormous, it shifts it to a large probability but keeps it between 0 and 1

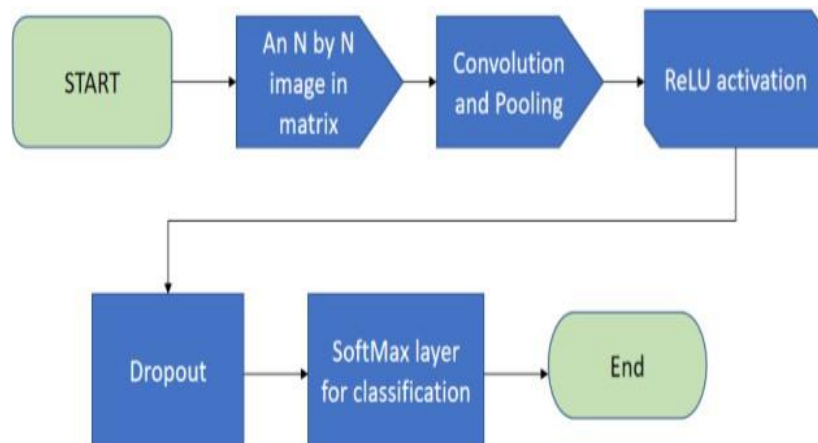


Figure .6 CNN Layers

Artificial neural network

Artificial Neural Networks (ANN) are cerebrum-inspired algorithms that are used to model intricate structures and anticipate problems. The Artificial Neural Network (ANN) is a deep learning method derived from the concept of Biological Neural Networks in the human brain. ANN was created because of an attempt to imitate the workings of the human brain. Although not identical, the workings of ANN are extremely reminiscent of biological neural networks. The ANN algorithm only accepts numeric and structured information

In this project, after applying the CNN layers, the generated output is fed into the ANN, which attempts to detect a pattern or relationship between them before generating the output

Dataset generation

In order to compare the photographs taken while using this technology for communication, a proper database of sign language motions must be created. Below are the steps followed to produce the data set used.

To create dataset, used an open computer vision library. For testing purposes, first took roughly 4500–5000 photos of every alphabet out of which used approximately 4000 images and 180–200 images of each letter. First, record every frame that the webcam on computer produces. As seen in the image below, designate a space of interest in each frame, which is indicated by a blue outlined box

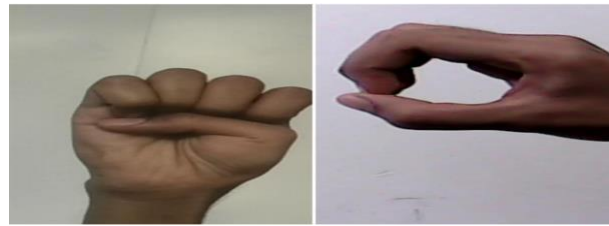


Figure. 7 Dataset Generation

3. RESULT AND ANALYSIS

This study has built an effective USA Sign Language detection based on eyesight in real time system for D & M users. By employing two algorithms layers that more symbols to validate and forecast similar to one another, can improve forecasting. With this technique, can virtually always identify symbols as long as they are correctly displayed, there is no background noise, and the lighting is adequate.

The system was tested on a 2000 image dataset and gave an accuracy of around 81%, which is approximately 1620 images matched with the input.

During the testing phase, a fresh dataset of over 2000 photos was updated, and with the previous dataset included, the system delivered around 84.20% accuracy, which is roughly 3360 images matched with input [Figure.7].

Successfully developed a model with an accuracy of about 84.20% as shown in testing phase using the CNN methodology, as well as several image categorization and verification method

Comparing to [1] they used manual signs and pre developed dataset for the experimental test and they achieved around 82.20% accuracy whereas compared to [2] they have used face detection, removal and hsv thresholding and for the SL they used CNN algorithm and have achieved accuracy of 93%. In the [3] the author's have used artificial neural network technology and have output of around 70 to 85%. In the paper[4] the used various contemporary technologies with the help of AI and NLP and resulted 91.1% accuracy.[5] In this they used NLP,CNN Methodology and siftin to create model in MATLAB after that they got 92% positive result

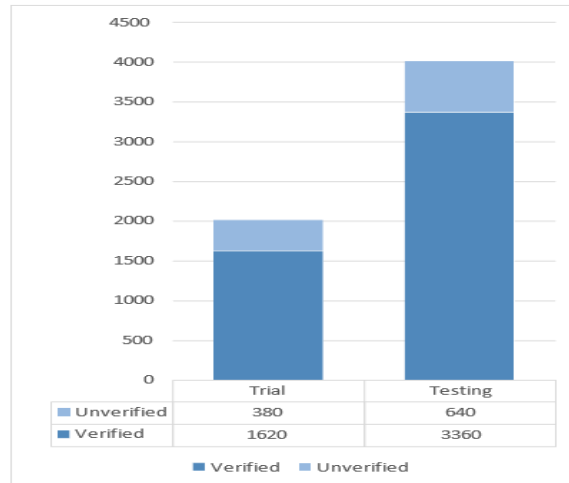


Figure. 8 Comparative Analysis – Training and Test Data

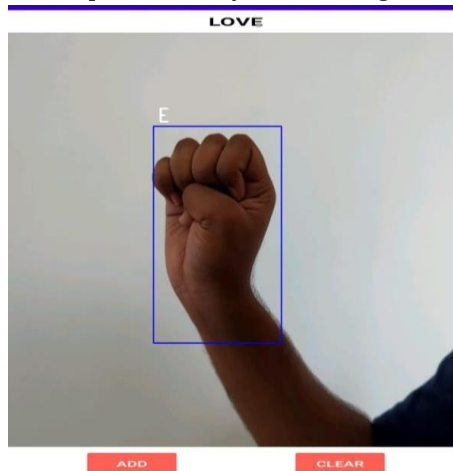


Figure. 9 Concluding Output

4. CONCLUSION

In this project at the beginning captured around 4000-4500 images for the dataset. After that used (CNN) algorithm for feature extraction where points on the hands are cited to detect the hand symbols. after that gesture classification step occurs where (CNN) is also used and the classification between different various symbols happens

In last step of the sign detection verification step takes place where system have used tensor flow and keras for training the model for detection the symbols. Also, these two technologies have also used for cross validation and verification

Successfully developed a model with an accuracy of about 84.20% in experimental test using the (CNN) (Convolutional neural network) methodology, as well as several image categorization and verification method Using the (CNN) algorithm, can incorporate handwriting recognition and text-to-speech conversion in future updates. This algorithm also allows us to use video capturing technology to create dynamic signs. Also intend to translate the results into local languages.

5. REFERENCES

- [1] Ashok K Sahoo, Gouri Sankar Mishra and Kiran Kumar Ravulakollu (2014)- Sign language recognition-State of the art
- [2] Anup Kumar, Karun Thankachan and Mevin
- [3] M. Dominic(2016)- Sign Language Recognition
- [4] Tülay Karaylian, Özkan KÖİÖÇ (2017)- Sign Language Recognition
- [5] Oyeniran, Oluwashina A, Oyeniyi, Joshua O, Sotonwa, Kehinde A,Ojo, Adeolu O.(2020)- Review Of The Application Of Artificial Intelligence In Sign Language Recognition System
- [6] R Rumana, Reddygari Sandhya Rani, Mrs.
- [7] R. Prema (2021)- A Review Paper on Sign Language Recognition for The Deaf and Dumb
- [8] V. N. T. Truong, C. K. Yang and Q. V. Tran (2016)- A Translator for American Sign Language To Text And Speech
- [9] M. M. Islam, S. Siddiqua and J. Afnan (2017)- Real Time Hand Gesture Recognition Using Different Algorithms Based On American Sign Language
- [10] I. C. Ani, A. D. Rosli, R. Baharudin, M. H. Abbas and M. F. Abdullah(2014)
- [11] Preliminary Study Of Recognising Alphabet Letter Via Hand Gesture
- [12] Rajshri Bhandra & Subhajit Kar- Sign Language Detection from Hand Gesture Images using Deep Multi-layered Convolution Neural Network. (2021)
- [13] Vishwa Hariharan Iyer, U.M Prakash, Aashrut Vijay, P.Sathishkumar – Sign Language Detection using Action Recognition(2022)
- [14] Tengfei Li Yongmeng Yan; Wenqing Du - Sign Language Recognition Based on Computer Vision(2022)
- [15] Tengfei Li Yongmeng Yan; Wenqing Du - Sign Language Recognition Based on Computer Vision(2022)
- [16] Rahib Abiyev; John Bush Idoko; Murat Arslan - Reconstruction of Convolutional Neural Network for Sign Language Recognition(2020)
- [17] Priyanka Pankajakshan; Thilagavathi B-Sign language recognition system(2015)
- [18] Nipun Jindal; Nilesh Yadav; Nishant Nirvan; Dinesh Kumar- Sign Language Detection using Convolutional Neural Network (CNN),(2022)