International Journal of Engineering Technology and Management Sciences
Website: ijetms.in Issue: 2 Volume No.6 March – 2022 DOI:10.46647/ijetms.2022.v06i02.006
ISSN: 2581-4621

Sign Language Detection System

Dr. Pallavi Chaudhari[1], Pranay Pathrabe[2], Umang Ghatbandhe[3], Sangita Mondal[4], Sejal Parmar[5]

[1] *HOD, Dept. of Information Technology, Priyadarshini College of Engineering, Nagpur, Maharashtra, India.*

[2] *Student, Dept. of Information Technology, Priyadarshini College of Engineering, Nagpur.*

[3] *Student, Dept. of Information Technology, Priyadarshini College of Engineering, Nagpur*

[4] *Student, Dept. of Information Technology, Priyadarshini College of Engineering, Nagpur.*

[5] *Student, Dept. of Information Technology, Priyadarshini College of Engineering, Nagpur.*

[1]pallavi.chaudhari@pcenagpur.edu.in
[2]pathrabepranay@gmail.com
[3]umangghathbandhe0@gmail.com
[4]mondalsangita794@gmail.com
[5]sejalparmar1916@gmail.com

***Abstract*— Sign language is a system for communication using visual gestures and signs for deaf and dumb people. This language is especially used by people who can't speak or hear. Sign language is oldest and natural form of language for communication, since most of the people are not aware of sign language, hence it is tedious to understand, for solving this problem we have built a real time method using convolution neural networks (CNN) for hand-based gestures to detect sign language. In our CNN model, first the hand is passed through a filter and then the filter is applied. Finally, the hand is passed through a classifier which predicts the class of the hand gestures. Our model provides 98% accuracy for the alphabet A-Z letters. On the other hand, the most common hurdle deaf and dumb come across is communication with the normal people and with the fact i.e., each and every normal person doesn't know the sign language. Another main feature of the project is to create a communication system for the deaf people. The functionality of these part of the system is to translate audio message to the corresponding sign language. This part of the system takes the audio message as input, and converts the audio recorded message into the respective sign image and videos and displays the relevant American sign language or GIFs which we have already defined. With the aid of this part of the system, communication between normal and deaf people gets feasible. Overall, the main idea of the project is to make a system which can help deaf and dumb to interact with normal people.**

***Keywords*—ANN, CNN, Pooling, Dataset Generation, ROI, RGB, Text to Sign Recognition, Hand Gesture Detection, Feature Extraction, ReLU, Machine-Learning, speech-to-text, machine translation, python, speech-recognition, NLP.**

I. INTRODUCTION

Sign languages are languages that uses the visual-manual mode to convey the meaning, Since the only difficulty faced by the deaf and dumb people is communication as they cannot speak languages, so the only way left is to communicate through sign language communication. Communication means to exchange ideas, feelings and messages. Communication can be established in many ways such as producing voice, signs, behaviors and perceptions. Speech and hear impaired humans express their feeling, thoughts, opinions to another human.

Deaf and dumb people express their ideas with other people by using their hands forming different gesture. Gestures are messages that are delivered indiscriminately and these signals are understood visually. Communicating with the deaf and dumb without using speech is called sign language. To bridge the gap between the normal people and hearing- impaired individuals they must know sign language.

Deaf and dumb people rely on vision-based communication. These Gestures can be easily recognized by the normal people if a common interface exists which can convert the sign language to text. So, many research has been done for vision-based communication interface system, where people who have speaking and hearing disability can enjoy feasible communication without having knowledge of each other's language.

Due to considerable time required in learning the Sign Language, people find it difficult to communicate with these people, creating a communication gap. Thus, we also propose a

International Journal of Engineering Technology and Management Sciences
Website: ijetms.in Issue: 2 Volume No.6 March – 2022 DOI:10.46647/ijetms.2022.v06i02.006
ISSN: 2581-4621

functionality which takes audio or speech as input in real time and converts it into text and displays the relevant sign language.

Our goal is to develop a user-friendly human computer interface, which understands the human sign language. There are various sign languages across the world, namely Indian Sign language (ISL), American Sign Language (ASL), British Sign Language (BSL), French Sign Language (FSL) and work had been held on other languages all around the world.

There are 3 major components of sign language (visual language):

1) Finger spelling- Spells the word letter by letter.
2) Sign vocabulary (word level)- Use for majority of communication.
3) Non-manual features- Body position, facial expression, tongue and mouth.

we created a model which can predict Fingerspelling based hand gestures in order to form a complete word by combining each gesture. The gestures we aim to train is in ASL standard which is given in the image below. And similarly, we can also train the model for ISL or any other sign language.

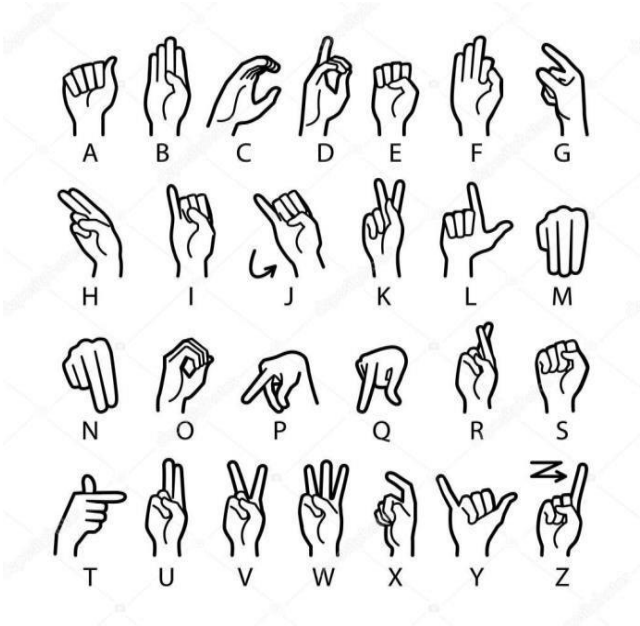


Fig.1: ASL signs

This paper also aims to develop an automated software system to translate speech to sign language, so that the language of a normal person can be understand by deaf person. This system first identifies the speech, next it converts speech to text, third it matches tokenized text with the visual sign word library (videos and images of signs), fourth this concatenates all the matched videos according to the text recognized and finally display the merged image and video to the deaf person.

II. LITERATURE SURVEY

There have been several researches carried out on the sign language system, and this has become the very popular subject and has been the topic of interest, and some of the methodologies have been elaborated as follows:

In method, developed by T. Shanableh to recognize isolated Arabic sign language. Glove is worn to get accurate details of hand outline to simplify process of segmenting out the hands of the signer via color segmentation. The efficiency of the proposed user-independent feature extraction was assessed by K-NN and polynomial networks. This research is mainly based on special sensor or electronic devices to predict the Sign Language.[1].

In technique created by P Subha Rajan and Dr G Balakrishnan for recognizing gestures for Indian signal Language wherein they proposed that each gesture could be recognized through 7-bit orientation and era process thru proper and LEFT test. the subsequent technique required approximately six modules and changed into a tedious method of recognizing signs and symbols [2].

Priyanka C Pankajakshan and Thilagavati B [4] have proposed a system of sign language with the use of ANN (Neural network of synthetic). The device involves holding a hand touch to create a path and select a signal using ANN. Das, A., Gawde, S., Suratwala, & Kalbande.D. and Rao, G. A., Syamala K, Kishore, P. V. V., & Sastry. [3] conducted basic research into sign language databases using CNN's set of rules to obtain positive results from school and database attempts. In [l] the authors proposed a gadget using CNN's Inception v3 in the database to check its accuracy and found it better than CNN. In [3] the proposed gadget is the use of the image language you have captured to manage images and is tested using stochastic integration. The CNN model used to train databases in [3] was designed for the use of specific window sizes and different set of images and the accuracy of the output was 92.88% compared to the contrasting techniques they researched.

[5] According to Neha V. Tavari A. V. Deorankar Dr. P. N. Chatur in his recording discusses that many people with physical disabilities rely on sign language interpreters to clarify their thoughts, so that they can communicate with others. The project presents a handshake taken using a webcam. The acquired image is processed and features are extracted. Functions are used as an algorithm type for detection. Visual touches are used to produce speech or text. In this gadget, the flexible sensor provides flexible analog output and requires multiple circuits so it has a very high value.

International Journal of Engineering Technology and Management Sciences
Website: ijetms.in Issue: 2 Volume No.6 March – 2022 DOI:10.46647/ijetms.2022.v06i02.006
ISSN: 2581-4621

III. METHODOLOGY

Sign to Text:

We have implemented a vision-based approach to our system. Every sign is displayed in front of the camera with an empty hand that completely removes the use of electrical or sensor equipment for use.

DATA SET GENERATION:

In the project we tried to use a predefined data set from various open sources, but we found that the images of these data set were not meeting our requirements. All the data images we could find were in RGB format. Therefore, we decided to generate our own set of data. The procedure we carried out to create our data set are as given below:

With the help of OpenCV library we successfully generated the data. The first step is to take approximately 800 pictures of each symbol in ASL for training purposes and about 250 pictures for each symbol for testing purposes. first, we captured each frame displayed on our machine's webcam. In each frame we present a region of interest (ROI) represented by a square with a blue border as shown in the picture below.

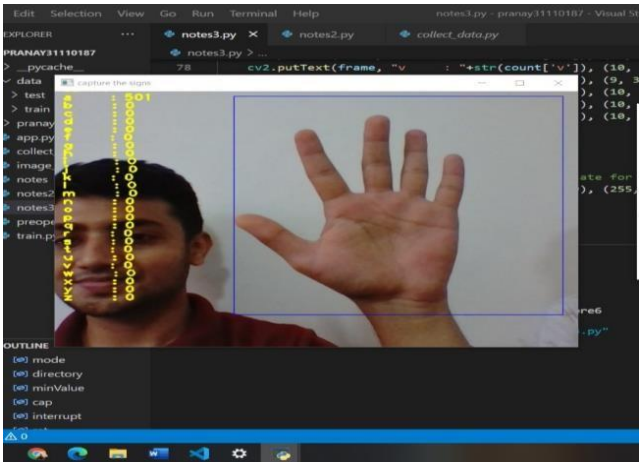


Fig.2: ROI (Region of Interest)

From the above window image, we extracted our ROI (Region of Interest) which is in RGB (Red Green Blue channel) and further it will be converted to grey scale Image as shown below.

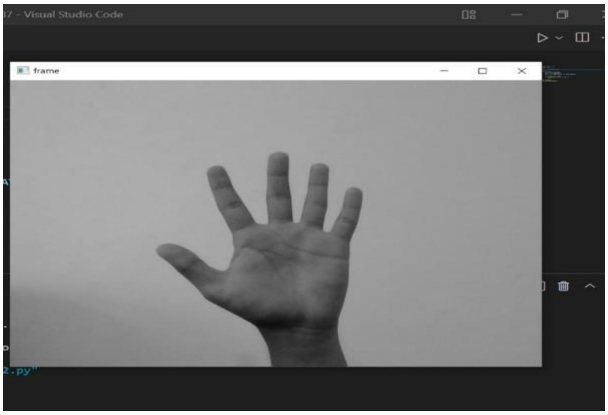


Fig.3: Gray scale image

Lastly, we applied a gaussian blur filter to the image which aided us in extracting different features of the image. Resulted image after applying gaussian blur filter is shown below.

Fig.4: output image after image pre-processing.

GESTURE CLASSIFICATION:

The technique which we are using in this project is as follows: The technique consists of two layers of algorithm to predict the hand gesture of the user.

Layer 1 Algorithm:

1. After taking the frame from OpenCV we apply gaussian blur filter and threshold to the image frame, to get the final processed image after feature extraction.
2. The output of step1 (image process) is sent to the CNN model for prediction. If a character is predicted in more then 50 frames, then the character is expressed and considered to build a word.
3. The space between word is indicated by a blank symbol and taken into consideration.

Website: ijetms.in Issue: 2 Volume No.6 March – 2022 DOI:10.46647/ijetms.2022.v06i02.006
ISSN: 2581-4621

Layer 1:

Convolution Model:

1. 1st CNN Layer: All the input images are of resolution 128x128 pixels. First it is processed in the first CNN (Convolution neural network) layer using weights of 32 filter (3x3 pixels each). Which will give a result of 126x126 pixel image, One Filter-weights for each.

2. 1st Pooling Layer: The down sampling of images is accomplished by max pooling of 2x2. In 2x2 square of array it keeps the largest value in it. As a result, our picture will be down sampled to 63x63 pixels.

3. 2nd CNN Layer: After pooling 63 x 63 pixels the picture is input to the CNN second layer. The input picture is processed in the second convolutional layer using weights of 32 filter (3x3 pixels each). which will give 60x60-pixel image in result.

4. 2nd Pooling Layer: Using max pool of 2x2 we again down sample the resulting images and as an output we get the image reduced to 30 x 30 resolution.

5. Densely Connected Layer 1: Now the output form the previous pooling layer we give it as embedded (input) in a fully integrated layer consist of 128 neurons at once, and the output from layer 2 of CNN is fixed to 30x30x32 =28800 values array. Input of this layer is consisted of a list of 28800 values. Result of the layer is provided to Densely Connected Layer 2. To avoid over fitting, we make use of a dropout layer which consist 0.5 value.

6. Densely Connected Layer 2: The output of densely connected layer 1 given as input to fully integrated layer which consist of 96 neurons.

7. Last layer: Last layer takes the input from 2nd Densely Connected Layer, which consist of the number of neurons which are equal to classes we are identifying (like alphabets and blank symbol).

Rectified Linear Unit Activation Function:

Using (Rectified Linear Unit) in every layer (convolutional and fully connected neurons). All input pixels max(x,0) are calculated by ReLu. It helps to more delicate feature because of additional nonlinearity to the formula. In addition to remove the gradient problem, the training is speed up by reducing calculation(computation) time.

Pooling Layer:

By the application of Max pooling to the image that is inputted which consist (2, 2) pool size along with ReLU activation function. This decreases the quantity of parameters, as a result it reduces the computation cost and overfitting.

Dropout Layers:
The hurdle of overfitting occurs right after the training, where the networking having weights are adapted to training examples, after giving new example as an input the performance of the network is not good. Set of activation is randomly selected in this layer by keeping value zero. Sake of accurate output for a specific example the network should provide the right output Even some of the activations are eliminated, the accurate result should be provided by the network [5].

Optimizer:
To update the model in response to the output of loss function we have used Adam optimizer. Combining the features of two extensions of two stochastic gradient descent algorithms.

Layer 2 Algorithm:

This layer works for differentiating the similar gesture.

1. If we come across the multiple sets of symbols which will show similar letter on detecting. Then apply next step.
2. Using classifier which are made for those sets of symbols, we then classify between these symbols which are similar.

Layer 2:

Using two layers of algorithms to verify and detect symbols which are more similar to each other so that we can get more detail to detect the symbol shown. By testing we realized that following symbols were not showing properly as they were mean to show, instead they were showing other symbols that were not to be shown for the given symbol:

1. For A: E and S
2. For P: N and Q
3. For I: T, D, K

For handling above errors that we come across, we created three different classifiers for classifying these sets:

1. {A, E, S}
2. {T, K, D, I}
3. {P, N, Q}

SENTENCE FROMATION

1. when the count of a letter is recognized, that it goes beyond the specific value and no other letter is near to it by a threshold, after that we print the letter and add it to the current string (In our python code we set the value as 50 and the value of difference threshold as 20).

Website: ijetms.in Issue: 2 Volume No.6 March – 2022 DOI:10.46647/ijetms.2022.v06i02.006
ISSN: 2581-4621

2. Otherwise, we remove the current dictionary which have the count of prediction of current symbol to avoid the probability of predicting a wrong letter.
3. The instance when the count of a blank (plain background) is detected exceeds a specific value and in addition to it, if the current buffer is empty no spaces are recognized.
4. In other case it detects the end point of a word by displaying a space and the current gets appended to the sentence.

Autocorrect Feature:
There is a python library for checking and autocorrecting the spellings knowns as Hunspell_suggest, which suggest synonyms or correct alternatives for the word which has to be printed. The user can alter or change his word by selecting a suitable word from the given suggestions by the Hunspell module. As a result, there is reduction in mistakes that may encounter in spellings and support in detecting complex words.

Training and Testing:

First, we convert our input RGB images to grayscale and also apply gaussian blur filter for removing unnecessary noise from the image. After that we apply adaptive threshold to extract only the details of our hand from the background and resize our images to 128 x 128 pixels. After pre-processing we feed the input images to our model for training and testing. After all the steps of image processing, the prediction layer we estimate and decide whether which image will be classified and belong to which classes. As a result, the output is normalized from the range 0 to 1 and therefore the sum of every value in each class adds up to 1. With the help of SoftMax function, we were able to obtained this. Firstly, the result of the prediction layer will be slightly differed from the real value. By application of labelled data, we have trained the networks to make it better. The performance measurement used in the classification is cross entropy. It gives a positive value that are not identical as labelled value and the value is zero exactly when it is same as the labelled value, it is all done by continuous function. To minimize the value near to zero, therefore we optimized the cross-entropy. For this purpose, we adjust the weight of our neural network to do this in our network layer. To calculate cross-entropy, TensorFlow has an inbuilt function. Finally, we have calculated the cross-entropy function. We optimized with the best optimizer known as Adam Optimizer using gradient descent.

Speech To Sign:

In this system, we added another new feature, which is audio to sign language translator using python. In this feature audio is given as input, and then search the corresponding recording using google api, and finally display the text on the system window and gives the sign of a given input using ASL (American Sign Language) generator. Now, all words in the sentence are checked against the words in the dictionary consist of images and GIFs denoting the words in a sentence. If the words are not found in the dictionary, it is then replaced with the synonym that is currently present in the dictionary. We have already defined Set of hand gestures in the system.

Algorithm:

Step1: speech to sign Conversion:

- Audio input is taken with the help of python library module named PyAudio.
- Translation of audio to text is done by using microphone.
- For analyzing grammar of the sentence and obtaining relationship between words dependency parser is used.

Step2: Text to Sign Language:

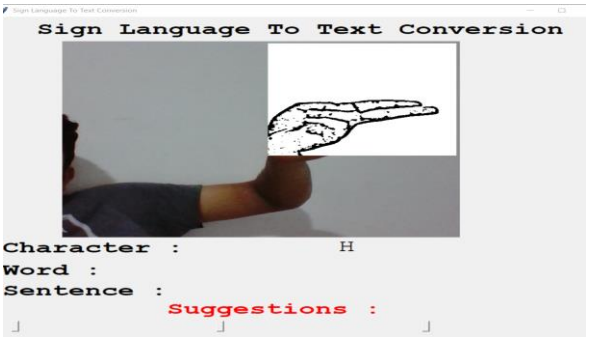
- Google Speech API is used to detect and recognized speech.
- By using NLP (Natural Language Processing), Text Pre-processing is done.
- applying machine conversion using dictionary
- ASL Generator: ASL of input sentence using ASL grammar rules.
- Generation of Sign language with gesture.

IV. IMPLEMENTATION AND RESULTS

Results for sign to text:

We achieved 95% Accuracy in our model using only 1 layer of our algorithm, and using a combination of layer 1 and layer 2 we get 98.0% accuracy, which is better than most current American Sign Language research papers

One thing to note is that our model does not use any background removal algorithm while the other models above do just that. Therefore, when we try to use different background in our project the accuracy may vary. many of the other projects use sensor devices but our main goal was to create a project that can be used with easily accessible resources. Below are demo images which shows how sign to text feature works. Fig 4(a) to 4(f)



International Journal of Engineering Technology and Management Sciences
Website: ijetms.in Issue: 2 Volume No.6 March – 2022 DOI:10.46647/ijetms.2022.v06i02.006
ISSN: 2581-4621

Fig.4(a): demo sign to text

Fig.4(b): demo sign to text

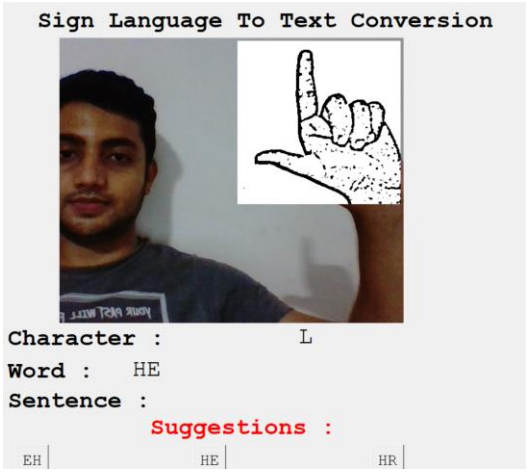


Fig.4(c): demo sign to text

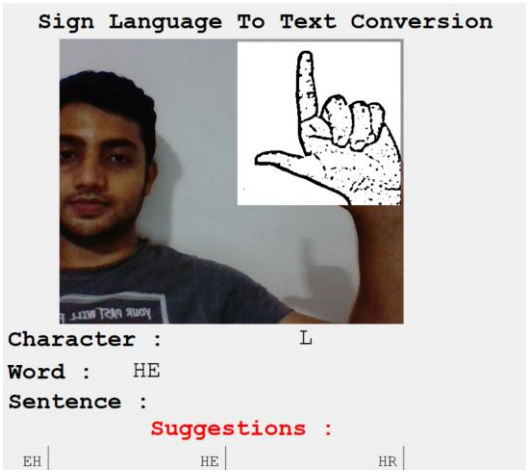


Fig.4(d): demo sign to text

Fig.4(e): demo sign to text

Fig.4(f): Text output for input signs gestures.

Result for Audio to text:

The effect of a particular English text is produced by producing its own equivalent sign language. The output of this program will be an ASL keyword. The pre-defined dictionary will have a video for each of the different names and the outgoing video will be an integrated video for such words.

1) Fig.5 shows the speech that is considered to be encoded using a PyAudio package

International Journal of Engineering Technology and Management Sciences
Website: ijetms.in Issue: 2 Volume No.6 March – 2022 DOI:10.46647/ijetms.2022.v06i02.006
ISSN: 2581-4621

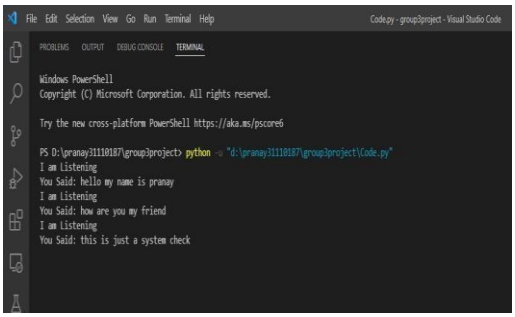


Fig.5: speech input.

2)Fig.5 shows the speech is recognized using Google Speech API.

The Google Speech-to-Text feature converts audio into text using neural network models in an easy-to-use API.

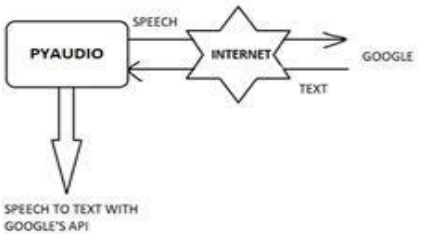


Fig.6: Google speech - text

3)Since we know that Machine can only understand two languages (i.e., 0 and 1) how can you understand our language. Therefore, to make the machine understand human language NLP was introduced. Indigenous Language Processing is a machine skill when you process and speak text. Understands the meaning of spoken words and appropriately produces a result. Pre-text processing consists of three components- Token Performance, Normal Performance and Audio Deletion as shown in Fig.6. Indigenous Language processing which is a mixture of synthetic ingenuity and integration languages. But really how it works with our project is very important. NLP can do extra work in our language. We will get our information after providing audio inputs based on NLP devices to understand human language. For example, Cortana and Siri.

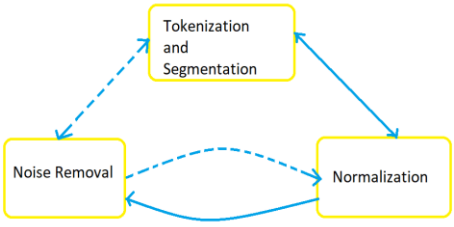


Fig.7: text pre-processing

It is not an easy task for the machine to understand our language but with the help of NLP, it is possible. Actually, how it works is shown below:

1. Provides audio as in-machine audio.
2. The machine records the input of the sound.
3. The machine then converts the sound into text and displays it on the screen.
4. The NLP system divides text into sections; understand the context of a conversation and execute based on the result of NLP

NLP is basically the process of creating an algorithm that translates text into a label based on location and word function in sentences. Human language is logically translated into numbers. This allows computers to understand the nuances that are clearly embedded in our language.

4) At last, Machine translation is executed by giving input as “Hello” into the microphone and the letter of the sentence will appear separately as output.

The output for speech to sign is shown in fig 8 a) to 8 e)

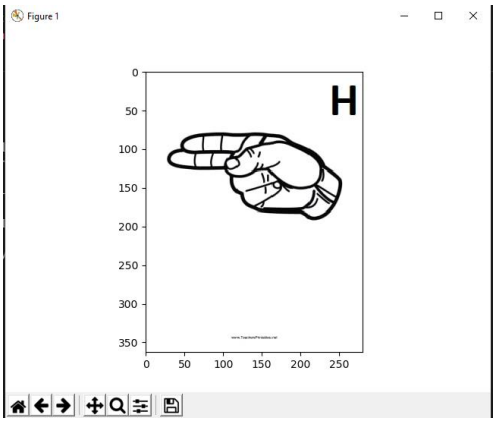


Fig.8 a): speech to sign output for letter H

International Journal of Engineering Technology and Management Sciences
Website: ijetms.in Issue: 2 Volume No.6 March – 2022 DOI:10.46647/ijetms.2022.v06i02.006
ISSN: 2581-4621

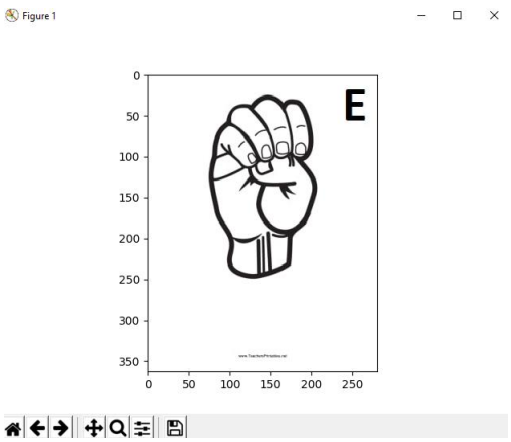


Fig.8 b): speech to sign output for letter E

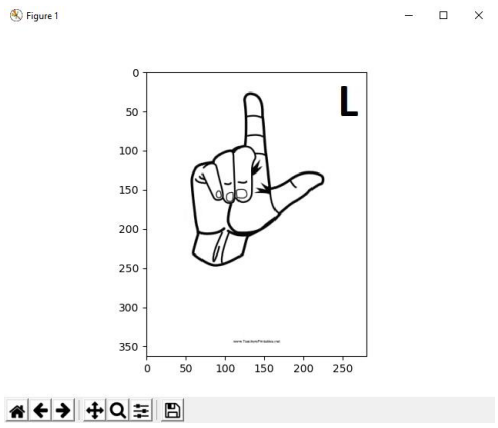


Fig.8 c): speech to sign output for letter L

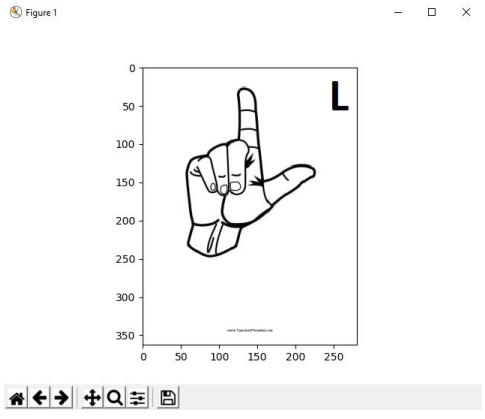


Fig.8 d): speech to sign output for letter L

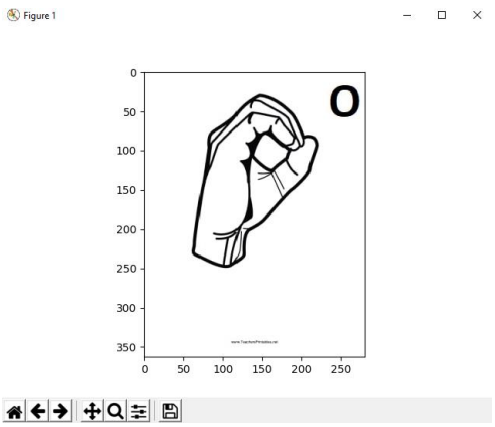


Fig.8 e): speech to sign output for letter L

Challenges Faced:

We faced many challenges in this project. We could not find any pre-existing database of ASL signs which is why we decided to create our own database and we also wanted the database images to be square images so that it gets easy to process it further. Another major problem was searching a filter that we can apply to our images to determine the appropriate image characteristics and we may be able to render that image as an application to a CNN model. We tried applying different filters like, median filter, canny filter, mean filter, binary threshold filter, gaussian blur, conservative smoothing etc. but finally we solved the problem by applying the gaussian filter. Additional problem faced after application of the filter was reduction in efficiency for detecting the model trained in the previous stage, which was eventually improved by increasing the size of the input image and improving the database.

V. CONCLUSION AND FUTURESCOPE

Sign to Text:

In this project, a real-time practical concept based on American Sign Language recognition for D&M people developed the alphabet ASL. We have found 98.0% final accuracy to detect the gesture and construct sentence using this model. We improved the model accuracy using two layers of algorithms in case if we counter verifying and predicting very similar signals. In this way we can see almost all the signs as long as they are properly displayed, there is no background noise and adequate lighting.

We plan to achieve high accuracy even if there is a complex domain by trying backlink removal algorithms. We also consider enhancing pre-processing to predict touch conditions in low light conditions with high accuracy.

Speech to Sign:

Website: ijetms.in Issue: 2 Volume No.6 March – 2022 DOI:10.46647/ijetms.2022.v06i02.006
ISSN: 2581-4621

sign language translator may be very useful in various regions. In faculties, schools, hospitals, universities, airports, courts anywhere, absolutely everyone can use this machine.

The news channels can use this feature of the system to report news by displaying sign language for the hear deficient person. Right now, only DD news Channel, use this kind of presentation but they use a person showing signs according to human speech giving live news. So, this will be the best idea we do give to the news channel. We look forward expanding project by re-adding facial expression to the system.

VI. REFERENCES

[1] Shanableh, Tamer, T. Khaled, "Arabic sign language recognition in user independent mode", IEEE International Conference on Intelligent and Advanced Systems, 2007, pp 597-600.

[2] Rajam, P. Subha and Dr G Balakrishnan, "Real Time Indian Sign Language Recognition System to aid Deaf and Dumb people", 13th International Conference on Communication Technology (ICCT), 2011, pp 737742.

[3] Rao, G. A., Syamala, 1<., Kishore, P. V. V., Sastry, A. S. C. S. "Deep convolutional neural networks for sign language recognition", (2018).

[4] Pankajakshan, P. C., & Thilagavathi B. Sign language recognition system. International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS), (2015).

[5] Vajjarapu Lavanya, Akulapravin, And M.S., Madhan Mohan “Hand Gesture Recognition and Voice Conversion
System Using Sign Language Transcription System" ISSN: