

Indian Sign Language Recognition System

Rupali Kadwade
Government College of
Engineering, Karad
Kolhapur, India

Akanksha Tangade
Government College of
Engineering, Karad
Kolhapur, India

Neha Pakhare
Government College of
Engineering, Karad
Kolhapur, India

Samiksha Kolhe
Government College of
Engineering, Karad
Kolhapur, India

Hajara Waikar
Government College of
Engineering, Karad
Kolhapur, India

S. J. Wagh
Government College of
Engineering, Karad
Kolhapur, India

Abstract—A 2019 poll found that 2.4 million people in India are deaf and mute, making about 20% of all deaf and mute people worldwide. They are unable to converse as easily as other folks. The usage of sign language is being employed to solve this difficulty. People can express themselves through hand gestures in sign language. Each nation has its own sign language that has evolved. Indian Sign Language is the name of the sign language that India has created. Language via Sign Recognition picks up on the hand signals and keeps learning until the correct text or speech is produced. There are two types of hand gestures: static and dynamic. In contrast to dynamic hand gestures, static hand motions are simpler to identify. To recognize hand motions, we can design a CNN (Convolutional Neural Network) architecture using computer vision. The vanishing gradient issue can also be solved using GRU (Gated Recurrent Unit) and LSTM (Long Short Term Memory). With these techniques, we can attain an accuracy of about 97%. others who have trouble speaking, hearing, or seeing will find this suggested remedy very helpful in establishing clear communication both inside themselves and with non-impaired others.

Keywords - Hand gestures, Indian Sign Language, Computer Vision, CNN, LSTM, and GRU.

I. INTRODUCTION

As is well known, there are several sign languages used in various nations, including American, Indian, and others. Body language, a kind of nonverbal communication, should not be confused with sign language. Around 2.4 million members of the Deaf and Hard of Hearing population (DHH) live in India, the second-largest nation in the world.[1] Speaking with

those who have hearing loss is really difficult. Since Deaf and Mute persons utilize hand gestures to communicate, normal people have difficulty understanding their language from the signals they make. Systems that can identify various indications and provide information to common people are thus necessary.[2]

A group of established languages known as sign languages employ a visual-manual modality to communicate.[3] The goal of this project is to create a system that can reliably translate Indian sign language into numerals so that those who are less privileged can speak with others in public settings like banks and train stations without the need of an interpreter.[4] We used gestures in the sign language recognition system so that the gesture model could be observed and provide output. The following hand motions are used for numbers and alphabets.



Figure 1 Gestures for numbers

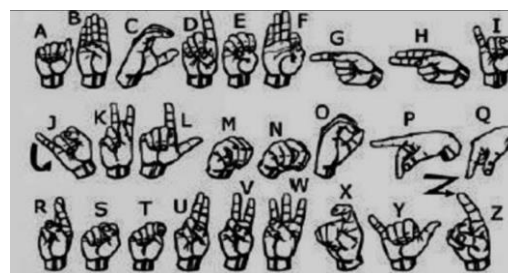


Figure 2 Gestures for alphabets

II. RELATED WORK

From the perspective of Satwik Ram Kodandaram, N. Pavan Kumar, Sunil GI [5](2021). Convolutional Neural Network model will be more effective to identify the pictures of hand motions across an epoch. Following the model's successful recognition of the gesture, English text is created that may later be translated into voice.

Madhuri Sharma, Ranjna Pal and Ashok Kumar Sahoo [6] (2014). The goal of this project is to create a system that can accurately translate Indian sign language into numerals, enabling the less fortunate to interact with others in public settings like banks, train stations, and other establishments without the assistance of an interpreter.

Another piece of work notable in this area is done by Juhi Ekbote, Mahasweta Joshi [7] (2017). They are working to create a system that can automatically recognise the numbers from 0 to 9 in Indian Sign Language.

Similarly, In [8] (2022). Deep Kothadiya, Chintan Bhatt, Krenil Sapariya, Kevin Patel, said the solution whose goal is to use deep learning-based model that detects and recognizes the words from a person's gestures.

Furthermore, several robust models have been suggested in order to address the challenges in Sign Language Detection. [9] (2022) Aman Pathak, Avinash Kumar, Priyam, Priyanshu Gupta, Gunjan Chugh propose the architecture trained on our own dataset. They had created and implemented of sign language recognition model based on a Convolutional Neural Network (CNN).

In our system, we have three files such as

- Data generation
- Application
- Model Training

In this file video streaming data for static and dynamic gesture is collected with the help of local camera. So we have collected separate datasets or folders for numbers, alphabets, and words. Also, during data collection, we have used two helper functions that are `cal_accum_avg` and `segment_hand`, `cal_accum_avg` applies background subtraction techniques to improve the resolution of the image and `segment_hand` is used to segment hand region found in frame of image which helps to make image processing faster.

In this file the front end of the application is built using the popular library of python that is tkinter. It gives us the ability to create GUI programs in Python and offers standard GUI components that can be used to create interfaces. These components include menus, entry fields of different types, display spaces, and buttons.

For the application usage we have provided three option live prediction, load image and classify the loaded image. Through live prediction, we provide real-time gesture detection if the user wants to make the prediction based on images, so we have provided the option to upload images. It utilizes computer vision

techniques and deep learning models to predict the sign language gestures.

For numbers, alphabets, and words we require separate models, so we have trained this model with different functionalities For all the process is same but by just changing no of layers, last fully connected layers as per the problem and some regularization parameters, and also a directory of datasets.

III.CONVOLUTION NEURAL NETWORK

One of the algorithms utilized in machine learning is the convolutional neural network. They share characteristics with artificial neural networks in that they feature nodes or neurons connected by weighted links that generate an output in response to the input. Convolutional networks are more suitable to visual classification, such as photographs, which is the primary distinction. Regular neural networks have an output layer that provides the classification output and a hidden layer that connects to the previous input layer. However, massive volumes of data cannot be handled by standard neural networks. Convolutional neural networks are therefore more effective for a large number of images [10]. Convolutional neural networks' neurons are shaped in three dimensions, having height, breadth, and depth. Here, not only are all the neurons in one layer connected to all the neurons in the layer above it, but there is also a tiny zone of neurons that are connected to one another. The network's output layer will transform the image into a single vector along the depth axis.

A popular CNN variant that rivals multilayers Perceptron (MLP) has many convolutional layers before subsampling (pooling) of layers and FC layers end.

The model architecture consists of several layers:

Conv2D: Convolutional layer with 32 filters, a kernel size of (3, 3), and ReLU activation. The input shape is (64, 64, 3) representing the image dimensions.

MaxPool2D: Max pooling layer with a pool size of (2, 2) and strides of 2.

Two more sets of Conv2D and MaxPool2D layers with increased filters (64, 128) and different padding configurations.

Flatten: Flattens the output from the previous layer into a 1D vector. Several Dense layers with different units and ReLU activation functions.

Dropout layers to reduce overfitting. The specified dropout rates are 0.2 and 0.7.

The final dense layer with 9 units (corresponding to the number of classes) and softmax activation for multi-class classification. The model is built sequentially by adding layers one after another using the `model.add()` method.

IV.METHODOLOGY

In our society there are multiple people who are challenged with speech, hearing and visually-

impaired. Our solution aims at establishing alternative and effective way of communication for these people with other people of society. By using sign language i.e. by using hand gestures, we are recognising the words which person want to convey and that specific word or sentence will get displayed in the form of text over screen and also will get the text in audio form. By this, their communication will get easier than usual. The collection of images of different signs will be used as a database for our model. We are giving sign images as an input. The gesture detection technique will be used for sign gesture detection. After gesture detection, the pre-processing will be done on the sign images using Scale Invariant Feature Transform, Shape Descriptors, and Histogram of Oriented Gradients algorithms. The output of pre-processing will be passed to the different layers of convolution neural network. The output of CNN will be compared with a gesture dataset and it will give output in text and after that, we give that text in the form of speech.

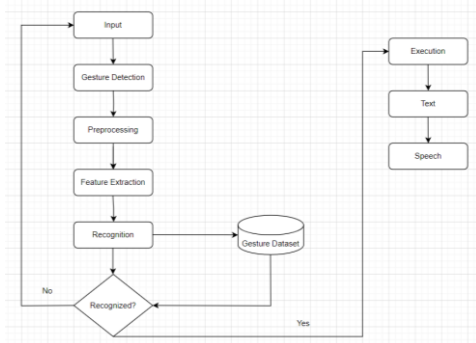


Figure 3: Proposed Methodology

1. To determine dynamic hand gestures using Neural Network and Computer Vision.

In this stage, we will recognise the hand gestures. In order to achieve this, first we will take image as input, then will perform pre-processing on that image. During pre-processing, noise will get removed by filtering. Then will do feature extraction based on the global and local features. Then will classify the image by using classification recognition algorithms such as CNN (Convolutional Neural Network), kNN (k-Nearest Neighbour) or Naive Bayes. After performing classification, we will get a resulting gesture (e.g., 'Hello', 'Thank You').

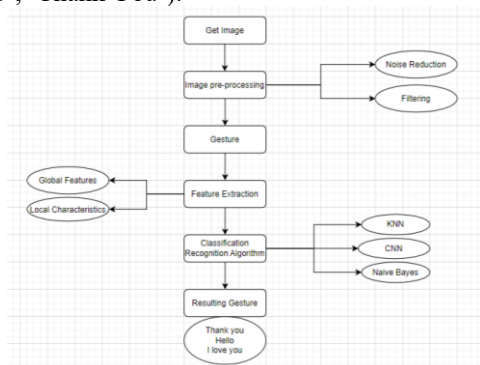


Figure 4: Dynamic Hand Gesture Methodology.

2. To achieve high accuracy using Convolutional Neural Network algorithm.

An important development in enhancing communication between the deaf and the general public is a real-time sign language detector. A Convolutional Neural Network (CNN)-based model for the recognition of sign language. Apply transfer learning to the problem by using a pre-trained SSD Mobile net V2 architecture that has been trained on a dataset. Furthermore, this tactic will be very helpful for sign language learners to practise sign language. Throughout the study, several human-computer interaction approaches for posture recognition will be investigated and evaluated. The optimal method will be determined using a number of image processing approaches with human movement categorization.

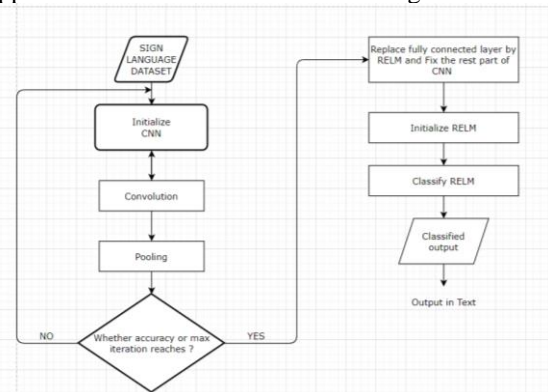


Figure 5: Convolutional Neural Network Methodology.

3. To convert the text to speech.

In this objective, at the very first step text is given to model which will apply text pre-processing using NLP. Text pre-processing used to remove functions, stop words and all pre-processing work. The output of pre-processing will give to the speech synthesis model. It is used to convert normal text into speech. The final result will be delivered as a speech.

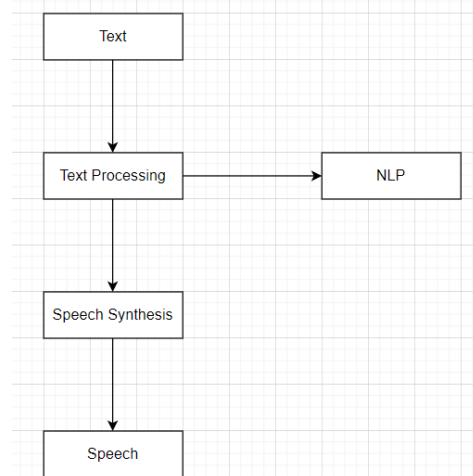
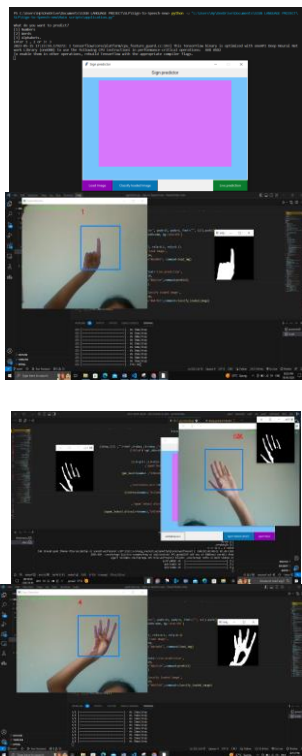


Figure 6: Text to speech conversion methodology.

V. RESULT

We have created a model so that it will detect Indian sign language gesture and give use output in text and audio form. In the sign predictor, we created three options Such as life prediction, load image and classify the loaded image. Through life prediction, we provide real-time gesture detection if the user wants to make the prediction based on images, so we have provided the option to upload images.

When we trained model, we got 98% accuracy for alphabet as well numbers and 96% accuracy for words. To get the best result of system we need fresh light when we give input to the system. For training purpose, we used 90% data and for testing purpose, we used 10% data. If we gave more data for training and testing so this system gives better output as well as accuracy.



VI. CONCLUSION AND FUTURE SCOPE

In this proposed solution, we will create a system which can recognize the ISL (Indian Sign Language) numericals, alphabets and some common words (e.g., 'Hello', 'Thank You'). The gestures will recognize by using Computer Vision, CNN, LSTM and GRU. By increasing layers in CNN, LSTM and GRU, we can achieve high accuracy. We will give output in the form of text and then will convert it into speech. So, at last we can get output in forms, text as well as speech. As an extension, we can enhance the dataset by

adding more vocabulary and can develop different datasets under ideal conditions. In future, we can collect more dataset so that we will get more accuracy.

REFERENCES

- [1] Varshney, Saurabh (2016-04-01). "Deafness in India". *Indian Journal of Otology*. 22 (2): 73. doi:10.4103/0971-7749.182281. ISSN 0971-7749. S2CID 78805217.
- [2] Madhuri Sharma, Ranjna Pal, Ashok Kumar Sahoo, "Indian sign language recognition using neural networks and kNN classifiers", August 2014. *Journal of Engineering and Applied Sciences* 9(8):1255-1259
- [3] Satwik Ram Kodandaram, N. Pavan Kumar, Sunil GI, "Sign Language Recognition", August 2021. *Turkish Journal of Computer and Mathematics Education (TURCOMAT)* 12(14):994-1009
- [4] Madhuri Sharma, Ranjna Pal, Ashok Kumar Sahoo, "Indian sign language recognition using neural networks and kNN classifiers", August 2014. *Journal of Engineering and Applied Sciences* 9(8):1255-1259
- [5] Juhi Ekbote, Mahasweta Joshi, "Indian sign language recognition using ANN and SVM classifiers", 2017 International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS)
- [6] Deep Kothadiya, Chintan Bhatt, Krenil Sapariya, Kevin Patel, "Deepsign: Sign Language Detection and Recognition Using Deep Learning", MDPI
- [7] Aman Pathak, Avinash Kumar, Priyam, Priyanshu Gupta, Gunjan Chugh, "Real Time Sign Language Detection", January 2022 *International Journal for Modern Trends in Science and Technology* 8(1):32-37
- [8] Rui Ma, Zengdong Zhang, Enqing Chen, "Human Motion Gesture Recognition Based on Computer Vision", February 2021
- [9] Nuwan Munasinghe, "Dynamic Hand Gesture Recognition Using Computer Vision and Neural Networks", April 2018.
- [10] Muneeb Ur Rehman, Fawad Ahmed, Muhammad Attique Khan, Usman Tariq, Faisal Abdulaziz Alfouzan, Nouf M. Alzahrani, Jawad Ahmad, "Dynamic Hand Gesture Recognition Using 3D-CNN and LSTM Networks", September 2021.
- [11] Chitralekha Mahanta, T. Srinivas Yadav and Hemanta Medhi, "Dynamic Hand Gesture Recognition System Using Neural Network".
- [12] Wenjin Zhang, Jiacun Wang, Senior Member, IEEE, and Fangping Lan, "Dynamic Hand Gesture Recognition Based on Short-Term Sampling Neural Networks", *Eee/Caa Journal of Automatica Sinica*, Vol. 8, No. 1, January 2021.
- [13] Keiron O'Shea , Ryan Nash, "An Introduction to Convolutional Neural Networks", December 2015.
- [14] Aijaz Ahmad Reshi, Furqan Rustam, Arif Mehmood , Abdulaziz Alhossan, "An Efficient CNN Model for COVID-19 Disease Detection Based on X-Ray Image Classification", 2021.
- [15] kavya Duvvuri, Harshitha Kanisetypalli, Sarada Jayan, "Detection of Brain Tumor Using CNN and CNN-SVM", 2022.
- [16] Ayushi Trivedi, Navya Pant, Pinal Shah, Simran Sonik and Supriya Agrawal, "Speech to text and text to speech recognition", *IOSR Journal of Computer Engineering (IOSR-JCE)*, March 2018.
- [17] Chaw Su Thu Thu, Theingi Zin, "Implementation of Text to Speech Conversion", *IJERT*, March 2014.