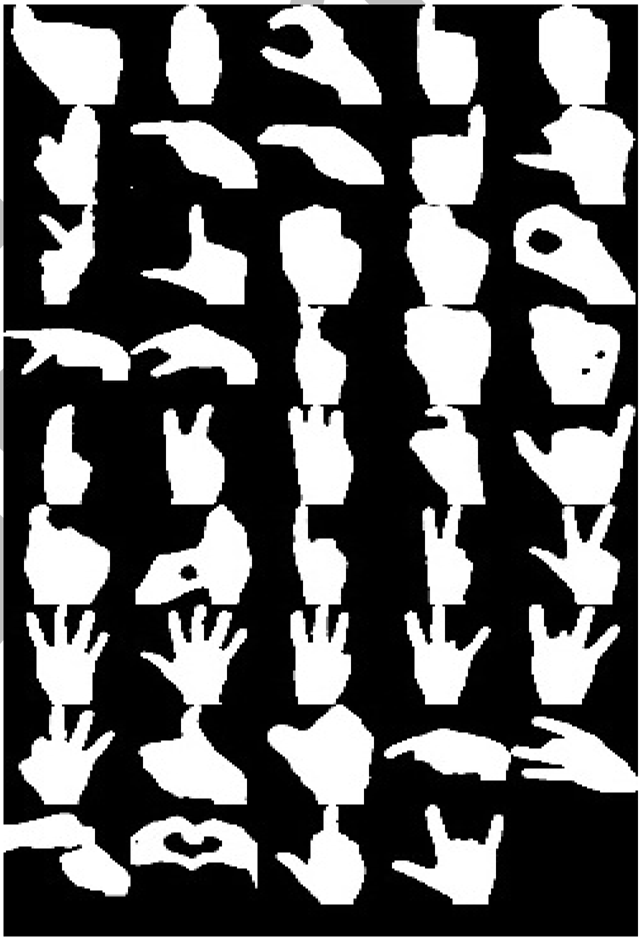
Sign Language Conversion into Text and Voice Using CNN for Vocally Impaired People

L. AARTHI , V. SENTHIL BALAJI , N. AISWARYA , N. PAVITHRA ,
N. PRASHANTHINI
DEPARTMENT OF INFORMATION TECHNOLOGY
SARANATHAN COLLEGE OF ENGINEERING
PANJAPPUR , TIRUCHIRAPALLI-620012, TAMIL NADU , INDIA.

ABSTRACT:

The main aim of this work to find unique solution/technique to help people with vocal impairment. People with vocal impairment uses sign language to communicate with others. It is impossible for all the people to learn sign language. This always leads to a communication gap between people with vocal impairment and normal people. In order to make it easy for both of them to communicate , this system helps vocally impaired people to communicate easily with others. In this system , user (vocally impaired people) can do hand gesture(sign language) which they want to convey to others. That sign language will be converted into text . Further the text which has been obtained is converted into voice which helps other people to easily understand the message which vocally impaired people meant to say. Thus the purpose of creating this system is that it will serve as the learning tool for those who want to know more about the basics of sign language such as alphabets, numbers and to break the communication gap between people with vocal impairment and normal people.



I.INTRODUCTION :

Sign language is a unique type of communication that often goes misunderstood or understudies. The process of translation of signs into spoken or written languages is formally known as 'interpretation'- the function which interprets plays is the same as that of translation for a spoken language. This work is mainly focussed on American Sign Language (ASL) which is used in USA. There are 26 hand gestures which corresponds to the 26 letters of the alphabet and it has also hand gesture for 10 digits. In the translation of sign language into spoken or written language 'fingerspelling' plays a major part. Fingerspelling is a method which is used for spelling words only through hand gestures. The reason fingerspelling plays a major part is that in sign language is that signers(vocally impaired people) used it to spell out names of anything for which there is not a sign. The names can be of anything , it may be a people's names, places, titles, brands, new foods, and uncommon animals or plants all fall broadly under this category, and this list is by no means exhaustive. Due to this reason, the recognition process for each individual letter plays quite a crucial role in its interpretation.

The Sign Language Recognition (SLR) architecture can be categorized into two main classifications based on its input as ‘data gloves-based’ and ‘vision-based’. ‘Data gloves-based’ refers to use of smart gloves to acquire measurements such as the positions of hands, joints orientation, and velocity using microcontrollers and specific sensors, i.e., accelerometers, flex sensors, etc.. The advantage of this approach is high accuracy, and the disadvantage is that it can be used only for limited movements. The ‘vision based’ refers to the use of cameras to obtain the gestures and translating them. This vision based becomes more popular in recent years.

This work is ‘vision based’ . This ‘vision based’ approach is categorized into two main parts. The first part is the ‘feature extraction’ which extracts the desired features by using image processing techniques or the computer vision method. It uses web camera to obtain the data (hand-gesture) from the user. The second part is that from the extracted and characterized features , the ‘recognizer’ should be learning of the pattern from the training data and correct recognition of testing data on which machine algorithms were employed. In this work Convolutional Neural Network (CNN) is used as recognizer of the system. Thus the main features of this work is creating words by fingerspelling method without the use of sensors or any other external technologies.

II. **RELATED WORK:**

For the past decades, Many studies shows that sensor-based devices such as SignSpeak are used for Sign Language Recognition. SignSpeak devices uses different sensors such as flex and contact sensors for finger and palm movements .For the hand movement accelerometer and gryos were used.Then, using Principal Component Analysis, the gloves were trained to recognize different gestures. And alphabets were recognised by the gesture. This device also uses Android phone to display the text and word received . the device SignSpeak was found to have 92% accuracy.

Sign dialect to content and discourse interpretation in genuine time utilizing convolutional neural arrange ,Making a desktop application that employments a computer’s webcam to capture a individual marking motions for American sign dialect (ASL), and decipher it into comparing content and discourse in genuine time. The interpreted sign dialect signal will be obtained in content which is more distant changed over into sound. In this way we are actualizing a finger spelling sign dialect interpreter. To empower the discovery of signals, we are making utilize of a Convolutional neural arrange (CNN). A CNN is exceedingly effective in handling computer vision issues and is competent of recognizing the required highlights with a tall degree of precision upon adequate preparing.

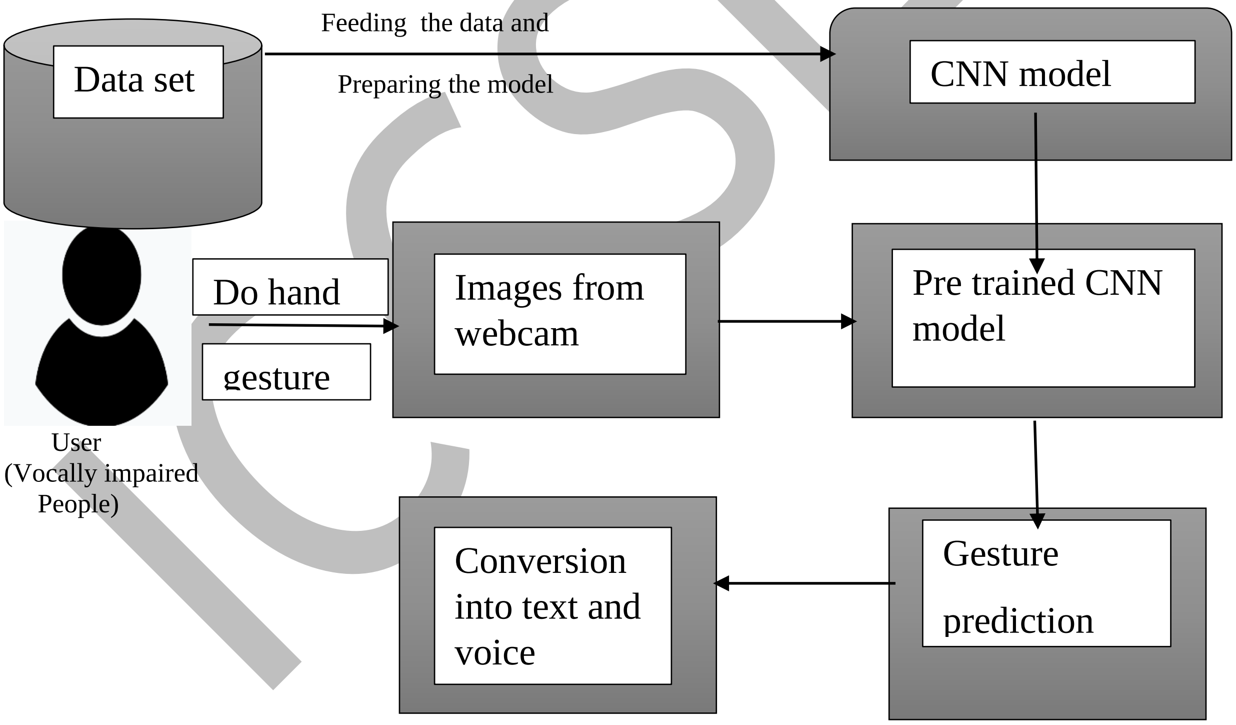
This inquire about centers on the improvement of sign dialect interpreter application using OpenCV Android based, this application is based on the contrast in color. The creator also utilizes Bolster Machine Learning to anticipate the name. Comes about of the investigate appeared that the coordinates of the fingertip look strategies can be utilized to recognize a hand signal to the conditions contained open arms whereas to figure signal with the hand clenched utilizing search methods Hu Minutes esteem. Fingertip strategies more flexible in signal acknowledgment with a higher victory rate is 95% on the remove variety is 35 cm and 55 cm and varieties of light intensity of roughly 90 lux and 100 lux and light green foundation plain condition compared with the Hu Minutes strategy with the same parameters and the rate of success of 40%. Whereas the foundation of open air environment applications still can not be used with a victory rate of as it were 6 overseen and the rest fizzled.

In our day to day life, communication plays vital part for passing on data from one person to another individual. But it gets to be exceptionally troublesome for the individuals who are hard of hearing and dumb to communicate with typical individuals. Sign dialect is the as it were one way to communicate with them. But ordinary individuals are ignorant of sign dialect. So there's only one way which is to covert sign dialect into text & discourse & bad habit versa. That's known as sign acknowledgment. Sign language could be a combination of body dialects, hand gestures and facial expressions. Among those hand motions are gives lion's share of the data .In this paper we are going examine the method proposed by creator for Sign Dialect Translation and its change to content.

As the internet has created it has gotten to be a put where individuals connected. They post opinions, modify and improve each other's commitments and share data. The issue we are investigating is sign dialect acknowledgment through administered include learning. With the technological slant in man-machine interfacing and the machine insights, misusing these powers has ended up a challenge in numerous areas. In specific, it was watched that the body gesture-based intelligent of human to human and human to machine are quickly increasing, especially within the range of sign dialect elucidation. Measurements emphatically propose that the population of deaf and quiet people is on the rise and there's a have to be prepare more individuals the American Sign Dialect (ASL) to bridge the hole. Moreover, the electronic gadgets such as TVs, PCs, PDAs Robots, cameras, etc. are progressed and built to studied clients motions and respond to their commands.

II. METHODOLOGY:

The system will be implemented through desktop. Initially the camera will capture the images of the hand which will be fed into the system. When the camera has captured the gesture from the user ,the system classifies the sample and compares them with the stored gestures. If the gestures matches it will display the corresponding output (text) on the screen for the user. The text which are obtained through hand gesture will be converted into voice using pyttsx3 (text to speech conversion library).



4 . ARCHITECTURE DIAGRAM

A. GATHERING OF DATA :

Keras comes with a library called 'datasets'. We can use them to load the datasets by downloading the data from the server and speeds up the process. We can also create gestures , Images will be captured and those images are converted to a

50 ×50 pixels black and white sample automatically. Each class contained 1,200 images .The script flips every image along the vertical axis. Hence each gesture has 2400 images. The train and test images along with the labels are loaded and it has been stored in variable.

B. HAND SKIN COLOR DETECTION

The Hand skin color can be detected using Image processing. Initially the signer need to have clear background for improvised skin color detection. Skin can be detected using cv2.cvtColor. Images which were obtained are converted from RGB to HSV. The HSV frame was supplied through the cv2.inRange function, with the lower and upper ranges as the arguments. The output from the cv2.inRange function was the mask. White pixels in the mask were considered as the region of the frame .Although black pixels are disregarded cv2.morph_ellipse function is used to remove small regions that may represent a small false-positive skin region. After dilations and erosions using morph_ellipse function , the resulting masks were smoothened using Gaussian blur.

C. NETWORK LAYER:

i. DATA PREPROCESSING:

The images which are obtained through datasets are grayscale images have pixel values that range from 0 to 255. Also, these images have a certain dimension. So the data must be preprocessed before we feed it into the model. At first convert each 50 x 50 image of the train and test set into a matrix of size 50 x 50 x 1 which is fed into the network. The data which has been obtained is in an int8 format. So before we feed it into the network it must be converted to float32 type. We have to rescale the pixel values in range 0 – 1 inclusive. Finally, for the model to generalize well, you split the training data into two parts, one designed for training and another one for validation.

ii. CREATION OF MODEL:

In Keras, the layers can be stacked up. We can add the desired layer one by one. First add a first convolutional layer with Conv2D(). The convolutional layer is composed of 16 filters of 2 X 2 Kernel. Then, a 2×2 pooling reduces spatial dimensions to 32 × 32. From 16 filters of the convolutional layers, filters are increased to 32, whereas that of the Max Pooling filters is increased to 5 × 5. Then, the number of filters in the Convolutional Neural Network (CNN) layers is increased to 64 . Next we need to add the Leaky ReLU activation function which helps the network learn non-linear decision boundaries. Since there need a nonlinear decision boundary that could separate these ten classes which are not linearly separable. Next, we need to add the max-pooling layer withMaxPooling2D(). But filters in maxpooling is still at 5 × 5. Dropout(0.2) functions are used to randomly disconnect each node from the current layer into the next layer. The model is now either flattened or converted into a vector format. The last layer is a Dense layer that has a soft max activation function with 10 units, which is needed for this multi-class classification problem. At last after creation of the model , compile the model and train the model using fit() function which in turn return a history object.

D. TRAINING THE SYSTEM :

Each dataset was divided into two ,training and testing. This has been done to see the performance of the algorithm used. The network was implemented and it has been trained through Keras and TensorFlow as its backend using a Graphics Processing Unit GT-1030 GPU. The network uses stochastic gradient descent(also known as the incremental gradient descent) as optimizer to train the network having a learning rate of 1 × 10−2 . The total number of epochs used to train the network is 50 epochs .Each epochs has a batch size of 500. The obtained images were resized to (50, 50, 1) for training and testing. This optimizer is used to minimize the batch size of large datasets. Redundant computations for large datasets has been performed by batch gradient. Then the gradients are recalculated before each parameter update for similar gestures. The stochastic gradient descent(SGD) optimizer eliminates this redundancy by performing one update at a time. It results in faster and it can be used for online learning.

IV. TESTING:

A. TESTING PROCESS:

Testing is the process of trying to discover fault or weakness in a work product. It provides a way to check the functionality of components sub assembles or finished product .It is the process of exercising software with the intent of ensuring that the software system meets its requirements and user expectations and does not fail in an unacceptable manner .There are various type of test, each test type address a software components is performed by the developers.

B. TESTING ACCURACY FORMULA

There are several rules that can serve as testing objective they are testing is a process of finding an error by executing it. specific testing requirement .Testing is not isolated to only one phase of the project but should be exercised in all phases of the project .After developing each unit of the software product an extensive testing process of the software is carried out by the developer. After the development of the software modules, a thorough unit testing and integration testing of each

High probability of finding an undiscovered error is a good test case. If testing is conducted successfully according to the objectives as stated above and it would uncover error in the program is a successful test.The accuracy of the letter , number , word can be obtained by the below formula.

$$\text{Accuracy rate} = \frac{\text{Total no of correct recognized letters or numbers}}{\text{Total no of users * Number of trials}}$$

V. RESULTS AND DISCUSSIONS:

A . LETTER RECOGNITION ACCURACY

In this system , accuracy rate of the gestures is the greatest concern. Figure 1 and Figure 2 represents the accuracy rate and average recognition of each letter from all the trials. The accuracy rate of each letter was obtained by the testing accuracy formula.

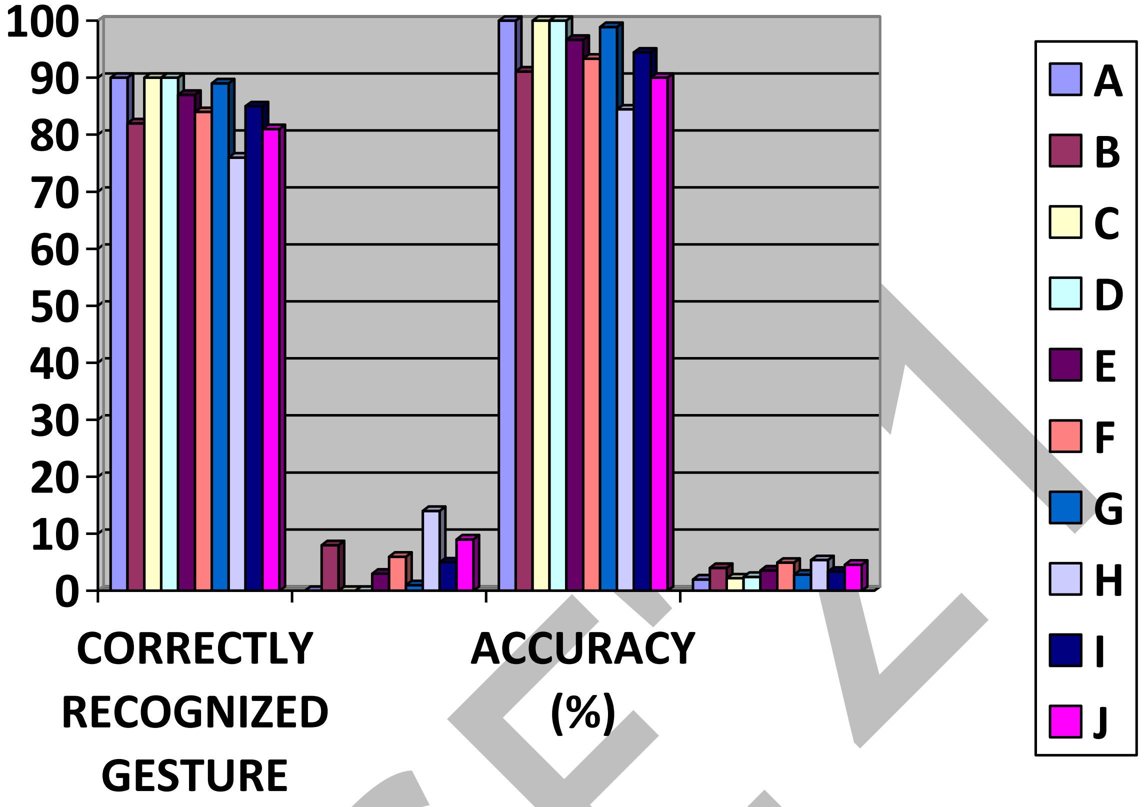


FIGURE 1 LETTER RECOGNITION ACCURACY (A-J)

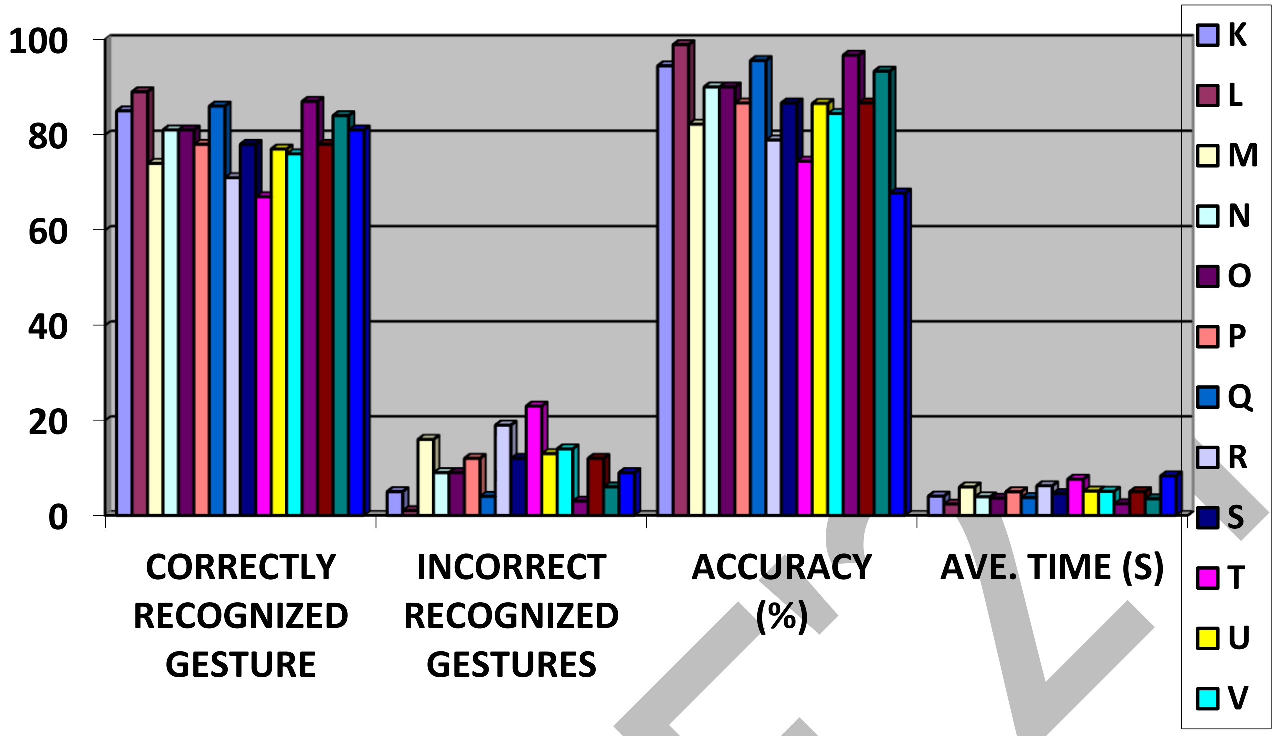


FIGURE 2 LETTER RECOGNITON ACCURACY(K-Z)

From the figures it can be seen that letters such as A ,C and D got the highest accuracy with 100 % rating and the letter Z got the lowest rating with 67 %. The overall letter recognition accuracy of the system was 90.04% accuracy attained in an average time of 4.31 seconds .This has been obtained by getting the average of each letter's accuracy.

NUMBER RECOGNITON ACCURACY:

The same computation can be performed for number gesture accuracy. Figure 3 represents the accuracy result of number.

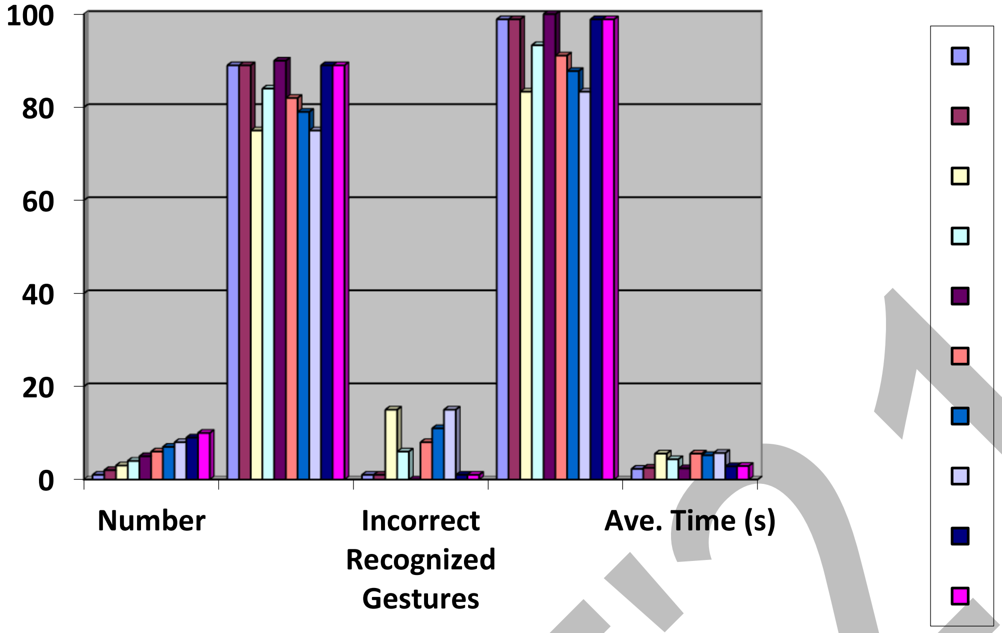


FIGURE 3 NUMBER RECOGNITION ACCURACY(1-10)

From the figure we can conclude that number 5 got the highest rating with 100 % and the number 8 got the lowest rating with 83.33%. The overall rating for the number recognition can be obtained by getting each number accuracy. The overall accuracy is 93.44% attained in an average time of 3.93 seconds.

VI . CONCLUSION:

Thus the project aims to break the communication barrier between vocally impaired people and normal people by converting sign language done by vocally impaired people into text format . Then the text will be converted into voice .In future this system can be used by visually impaired people and hearing impairment people to communicate with others.

VII. FUTURE ENHANCEMENT:

The project made ensures that the project could be valid in today's challenging real world. It has a vast scope in future .More functionalities can be added in accordance with the flexibility of the user requirement and specification. This project can be enhanced in few ways. In future, it could be built as a web or mobile application for the users to conveniently access the project. The existing project can be extended to work for other native sign languages with enough dataset and training.

REFERENCES:

1) Ming Jin Cheok , Zaid Omar , "A review of hand gesture and sign language recognition techniques"

2) Bheda, vivek and diannaradpour 'Using deep convolutional networks for gesture recognition in american sign language'. Arxiv abs/1710.06836 (2017): n. Pag.

3) Sharmila gaekwad, akankshashetty, akshayasatam, mihirrathod, pooja shah(2019), "recognition of american sign language using image processing and machine learning", IJCSMC, pg: 352-357,2019.

4) Tolentino, lean karlo S. Et al. "Static sign language recognition using deep learning." International journal of machine learning and computing 9 (2019): 821-82

5) Brandon Garcia Stanford University Stanford-CA, Sigberto Alarcon Viesca ,Stanford University Stanford-CA

,“Real-time American Sign Language Recognition with Convolutional Neural Networks”

6) Shubham Patil, Standford University “Sign Language Recognition and Transcription with Neural Networks”

7) Kacper Kania and Urszula Markowska-Kaczmar , Wroclaw University of Science and Technology ,Polland “American Sign Language Fingerspelling Recognition using Wide Residual Networks”