Sign Language to Text-Speech Translator Using Machine Learning

Akshatha Rani K [1], Dr. N Manjanaik [2]

[1] Student, Digital Communication and Networking, University BDT college of Engineering, Davangere,
Karnataka, India, akshatharani027@gmail.com
[2] Professor, Digital Communication and Networking, University BDT college of Engineering, Davangere,
Karnataka, India, manjubdt2009@gmail.com

ABSTRACT

Communication with deaf and dumb people is quite difficult task for others. So, through sign language can communicate with deaf and mute persons but it is difficult for normal people to understand the sign language hence it creates a huge gap between them and it's uneasy to exchange their ideas, thoughts with others. This gap has existed for years in order to minimize this, new technologies should be emerged. Therefore, an interpreter is necessary which acts as a bridge between deaf-mute and others. This paper proposed system which is a sign language translator. The system used American Sign Language (ASL) dataset which is pre-processed based on threshold and intensity. This system recognizes sign language alphabet and by joining the letters it creates a sentence then it converts the text to speech. As the system is based on hand, hand gesture is used in sign language recognition system, for that the efficient hand tracking technique which is given by media pipe cross platform is used and it exactly detects the hand after that by using the ANN architecture the model has trained and which classifies the images. The system has achieved 74% accuracy and recognize almost all the letters. The system which also converts sign text to speech so that it will also helpful for blind people.

Key words: ANN, ASL, deaf-mute, hand gesture, Sign Language.

1. INTRODUCTION

Communication is an important media to convey thoughts and expressions among the groups or between the individuals. Good communication leads to good thoughts and it helps for developments. In order to communicate, language is an essential tool, language means it not only to be in words but also can be an action. Sign language is used by deaf and mute people in order to communicate with others through body movement and hand gestures. All are unable to understand sign language so it becomes difficult for deaf, hearing impaired and speech disabled persons to communicate and express their thoughts with others. As a result, this challenge is a barrier between deaf, dumb people and others.

Therefore, in order to overcome this challenge sign language recognition system is a powerful tool and so many researches are carrying on in this field which are very helpful for the society. In this competitive world, day by day technologies are getting advanced so this interpreter plays a major role and by this system equal opportunities will be available for all regardless of their disabilities.

In this world numerous different languages are there, in different regions people will speak different languages like that sign language will also differ according to the regional language. In this paper American Sign Language (ASL) is used and communication is carried out in English. There are two groups in sign language recognition namely static and dynamic sign language. In this paper static sign language is used that is data is in the form of images and hand tracking technique is used which tracks the hand efficiently [1]. This system recognizes the hand gestures on real-time which are captured by the camera.

This system is built by using machine learning algorithm and data is processed that is given to the model which is built by deep learning neural network and then prediction will be taken place in real-time manner.

2. LITERATURE REVIEW

[2] proposed hand gesture recognition using Karhunen-Loeve (K-L) transform with this method they have also used CNN. For hands detection they used skin filtering, palm cropping to extract the palm area of hand and edge detection to extract the outline of palm. Then feature extraction of hand was carried out by using K-L transform method and image classification by using Euclidean distance. They tested for 10 different hand gestures with 96% of accuracy.

[3] proposed single hand sign language gestures recognition using contour tracing descriptor. In this paper, segmentation of hand contours from image background was carried out by using skin color detection with RGB and YCbCr color spaces, and threshold intensities of grey level. Contour tracing descriptor was used for gesture contours

detection by segmentation. They used SVM end KNN supervised machine learning techniques for image classification to evaluate the accuracy.

[4] proposed hand gesture recognition using PCA. This system color model approach and thresholding method with effective template matching for hand detection. Hand recognition is segmented with skin color modelling in YCbCr color space. Otsu thresholding is used for foreground and background separation. PCA is used for template matching for gesture recognition the system achieved accuracy of 91.43% for low brightness images.

[5] proposed ASL gesture recognition by using deep CNN for letters and digits. In this paper, images were pre-processed in which image background was removed by using background subtraction technique. The dataset was split into two, one for training and other for testing, and they have used CNN to classify images. The system achieved 82.5% accuracy on the alphabet gestures.

[6] proposed review of hand gesture & language recognition techniques. The system carried out data acquisition, pre-processing in which it used median and gaussian filter for noise reduction, morphological operation to remove unwanted information and histogram equalization, then segmentation in which it has skin color segmentation and tracking for hand detection and next step is feature extraction it used various methods and at last image classification. Overall, this paper provides comprehensive introduction in field of automated gesture & language recognition.

[7] proposed dynamic sign language recognition system. They used supervised learning algorithm called SVM for image classification, prediction and identification. This system recognizes sign gestures from a live video feed. In this system it extracts the hand contours from the frames of video by darkening the images and getting the white border of the hand this border is used to identify the hand contours.

[8] proposed sign language recognition for static signs using deep learning. This system used skin color modelling technique for hand detection and skin color range is predetermined that will extracts hand pixels i.e., foreground from non-pixels i.e., background. The system used CNN for image classification and images has uniform background. The system achieved accuracy of 90.04% for asl alphabet recognition and 93.67% testing accuracy.

[9] proposed hand gesture recognition for static images based on CNN. In the system, image pre-processing has morphological operations, contour extraction, polygon approximation and segmentation. They used different CNN architecture for training and testing to extract the features from images, classify them, then compared the results of all the CNN architectures.

[10] proposed sign language recognition system using CNN and computer vision. The system used HSV color algorithm for hand gesture detection and they set the background black. Image pre-processing consists of grayscale conversion, dilation, mask operation and hand gesture was segmented. The CNN architecture was used for feature extraction in the first layer and then for image classification. This system was able to recognize 10 alphabets and it achieved 90% of accuracy.

3. SYSTEM METHOD

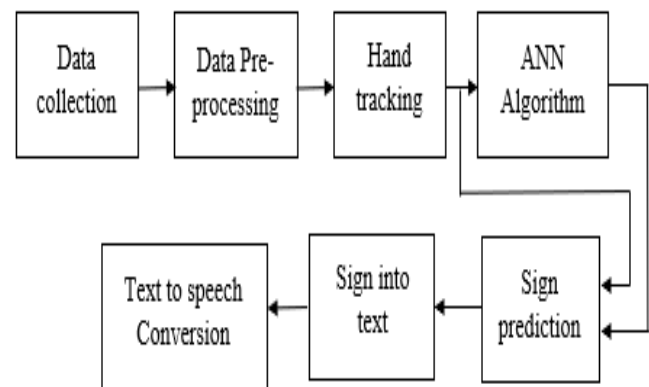The below figure 1 shows the block diagram of the proposed system.

Figure 1: System block diagram

This system consists dataset of images are available in Kaggle website which are captured by camera. These images are pre- processed in which thresholding and intensity rescaling operations are carried out, and after the pre-process by using hand tracking technique the system will consider the images in which the hand is detected. Then images are saved in the form of file. After that these images are trained by using ANN architecture and the model is saved. By using this model further, the system can predict the sign language alphabet in real time, one by one by joining the letters system can create a sentence. Then the text is converted to speech.

3.1 Sign Language Dataset

This is the first most and one of the crucial steps in machine learning. Data is collected from the Kaggle website, is an online community for machine learning practitioners. Here the data set used is an American Sign Language (ASL) alphabet, it is partitioned into two, for training and testing. The training folder which consists of 26 folders of ASL alphabet with one folder of 'space' character. Each folder consists of 2000 RGB images and these are all static images. In order to get the higher consistency, these images are captured with the same background and images are in RGB color space with the size of 200 x 200 and these are in JPG format.
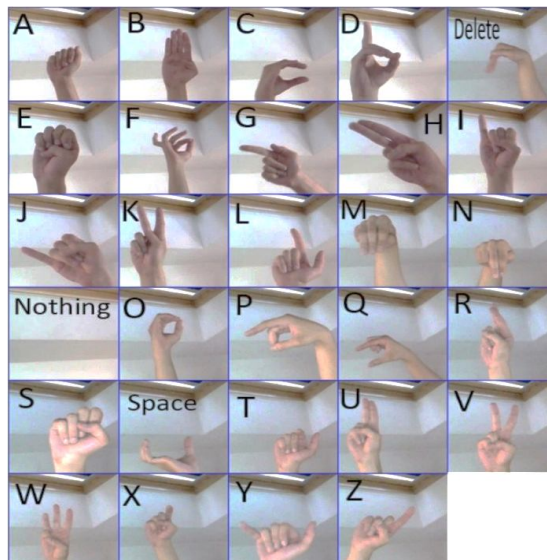
Figure 2: Sign Language hand gestures

3.2 Data Pre-processing

Image pre-processing step consists threshold setting and rescaling the intensity of images. Before these, captured images are in RGB form so first have to convert these RGB images to BGR form, then thresholding and intensity rescaling operations are carried out. Threshold operation in which automatic multilevel thresholding of colour images taken place and it searches for upper threshold value, then pixels which has intensities lower or equal to this value are assumed as foreground. Intensity rescaling operation which is used to stretch or shrink the intensity range of the given image. After these pre-processes, in the resulted images system tries to detect the hand and it will consider the images in which it can track the hand and it form a final dataset which is used for further process.

3.3 Hand Tracking Technique

In this paper, mediapipe hand tracking technique is used. Mediapipe is a cross platform framework which facilitate to build multimodal applied ML pipelines. Mediapipe hand is a high-Fidelity hand tracking solution, it works on real time which recognize hand skeleton of input image captured by the camera. This technique involves two models: palm detector model and hand landmark model [1].
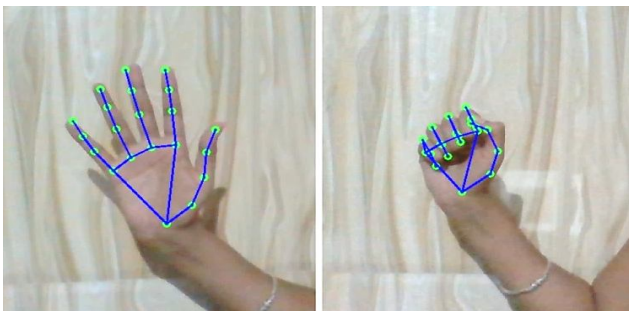

Figure 3: Hand-tracking using mediapipe

Palm detector model which provides a bounding box of a hand and recognises the palm through that bounding box in an input image [12]. Hand detection is a quite Complex task because there are variety of hands with different sizes so the system should be able to detect the hand. Here, palm detector is trained instead of hand detector because estimating the bounding boxes of palm and fist are simpler than hand fingers. Next encoder-decoder feature extractor is used and minimise the focal loss during training.

Hand landmark model which the predicts the hand skeleton on an input image which is in the bounding box provided by palm detector, in turn hand landmark model results 3D landmarks. After executing the palm detector on an input image, hand Landmark model locates the landmarks of 21 3D points on the hand which is detected in the hand area. So that the model consistently learns the hand poses and becomes robust, even it can detect the partially visible hands [11]. Figure 3 shows the hand tracking with hand landmarks.

3.4 ANN Architecture

Artificial neural network is used for classification. The images to be classified are given to the network through neurons at the input layer. These activation function process the images and output will be given at the output layer [13]. Here, the ANN which has multilayer perceptron (MLP) that is it consists input layer, hidden layers and an output layer. Training of neural network has calculated the weights [14].
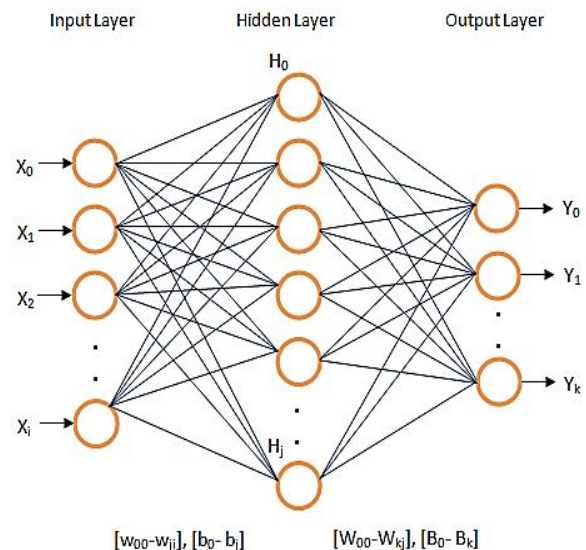

Figure 4: ANN Architecture

This system used ANN model with Keras and sequential model is used by arranging the Keras layers sequentially. First, dense layer is added with activation function ReLu and next dropout player with activation function ReLu is added, likewise alternatively dense layer and dropout layer is kept on added with 1024, 512, 256 ,128 and 64 filters. Then the model is compiled by using the categorical cross entropy as loss function and Adam as optimizer.

4. RESULT AND DISCUSSIONS

In the training phase, the system is trained by using 2000 images with the ANN architecture and the model is saved. now the prediction of letters takes place by using the model. the system first detects the hand in the live video frame, when the hand tracking is done then it recognises the sign and display it on the screen in the text format.

Figure 5: Predicted sign as letter 'A'

The fig. 5 shows the sign of letter 'A'. The system which tracks the hand, and it compares the hand pattern with trained images and then it predicts the sign is letter 'A' with the probability percentage of prediction.

In this system, we can also create the words by joining the letters one after other. The fig. 6 shows that first the system tracks the hand pattern, and it predicts the sign is 'A' then it predicts the next sign 'I' which is shown to the camera and after that, predicts the sign 'M', so it formed the word 'AIM'. Likewise, we can form any word.

By using the 'space' which is also trained under the model, we can form a sentence. After the sign to text conversion, the text can be converted to speech which is helpful for blind people. The system pronounces the word or text.
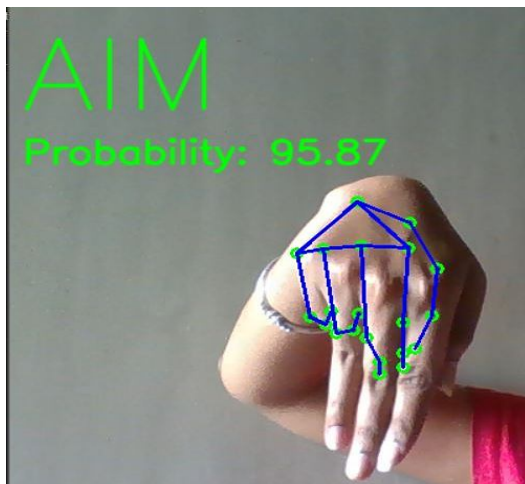

Figure 6: Word formation

The model has achieved 74% of validation accuracy with efficient hand tracking technique. Its graph is plotted with number of epochs against validation accuracy of the model, as shown in the below fig. 7.
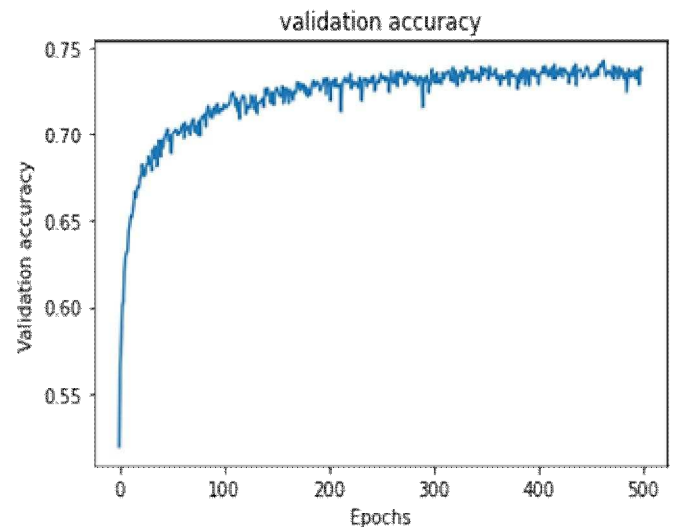

Figure 7: Graph of validation accuracy

5. CONCLUSION

There are many researches have been carried out in the field of machine learning and computer vision. They have contributed effective works which are very necessary and helpful for everyday life. Likewise various research has been done on sign language recognition using different methods like neural networks, KNN, SVM and LSTM. In this paper, the proposed system concentrated on hand tracking technique which is very effective technique. It also detects the hand for different skin colours and lighting condition, and it also detects the hand in low-light condition. We used ANN to classify images of asl alphabet, the system recognises almost all the letters and achieved 74% of accuracy. The system which also incorporates the speech which converts the recognized sign text to speech, so that it will also be helpful for blind people.

6. FUTURE WORK

The system model can be improved in terms of accuracy by using different classification methods so that the model will recognize the alphabet even more accurately.

REFERENCES

1. https://google.github.io/mediapipe/solutions/hands.html
2. Singha, J. and Das, K. **Hand Gesture Recognition Based on Karhunen-Loeve Transform**, Mobile and Embedded 232 Technology International Conference, January 17-18, 2013.

3. R. Sharma, Yash Nemani, Sumit Kumar, Lalit Kane, Pritee Khanna. **Recognition of Single Handed Sign Language Gestures using Contour Tracing descriptor**, Proceedings of the World Congress on Engineering 2013 Vol. II, WCE 2013, July 3 - 5, 2013, London, U.K.

4. Mandeep Kaur Ahuja, Dr. Amardeep Singh. **Hand Gesture Recognition Using PCA**, IJCSE, Vol 5, July 2015, Issue 7,267-271.

5. Vivek Bheda and N. Dianna Radpour. **Using deep convolutional networks for gesture recognition in American sign language**, arXiv preprint arXiv:1710.06836

6. Ming Jin Cheok, Zaid Omar, Mohamed Hisham Jaward, **A review of hand gesture and sign language recognition techniques**, Springer-Verlag GmbH Germany 2017.

7. S. Saravana Kumar1, Vedant L. Iyangar. **sign language recognition using machine Learning**, International Journal of Pure and Applied Mathematics, Volume 119 No. 10, 2018, 1687-1693.

8. Lean Karlo S. Tolentino, Ronnie O. Serfa Juan, August C. Thio-ac, Maria Abigail B. Pamahoy, Joni Rose R. Forteza, and Xavier Jet O. Garcia. **Static Sign Language Recognition Using Deep Learning**, International Journal of Machine Learning and Computing, Vol. 9, No. 6, December 2019.

9. Raimundo F. Pinto Jr., Carlos D. B. Borges, Antˆonio M. A. Almeida, and I ´alis C. Paula Jr., **Static Hand Gesture Recognition Based on Convolutional Neural Networks**, Volume 2019, Article ID 4167890, Published 10 October 2019.

10. Mehreen Hurroo, Mohammad ElhamWalizad, **Sign Language Recognition System using Convolutional Neural Network and Computer Vision**, International Journal of Engineering Research & Technology (IJERT), Vol. 9 Issue 12, December-2020

11. Fan Zhang, Valentin Bazarevsky, Andrey Vakunov, Andrei Tkachenka, George Sung, Chuo-Ling Chang, Matthias Grundmann, **MediaPipe Hands: On-device Real-time Hand Tracking,** arXiv:2006.10214v1 [cs.CV] 18 Jun 2020.

12. Wei Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng-Yang Fu, Alexander C. Berg, **SSD: Single Shot MultiBox Detector**, arXiv:1512.02325v5 [cs.CV] 29 Dec 2016.

13. https://medium.com/@gongster/building-a-simple-artifi cial-neural-network-with-keras-in-2019-9eccb92527b1

14. Z. Zhang, Multivariate Time Series Analysis in Climate and Environmental Research, in this Chapter 1-**Artificial Neural Network**, Springer International Publishing AG 2018.