Sign Language Interpreter for Communication with deaf and dumb Individuals

Sumit Yadav[1, *], Mustafa Master[2], Ninad Shende[3],
Vedant Dadurwade[4], Vipul Lande[5]

[1-4] B.E. Students, Department of Electronics and Communication Engineering, Shri Ramdeobaba College of Engineering and Management Nagpur, 440013

[5] Asst. Professor, Department of Electronics and Communication Engineering, Shri Ramdeobaba College of Engineering and Management Nagpur, 440013

---***---

Abstract

Communication is an essential need for a human being. A part of the human population is born with deaf and dumb as a disability which hampers and affects their daily life. Sign language is a mode of communication for deaf and dumb individuals. In which specifically American Sign Language (ASL) is used all around the world for communication. The Continuous growth in technology and advancements helps us achieve and overcome the problem of communication. In this paper we created our own dataset for gestures. saving the hand gestures individually, and then training the system with data sets using Convolutional neural network for the hand gesture graphics. The application recognizes the gestures registered in the datasets when a user runs it, and the words associated with the gestures are then displayed on the screen. Also, the system is implemented on Raspberry Pi 4 for making it portable, reliable device. By using this technique, it is feasible to help people who have disability to communicate while also making it easier for everyone else to communicate with them.

Keywords: *ASL, Sign Language, Convolutional Neural Network, Raspberry Pi.*

Introduction

Communication using gestures, or hand-talk, has become a popular method of communicating for those who are unable to speak vocally. Sign Language around the world is used effectively by people with speech disability to communicate. Understanding of sign language helps individuals to communicate and convey their message in a hassle-free manner. The statistical data suggests that in India the 63 million individuals that have deaf and speech disability and 5 million are children. Although sign language is not very commonly known and understood by common public it become quite difficult to communicate with persons who do not understand or know sign language. Because of this deaf and dumb people may have difficulty communicating in such situations and this difficulty become hindrance in their communication which may lead to some false assumption.

Different sign languages are used in different countries or different regions. ASL is a language completely separate and distinct from English. It contains all the fundamental features of language, with its own rules for pronunciation, word formation, and word order. While every language has ways of signalling different functions, such as asking a question rather than making a statement, languages differ in how this is done.

Deaf and dumb individual often communicate through sign language interpreters. Finding a skilled interpreter, on the other hand, is difficult and frequently expensive too. As a result, a computerized interpreter might be a more dependable and less expensive option. Therefore, a system is proposed which can convert one's gestures into text in real-time.

To communicate with the deaf and dumb, sign language is a language of physical gestures used to convey different emotions. Although translation of sign language into speech or text is a difficult task. Here we aim to affectively address this issue of gesture recognition. In this research we used Convolutional neural network to train the model for sign language (SL) that directly converts sign language into alphabets. In this trained model the data sets that have been used were created by us. Image preprocessing was also effectively used to acquire high accuracy. As a result, we are able to translate sign language using images as input and text as an output.

Literature Survey

Ankit Ojha, Ayush Pandey, Shubham Maurya, Abhishek Thakur and Dr. Dayananda P proposed Sign Language to Text and Speech Translation in Real Time Using Convolutional Neural Network, paper presents a real-time sign language recognition system using CNNs. It presents developing an application that captures a person's signing motions for American sign language (ASL) and converts them into appropriate text and audio in real time using a computer's webcam. It discusses the architecture of the CNN model, which includes multiple convolutional and pooling layers followed by fully connected layers, the system uses CNN model of 11 layers and how it is trained using a large dataset of sign language gestures. The system works on the functionality and flow of Formation of gestures from the camera, training of the system learning model for image to contextual training, formation of words, formation of sentences and the content and obtaining the audio output. This system attains the accuracy of 95%. [1]

Rajesh George Rajan and Dr. P. Selvi Rajendran proposed Gesture Recognition of RGB-D and RGB Static images using ensemble- based CNN Architecture, system proposed works on Recognition of RGB-D and RGB static images based on CNN Architecture. The most common sign language is American Sign Language (ASL). The proposed method utilizes a SoftMax classifier and multiclass SVM that has been pretrained on VGG-19 (Visual Geometry Group). The pretrained network's completely linked layers produced a model with 4096 features. Thus, extraction is performed using the pretrained networks, while fitting and rectification are performed using a SoftMax layer. Except for Z and J, the system was trained by five separate users to recognized twenty-four static signs. A total of 95,967 photographs were gathered, with about 4000 images alphabetically collected for each user. Using deep learning models and VGG-19 training models, an integrated CNN model is utilized to assess American sign language, resulting in an accuracy of 99.93%. [2]

Arpita Halder, Akshit Tayade proposed Real-Time Vernacular Sign Language Recognition using MediaPipe and Machine Learning, the proposed system binds the communication gap for deaf and dumb individuals. The system works on Support Sector Machine (SVM) and a hand tracking method utilizing MediaPipe, an open-source project, and a machine learning algorithm provide real-time precise detection. The system collected data sets from different languages namely American, Indian, Italian, Turkish. The data sets comprising of numbers and alphabets for the language. The system used MediaPipe and Palm detection system for recognition of 21 points on the palm of the user for tracking of the hands. The system compared and utilized the MediaPipe framework for obtaining better results when compared to other architectures Artificial neural networks (ANN), KNN, Random Forest, Decision Tree, Naive Bayes, and multi-layer perceptron (MLP) are some examples. The SVM outperformed the other networks and achieved an accuracy of 99 percent for the majority of the data sets. Utilizing MediaPipe allowed the system to outperform previous networks and work better with various data sets. [3]

Jyotishman Bora, Saine Dehingia, Abhijit Boruah, Anuraag Anuj Chetia, Dikhit Gogoi proposed Real Time Assamese Sign Language Recognition using MediaPipe and Deep Learning, system proposed works on recognition of Assamese Sign Language Recognition using MediaPipe and Deep Learning. India is a very vast country with multiple languages the system works on sign language recognition for Assamese Sign Language. The system used Kscan3D software for capturing 3D images capturing with the Microsoft Kinect sensor. OpenCV library was used for capturing 2D images using Red Green Blue (RGB) webcam and extracting landmarks by MediaPipe hand tracking model, deep learning was implemented on the Jupyter Python notebook web application. The system used 2D and 3D images from multiple angles and distances to train the system. The palm detector detected 21 points on the system for detection, the system was trained for 30000 images, the files were converted to csv file and moved to pandas library. The gestures comprised of 9 selected alphabets from Assamese sign language from both hands, working with feedforward neural network with hidden layers and eventually ReLU activation layer as an output layer. The system was able to obtain 99 percent accuracy with the use of Mediapipe for Assamese Sign Language Model. [4]

Prangon Das, Tanvir Ahmed, Md. Firoj Ali proposed Static Hand Gesture Recognition for

American Sign Language using deep convolutional Neural Network, proposed system works recognition of American sign Language using convolutional neural network. The system used 2515 images from five different with variations in lighting conditions. The system used CNN model comprising of multiple convolutional and deep layers, using four groups of two convolutional layers taken after by a max pool layer, a dropout layer and one final output layer. The image sizes were specified to 200x200 and fed to the neural network. The pooling layers were added for reducing parameters and computations in the model. Similarly the system adds more convolutional layers with SoftMax Activation apart from ReLU layer at the system, Keras and tensorflow were also applied in order to reduce the training time. The images were then fed to existing models of of American Sign Language (ASL). The models which were compared were AlexNet, SUNY deepCNN, Stanford deepCNN, Rao deepCNN, RF-JA+C, ESF-MLRF and deep CNN of the system. Through experimental bases the system when compared obtained the maximum recognition rate of 94.34 percent and less sum of training and validation loss is also observed. The system was successfully able to deploy CNN and obtain the results. [5]

Impact on Society

The communication barrier that is the gap between people with deaf and dumb disability is bridged by this proposal. Sign language interpreting is used to guarantee that people who are deaf or hard of hearing have equal access to information and can interact effectively. It is now easier for those who have difficulty communicating to use sign language to interact with others. As a result, the mediator who usually acts as a translator is dismissed. This would provide a user-friendly environment for the user by speaking or writing in response to a sign gesture input.

The ability to lack of communication affects a major population in a way for not only in their day-to-day life but also affecting their lives in the longer run. Education, work places and communication in any possible manner to convey a message is difficult. The system proposed gives an individual with this disability a medium and platform to communicate with the world and makes their life easier. The reduced time and efforts and ability to use technology for the human life and its betterment is in which this project will majorly affect the society.

Proposed Methodology

We created our own dataset for each symbol of ASL with respect to its alphabet from A to Z. Then we preprocessed the captured gesture by subtracting the background remaining only skin part, then MediaPipe is used to identify hand landmarks and draw those landmarks getting skeleton images of each gesture. The model was trained on the data sets using CNN (Convolutional Neural Network). The preprocessed 150 gestures will feed to Keras CNN model. CNNs are a type of neural network that is useful for tackling computer vision problems. CNN includes layers such as a convolution layer, a maximum pooling layer, a flatten layer, a dense layer, a dropout layer, and a fully connected neural network layer. These layers, when combined, form a very strong tool for identifying features in an image.

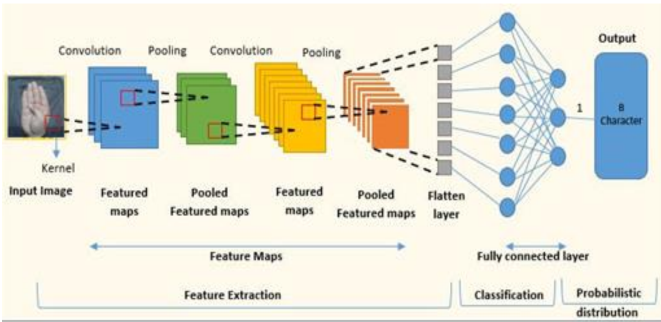


Fig: CNN

After Image preprocessing and gesture classification, a user-friendly GUI is created using Tkinter python library for capturing the gesture. The application takes input from user and predicts the character, forms the sentence and gives suggestion of words too. This helps in sentence building and gives the desired output to the respective user. Application also converts one's gesture into speech, for conversion of text to speech python library named pyttsx3 is used.

Block Diagram:

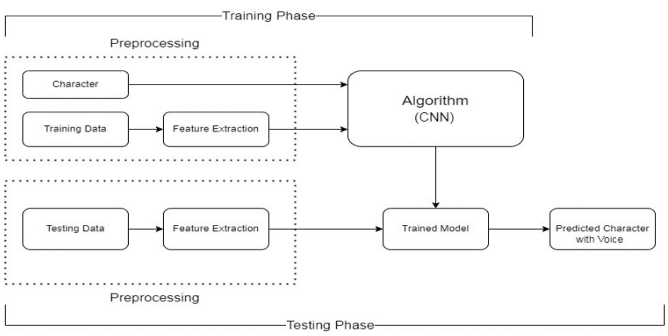


Fig: Block Diagram

Training Phase:

Step 1: For every gesture a video was created while the hand was moving, as the size of the gesture would vary while with movements. The video was further broken down to frames to extract 150 images per gesture.
Step 2: Further these were taken and put into image preprocessing and from the same remaining was the skin and the other remains were black in the background from which the only part remaining would be the gesture in the image.
Step 3: Then we use MediaPipe to identify hand in frames and get the hand landmarks in that image, then we draw and link those landmarks. This provides the skeleton image which helps to increase the accuracy using MediaPipe python library.
Step 3: The machine was then trained on the data sets using Convolutional Neural Network and the trained model file was then use to predict the gesture into character.

Testing Phase:

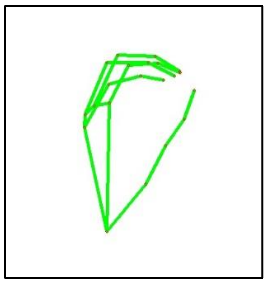
Step 1: Once the Graphical User Interface is started it captures the gesture of user. Further the captured image was put into preprocessing.
Step 2: The algorithm used for preprocessing earlier in the training phase is also used for testing.
Step 3: The preprocessed image is them given to trained model and then the system will recognize which gesture it is and character assigned for that particular gesture will be displayed on GUI as a 'character'.
Step 4: If the output character is correct as required then it is saved based on performing the save gesture, it will reflect the character in text area named 'sentence' otherwise capture same gesture again. There is also a 'space' gesture which add space between the words thereby forming a sentence. Auto suggestion of word is also provided for system to be fast, accurate, and reliable. The application provides two buttons: "clear" and "speak." The "clear" button is used to clear the sentence and start over again. The "speak" button is used to speak out the sentence that has been formed.

The training data set was created using around 150 images per gesture. Then with the help of preprocessing the skin was extracted and the background was removed completely and the remaining part was only the gesture. The preprocessed data set was trained using CNN algorithm. For the testing phase, the gesture image was captured using web cam and the same image was preprocessed and the output character was obtained on the GUI.

Observations


Captured Image


Preprocessed Image


Skeleton Image

Recognized Gesture

First, we capture the image of alphabet 'O' and then we preprocessed that image to extract only skin area and blackout the other part of the image and got only hand gesture at the output. Then using MediaPipe we got the landmark of the preprocessed image. In the result we got the actual alphabet that we have made for it along with the gesture image.

Fig: GUI

We made gestures for every single character of word 'HELLO' and saved the correct character accordingly. Combining each character got the word HELLO at the output. Also, we get the suggestion of word HELLO by python library named Enchant.

Text to Speech Conversion:

Project translates one's gestures into character but it also converts recognized character into its respective speech. For this, python pyttsx3 library is used which converts text to speech. This feature is useful for the individual who is blind and also replicates a real-life conversation.

Hardware Implementation:

To use a trained model on a Raspberry Pi 4, various Python libraries such as Keras, TensorFlow, and OpenCV were installed on it. In order to view the video output along with the predicted sign, a 7-inch display was connected to the Raspberry Pi. Portable power adapter of type C and Pi camera of 5MP is connected with the raspberry pi to make it a portable device for translating gestures. Once all the components were assembled correctly, the trained model was tested to identify gestures.

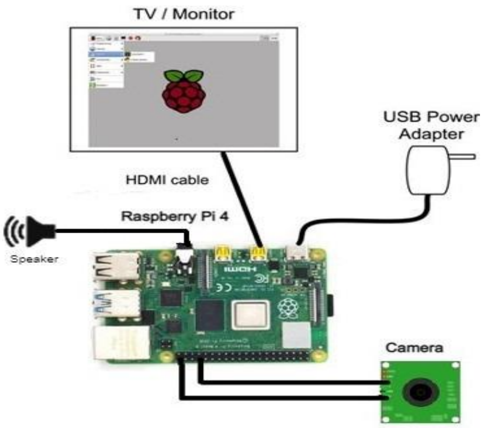


Fig: RPI4 Configuration

To capture gestures from a person who is deaf and dumb, a Raspberry Pi 4 processor and Pi camera were utilized. The image processing program has been written using Python and runs on the Raspbian Buster Operating System. It compares the image with other images stored in the database and uses a Convolutional Neural Network to predict the character. The resulting code is utilized to display the corresponding text on an LCD screen connected to the Raspberry Pi. Additionally, it is also used to generate audio output.

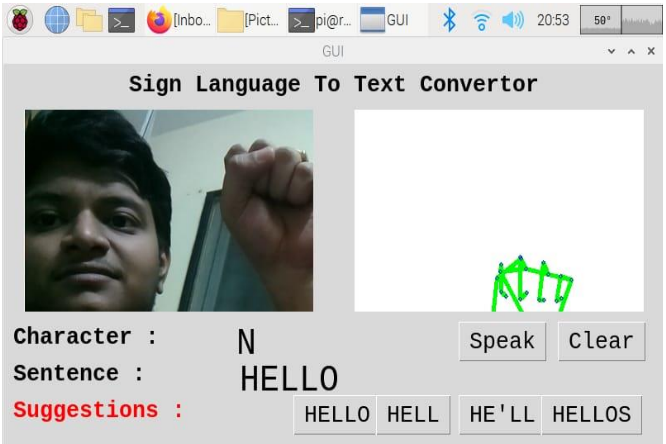


Fig: Raspberry Pi Implementation (GUI)

Results:

The present study proposes a gesture to text conversion system for deaf and dumb individuals using machine learning techniques. The primary objective of the study is to improve the accuracy of the system compared to the previous version of the project which employed MATLAB and AlexNet convolutional neural network (CNN). The current version employs a machine learning algorithm and has achieved an accuracy of 97% for 26 English alphabets, which is significantly higher than the previous project's accuracy of 83%. The accuracy of the proposed system is tested and verified by plotting an accuracy table for each alphabet to compare its performance. The results demonstrate that the proposed system outperforms the previous system in terms of accuracy

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
A	0.9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.1	0	0	0	0	0	0	0
B	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
C	0	0	0.9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
D	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
E	0	0	0	0	0.9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
F	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
G	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
H	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
I	0	0	0	0	0	0	0	0	0.9	0.1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
J	0	0	0	0	0	0	0	0	0.1	0.9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
K	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
L	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
M	0	0	0	0	0	0	0	0	0	0	0	0	0.9	0.1	0	0	0	0	0	0	0	0	0	0	0	0
N	0	0	0	0	0	0	0	0	0	0	0	0	0.1	0.9	0	0	0	0	0	0	0	0	0	0	0	0
O	0	0	0.1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
P	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
Q	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
R	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
S	0.1	0	0	0	0.1	0	0	0	0	0	0	0	0	0	0	0	0	0	0.9	0	0	0	0	0	0	0
T	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
U	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
V	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
W	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
X	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
Y	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
Z	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1

Fig: Accuracy Table

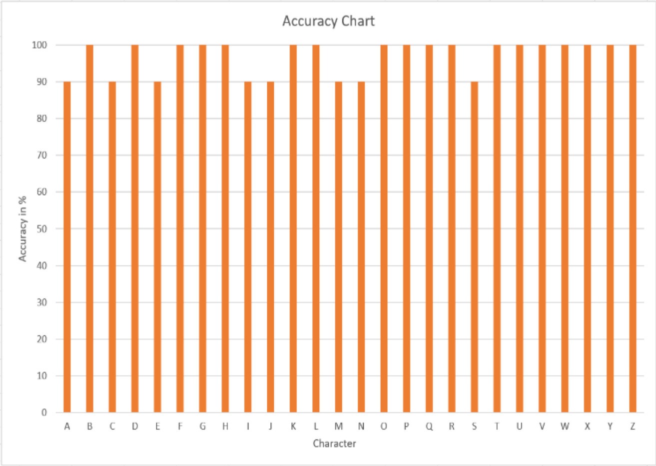


Fig: Accuracy Chart

Conclusion & Future Scope

The proposed approach supports the American Sign Language (ASL) to text on GUI Application and also implemented on raspberry pi 4 and got very good accuracy. We trained the model using a smaller number of training images and we attained a light weight model using CNN architecture. Therefore, CNN is efficient in getting high accuracy with a smaller number of training images. Although there can be provisions in future to change it to different sign languages as well. The sign language uses fingers to convey the alphabets although they can also be giving recognitions for objects or actions. As a result, recognizing this type of modulation would require more processing.

References

[1] A. P. A. M. S. T. A. &. D. P. Ojha, "Sign language to text and speech translation in real time using convolutional neural network," *Int. J. Eng. Res. Technol.(IJERT)*, vol. 8, no. 15, pp. 191-196, 2020.

[2] R. G. R. a. P. S. Rajendran, "Gesture recognition of RGB-D and RGB static images using ensemble-based CNN architecture," in *5th International Conference on Intelligent Computing and Control Systems (ICICCS)*, Madurai, India, 2021.

[3] A. a. A. T. Halder, "Real-time vernacular sign language recognition using mediapipe and machine learning," *International Journal of Research Publication and Reviews*, vol. 2582, p. 7421, 2021.

[4] S. D. A. B. A. A. C. D. G. Jyotishman Bora, "Real-time Assamese Sign Language Recognition using MediaPipe and Deep Learning," *Procedia Computer Science*, vol. 218, no. 1877-0509, pp. 1384-1393, 2023.

[5] T. A. a. M. F. A. P. Das, "Static Hand Gesture Recognition for American Sign Language using Deep Convolutional Neural Network," *IEEE Region 10 Symposium (TENSYMP)*, Vols. 1762-1765, pp. 1762-1765, 2020.