International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 07 Issue: 04 | April - 2023 Impact Factor: 8.176 ISSN: 2582-3930

Sign Language Recognition System

Trupti Deshmukh[1], Arya Hariharan Nair[2], Eshwariya Sapkal[3] ,Mehnaz Uluwar[4] ,Shubhangi Chavan[5]

[1-5]Department of Computer Engineering & Pillai College of Engineering(Autonomous), New Panvel, India

---

Abstract - Sign language is a visual language that conveys meaning through a combination of hand gestures, body language and facial expressions. Sign languages are used by people who are deaf or mute or are hearing impaired, to communicate with each other, as well as with hearing people who understand the sign language. To facilitate and aid the communication between a wide range of hearing, hard of hearing, or deaf individuals and those who don't understand sign language, we propose a ML based sign language recognition system that uses feature extraction to identify patterns and make predictions about the user's hand characteristics captured through the webcam. We have proposed a system that utilises the convolutional neural network to recognize Fingerspelling gestures of the American Sign Language. We use Tensorflow and Keras to build and train the CNN model for gesture recognition and classification. The recognised alphabets can then be combined in order to form a coherent word. NLP techniques are applied to suggest words from the detected alphabets. The words can further be appended together to form a sentence.

Key Words: Deaf and Hard of Hearing Community(DHH), Convolutional Neural Network (CNN) , Sign Language

1. INTRODUCTION

According to the World Health Organization (WHO), as of 2021, approximately 466 million people worldwide have disabling hearing loss, which corresponds to about 6.1% of the world's population. A primary means of communication for the DHH community is using sign languages. Sign languages are fully developed and established natural languages that are structurally different from the spoken languages. People who are unable to hear or speak, can communicate by visually reading the hand gestures during signing. They are essential for promoting inclusivity and ensuring that all people, regardless of their hearing ability, have access to communication, information, and services.

Sign languages are distinct from spoken languages, and each sign language has its own grammar and vocabulary. For example, American Sign Language (ASL) and British Sign Language(BSL) are both sign languages that vary from each other in respect to vocabulary and gestures used, the former is widely used in the United States and some parts of Canada, while the latter is used by the people of the United Kingdom. Other sign languages include Auslan (Australian Sign Language), LSF (French Sign Language), and Japanese Sign Language (JSL).

Sign languages are not universal and are not based on spoken languages. Instead, they are natural languages that have evolved over time within deaf communities. They are not simply a system of manual gestures; they are complete languages that have their own vocabulary, grammar and syntax.

Sign languages face several challenges, many of which are related to their recognition as legitimate languages and the provision of accessible services to deaf communities. Some of the challenges are:

1. Sign languages are often not recognized as official languages, which limits access to education, employment, and other opportunities for deaf individuals.
2. Many deaf children do not have access to quality education in sign language, which can limit their academic and career opportunities.
3. Deaf individuals may have limited access to healthcare services due to communication barriers with healthcare providers who do not know sign language.
4. Many public spaces and services, such as government offices and transportation, do not provide adequate sign language interpretation or other accommodations for deaf individuals.
5. Sign language research and development are still relatively underfunded and underdeveloped, which limits the availability and quality of sign language technologies and resources.

It is evident that sign languages play an important role in the communication and social lives of millions of people around the world . Yet there are many barriers to realising fair and

equal opportunities for all irrespective of any disabilities. Our goal is to create a project that can be substantive to the community and be of some assistance in making sign languages a little more accessible, and aid sign language users to communicate efficiently with others even in the absence of an interpreter.

2. LITERATURE SURVEY

In this section, we have cited some of the previous literature that use the various image recognition or classification techniques, and are relevant to our project . Most researchers have employed the Convolution Neural Network for recognition and classification mainly due to the several advantages it holds over other Machine Learning models.

A. Sunitha Nandhini,et.al. [9] have proposed a system that can predict some common English words using CNN. More informative features from the images were finely extracted and utilised for classification and training. The system can recognize 125 words, including alphabets, using a total of 1750 static images for each sign.

Mehreen Hurroo,et.al. [7] have proposed a system that predicts input gesture of user and displays result. In this model, HSV is used to identify and remove the user input backgrounds, perform morphological operations, and apply masks. The images are resized for classification and the hand gestures are segmented.

Pathak, et.al. [10] have addressed the development and implementation of a model for sign language recognition. The model was based on a convolutional neural network (CNN) and a pre-trained SSD Mobile net v2 model for sign recognition. The Mobile Net SSD model is a single-shot multibox detection (SSD) network and it was used to conduct object detection.

Rachana Patil, et.al. [4] have introduced an Indian Sign Language recognition system. Their aim was that the system should be able to predict and display the name of the captured image, and the user must be able to use a web camera to take pictures of hand gestures. The captured image is processed through a series of steps that include computer vision techniques like dilation, mask operation, and changing the image to grayscale.

Shagun Katoch, et.al. [6] have proposed a real-time recognition utility that utilises a unique data set, ensuring that the system is rotation-invariant, and resolves the background dependence issue. The system is trained on all 36 ISL static alphabets and digits.

Suharjito, et.al. [14] have proposed a Video based recognition performed on the basis of action rather than shape. The implementation of the i3D inception model for Sign Language Recognition via transfer learning is the objective of this paper. Based on the test, they were able to train with 10 words and 10 signers and 100 classes at a 100% accuracy, but the validation accuracy was pretty low. The model encountered the problem of overfitting.

Sung,et.al. [15] have introduced a real-time hand gesture recognition (HGR) system, which identifies a bunch of predefined static motions through a RGB camera, implemented using the open source MediaPipe framework. A hand skeleton tracker improved from MediaPipe Hands and a gesture classifier was used for their model.

Wanbo Li, et.al. [16] have proposed an improved sign language recognition system that combines a convolutional neural network (CNN) and a long short-term memory (LSTM) neural network. The graphical user interface (GUI) for the proposed system was developed with PyQt.

3. EXISTING SYSTEM ARCHITECTURE

The existing system architecture takes an input image of the hand figure and further processes it by skin Filtering and Hand cropping, which helps in focusing on only the necessary data. It is further converted into Binary Image and Feature Extractions are applied.Further It was trained using the LSTM model.

4. PROPOSED SYSTEM ARCHITECTURE

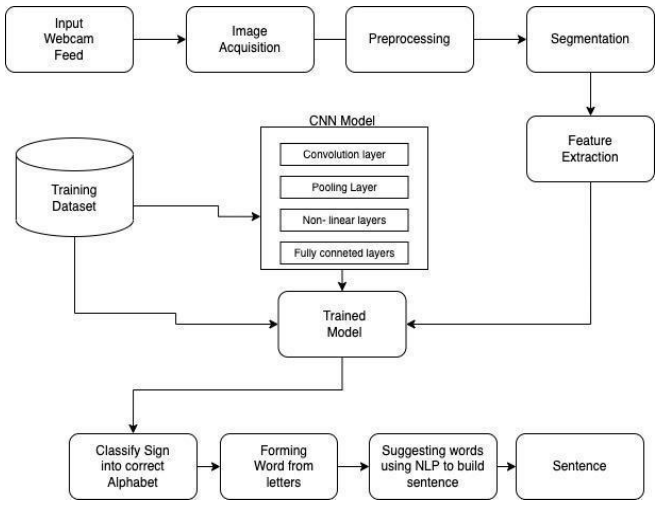

Fig 1: Proposed System Architecture

5. DATASET USED

An image dataset containing separate classes for each letter was used to train the model. Each alphabet class includes a set of 1750 masked images in the training set and 250 images in the test set.

Fig 2: Sample Dataset showing each alphabet

6. IMPLEMENTATION PROCESS

A) Image acquisition

The hand gestures are captured by the webcam.We use a bounded square to indicate the Region of Interest; the user input is captured from the ROI and a gaussian blur filter is applied to the image in order to extract the various features of the image.

B) Data preprocessing

Skin colour and its tone heavily depend upon the lighting conditions. As a result, the output we obtain after the segmentation is not quite remarkable. For this reason, our input images (RGB) are converted into grayscale and a gaussian blur filter is applied upon it to help remove the irrelevant noise that may be present. We then perform an adaptive thresholding to accurately extract the hand from its background. The images are further resized to 64 x 64 pixels.

C) Feature extraction and representation

Images are often represented as 3D matrices in image processing and computer vision approaches. In these matrices, the value in the matrix corresponds to a pixel in the image. The three dimensions of the matrix represent the height, width, and colour channels of the image.

The 3D matrix representation is useful because it allows for the efficient processing of images using convolutions(in CNN), which are essentially a series of matrix operations, to extract features from images. For grayscale images, the matrix has a height and width corresponding to the size of the image, and a single colour channel. Each element in the matrix corresponds to the grayscale intensity of the pixel at that location.For RGB images, the matrix has a height and width corresponding to the size of the image, and three colour channels (red, green, and blue). Each element in the matrix corresponds to the intensity of the corresponding colour channel of the pixel at that location.

D) Model creation and training

We create a CNN model with the required layers and assign some weight to each layer. The model is first trained on the images from the dataset and the cross entropy of the model is calculated. We then focus on optimising the model by minimising the cross entropy function. This is done by updating the weights by backpropagation. The trained model is then further used for classification of real time data.

E) Gesture classification

In this step, input images are fed to the model after performing the necessary preprocessing steps, for training and testing. The prediction layer estimates how likely the image will fall under one of the classes.

F) Text conversion:

The detected label for the input gesture captured through the camera is then converted into text.

G) Forming words and sentences:

Each predicted alphabet can be appended to form a word. NLP is used to suggest possible words from the detected letter. Words can be joined in a string to form sentences..

7. ALGORITHM

Gesture classification Algorithm

The algorithm followed by our model to perform the label prediction can be explained in two layers:

Layer 1:

1. Application of the gaussian blur filter and adaptive threshold to the user input frame captured using OpenCV for feature extraction.
2. Passing the processed image from the previous step to the trained CNN model for prediction. If an alphabet is detected for at least 50 frames, it is printed into the

stack and taken into consideration for forming the word.

3. Space between the words can be conveyed using the space gesture from the dataset.

Layer 2:

1. Alphabet is recognised and printed into the stack after being detected for a threshold of 30 seconds
2. Letters in stack are appended into a sentence when a legible word forms.
3. Based on the letters recognised, suggestions for probable words are made which can be appended into the sentence by clicking on it.

8. USE CASE DIAGRAM

Use case diagram is used to describe dynamic behaviour of a system. It depicts how a system interacts with external actors. A stick figure with the name of the actor is used to represent an actor, and a circle with the use case's name inside it represents a use case. The use cases are used to categorize and divide system functionality for better analysis of the project. Actors portray the parts taken on by users of the system.

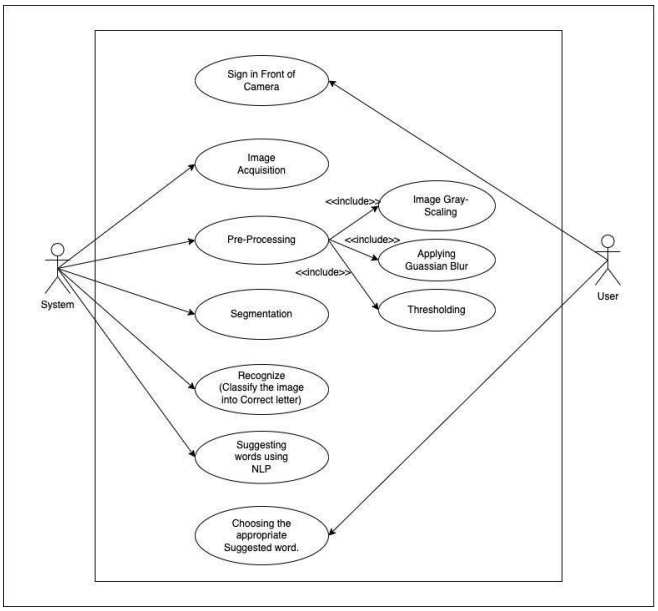

Fig -3: Use Case Diagram

9. PERFORMANCE EVALUATION

The proposed Sign Language Recognition System using CNN was analysed using the various evaluation methods mentioned before. The result of the analysis is as follows:

Evaluation Metrics	Model evaluation results
Accuracy	0.9785
Precision	0.96
Recall	0.96
F1 Score	0.96

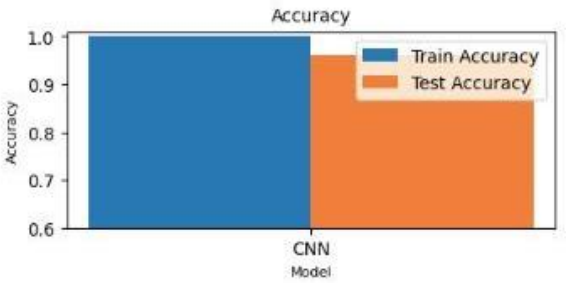

Fig -4: Accuracy of learning model used

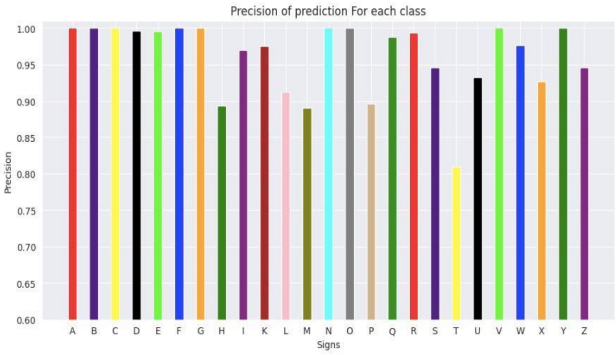

Fig -5: Precision of prediction for each letter of the ASL

10. RESULT AND DISCUSSION

The experimental results of the proposed system are provided in this given section. This presents the graphical user interface of the application, depicting all the provided features.

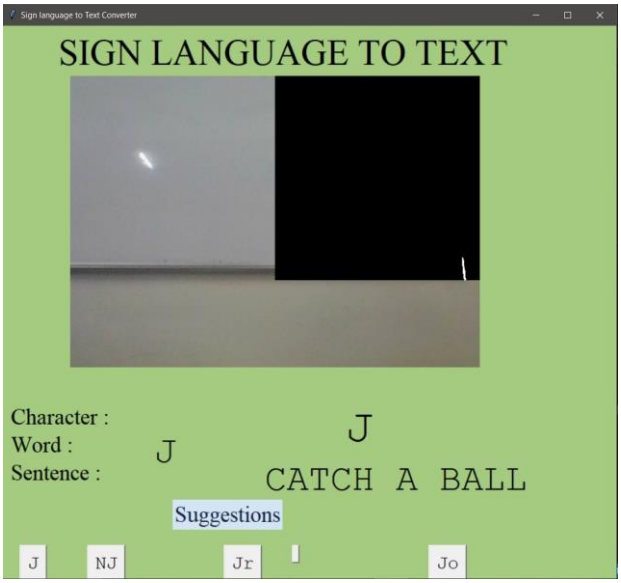

Fig -6: User Interface of the application

11. CONCLUSIONS

The main purpose of the sign language detection system is to provide a viable tool for communication between the Deaf and Hard of Hearing community, phonics and those unable to use sign language. Any webcam or device built in camera can be used to access the proposed system for performing gesture detection and recognition. Given that CNN requires significantly less preprocessing than other classification algorithms, we have chosen to employ it for building our model. CNN can learn these filters/characteristics with adequate training. CNN model also exhibits higher accuracy than other image classification models. CNN's ultimate aim is to make image processing effortless by extracting all the relevant characteristics from the image while preserving crucial information that is essential for making accurate predictions. The goal of the project is to achieve good accuracy for the sign recognised through the camera and the respective mapped output.

REFERENCES

[1] Abraham, Ebey et al. "Real-Time Translation of Indian Sign Language using LSTM." *2019 Global Conference for Advancement in Technology (GCAT)* (2019): 1-5.

[2] Alsaffar, Ahmed & Tao, Hai & Talab, Mohammed. (2017). Review of deep convolution neural network in image classification. 26-31. 10.1109/ICRAMET.2017.8253139.

[3] Chandandeep Kaur, Nivit Gill, An Automated System for Indian Sign Language Recognition, International Journal of Advanced Research in Computer Science and Software Engineering.

[4] Indian Sign Language Recognition using Convolutional Neural Network
Rachana Patil, Vivek Patil, Abhishek Bahuguna, Gaurav Datkhile
ITM Web Conf. 40 03004 (2021)
DOI: 10.1051/itmconf/2021400300

[5] JOYEETA SINGH, K. D. Indian sign language recognition using eigen value weighted euclidean distance based classification technique. International Journal of Advanced Computer Science and Applications 4, 2 (2013)

[6] Katoch, Shagun & Singh, Varsha & Tiwary, Uma Shanker. (2022). Indian Sign Language recognition system using SURF with SVM and CNN. Array. 14. 100141. 10.1016/j.array.2022.100141.

[7] Mehreen Hurroo , Mohammad Elham, 2020, Sign Language Recognition System using Convolutional Neural Network and Computer Vision, INTERNATIONAL JOURNAL OF ENGINEERING RESEARCH & TECHNOLOGY (IJERT) Volume 09, Issue 12 (December 2020)

[8] Muthukumar, K. A. et al. "Vision Based Hand Gesture Recognition for Indian Sign Languages Using Local Binary Patterns with Support Vector Machine Classifier." (2017).

[9] Nandhini, A & Roopan, D & Shiyaam, S & Yogesh, S. (2021). Sign Language Recognition Using Convolutional Neural Network. Journal of Physics: Conference Series. 1916. 012091. 10.1088/1742-6596/1916/1/012091.

[10] Pathak, Aman & Kumar|priyam|priyanshu, Avinash & Chugh, Gupta|gunjan & Ijmtst, Editor. (2022). Real Time Sign Language Detection. International Journal for Modern Trends in Science and Technology. 8. 32-37. 10.46501/IJMTST0801006.

[11] Rajalingam, B. et al. "A Smart System for Sign Language Recognition using Machine Learning Models." *2022 4th International Conference on Advances in Computing, Communication Control and Networking (ICAC3N)* (2022): 1125-1131.

[12] Sagar P.More, Prof. Abdul Sattar, Hand gesture recognition system for dumb people

[13] S. Shirbhate1, Mr. Vedant D. Shinde2, Ms. Sanam A. Metkari3, Ms. Pooja U. Borkar4, Ms. Mayuri A. Khandge/Sign-Language-Recognition-System.2020 IRJET Vol3 March,2020.

[14] Suharjito, H. Gunawan, N. Thiracitta and A. Nugroho, "Sign Language Recognition Using Modified Convolutional Neural Network Model," 2018 Indonesian Association for Pattern Recognition International Conference (INAPR), Jakarta, Indonesia, 2018, pp. 1-5, doi: 10.1109/INAPR.2018.8627014.