Use of Machine Learning Applications for Speech Impaired People

Dr. Rajeshri Pravin Shinkar
Asst. Prof.
SIES (NERUL) College of Arts, Science & Commerce

Abstract **- While sign language is very important to deaf-mute people to communicate both with normal people and with themselves; it is still getting little attention from the normal people. We, as normal people, tend to ignore the importance of sign language unless there are loved ones who are deaf-mute. One of the solutions to communicate with the deaf-mute people is by using the services of sign language interpreters. But the usage of sign language interpreters can be costly. A cheap alternative to replace the interpreters is the use of a model that can automatically translate their actions into words. This paper covers the development of such a model which helps to automatically detect our actions in real time using the Mediapipe model and translate the same into respective text format.**

Keywords: Mediapipe, Gesture, sign language

I. INTRODUCTION

Deaf is a disability that impair their hearing and make them unable to hear, while mute is a disability that impair their speaking and make them unable to speak. Both are only disabled at their hearing and/or speaking, therefore can still do many other things. The only thing that separates them and the normal people is communication. If there is a way for normal people and deaf-mute people to communicate, the deaf-mute people can easily live like a normal person. And the only way for them to communicate is through sign language. Sign language is the mode of communication which uses visual ways like expressions, hand gestures, and body movements to convey meaning. Sign language is extremely helpful for people who face difficulty with hearing or speaking. Sign language recognition refers to the conversion of these gestures into words or alphabets of existing formally spoken languages. Thus, conversion of sign language into words by an algorithm or a model can help bridge the gap between people with hearing or speaking impairment and the rest of the world. Vision-based hand gesture recognition is an area of active current research in computer vision and machine learning. Being a natural way of human interaction, it is an area where many researchers are working on, with the goal of making human computer interaction (HCI) easier and natural, without the need for any extra devices. So, the primary goal of gesture recognition research is to create systems, which can identify specific human gestures and use them, for example, to convey information. For that, vision-based hand gesture interfaces require fast and extremely robust hand detection, and gesture recognition in real time. Hand gestures are a powerful human communication modality with lots of potential applications and in this context, we have sign language recognition, the communication method of deaf people[1][3].

II. OBJECTIVES

- The Sign Language Recognition model here is a real-time vision- based system whose purpose is to recognize the Universal Sign Language.
- The purpose of the model was to test the validity of a vision-based system for sign language recognition and at the same time, test and select body features(key points) that could be used with machine learning algorithms allowing their application in any real-time sign language recognition systems.
- This system once deployed on a full scale can replace the existing translation of using human interpreters and the process can be automated saving time and resource constraints.

III. METHODOLOGY AND STEPS

The system is a vision-based approach. All the signs are represented with bare hands and so it eliminates the problem of using any artificial devices for interaction.

Following is the FlowChart of steps involved in building the proposed system.

Import and Install Dependencies
⇓
Keypoints using MP Holistic
⇓
Extract Keypoint Values
⇓
Setup Folders for Collection
⇓
Collect Keypoint Values for Training and Testing(Dataset Generation)
⇓
Preprocess Data and Create Labels and Features
⇓
Build and Train LSTM Neural Network
⇓
Make Predictions
⇓
Save Weights
⇓
Evaluation using Confusion Matrix and Accuracy
⇓
Testing in Real Time

A. Import and Install Dependencies:

Tensorflow, OpenCV and Mediapipe are the primary libraries used here. Apart from these, Sklearn, Matplotlib, Numpy, OS and Time are some other libraries that are imported and used while building this system.

B. Keypoints using MP Holistic & Extract Keypoint Values:

Mediapipe is a cross-platform library developed by Google that provides amazing ready-to-use ML solutions for computer vision tasks.

MediaPipe Holistic utilizes the pose, face and hand landmark models in MediaPipe Pose, MediaPipe Face Mesh and MediaPipe Hands respectively to generate a total of 543 landmarks (33 pose landmarks, 468 face landmarks, and 21 hand landmarks per hand).

The MediaPipe perception pipeline is called a Graph. Let us take the example of the first solution, Hands. We feed a stream of images as input which comes out with hand landmarks rendered on the images.

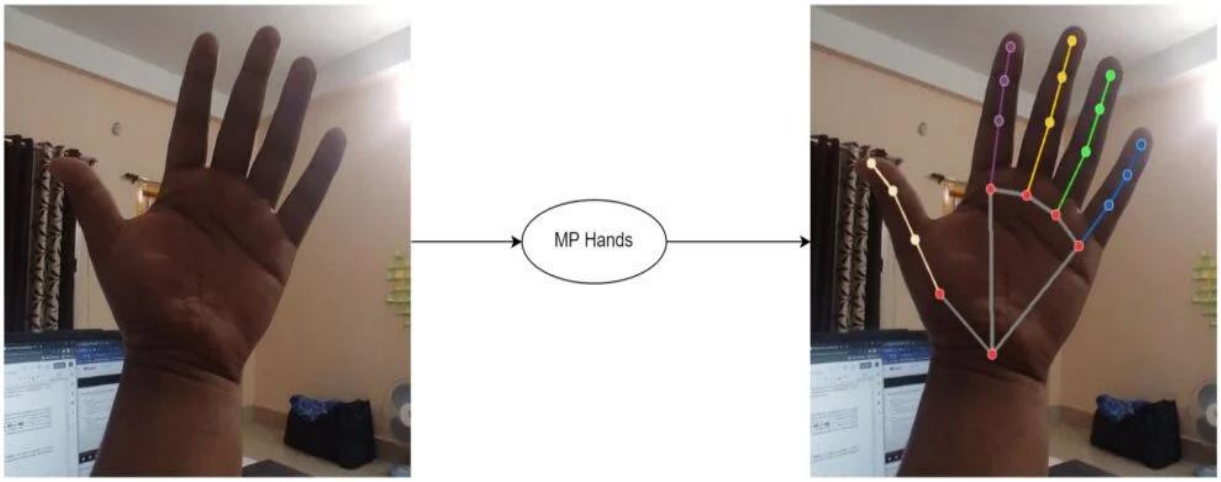


Fig. 1. Hand Landmarks

The flowchart below represents the MP hand solution graph.

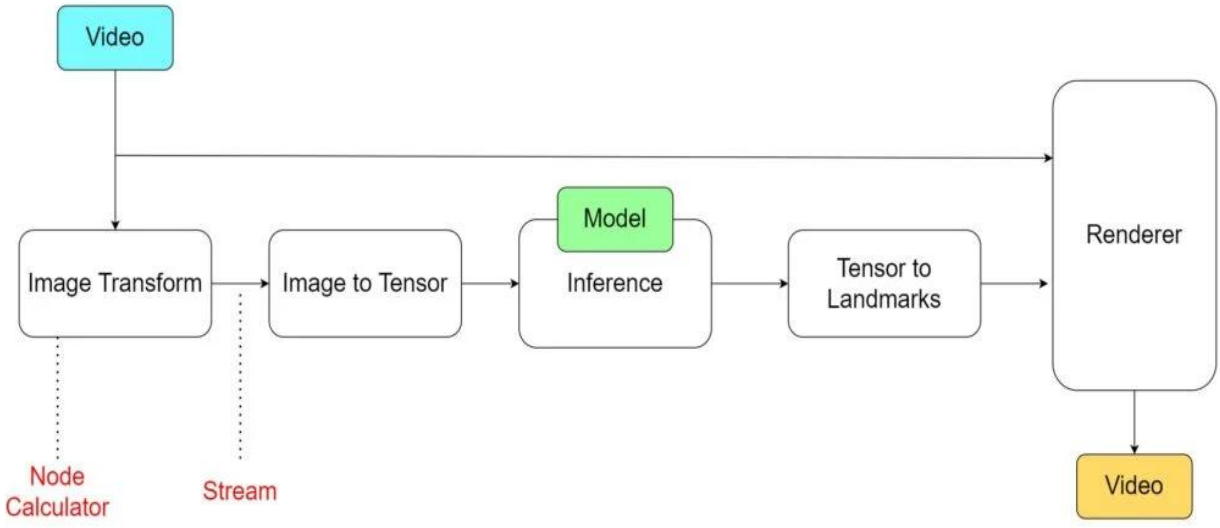


Fig. 2. Hand Solution Flowchart

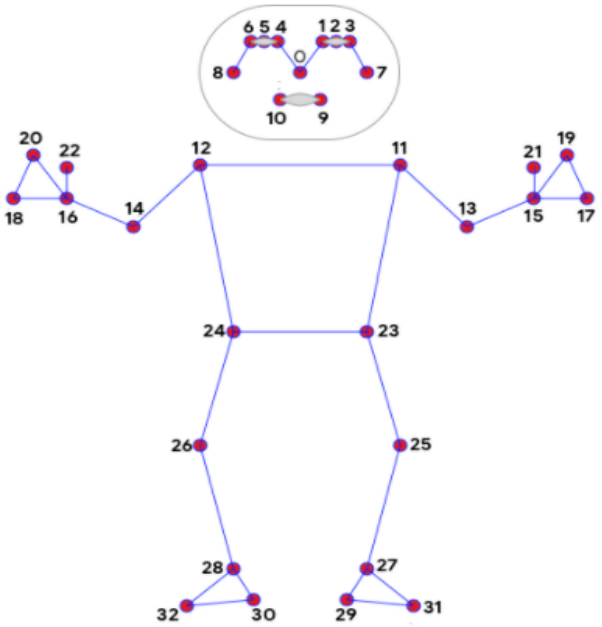


Fig. 3. Pose Landmarks keypoints

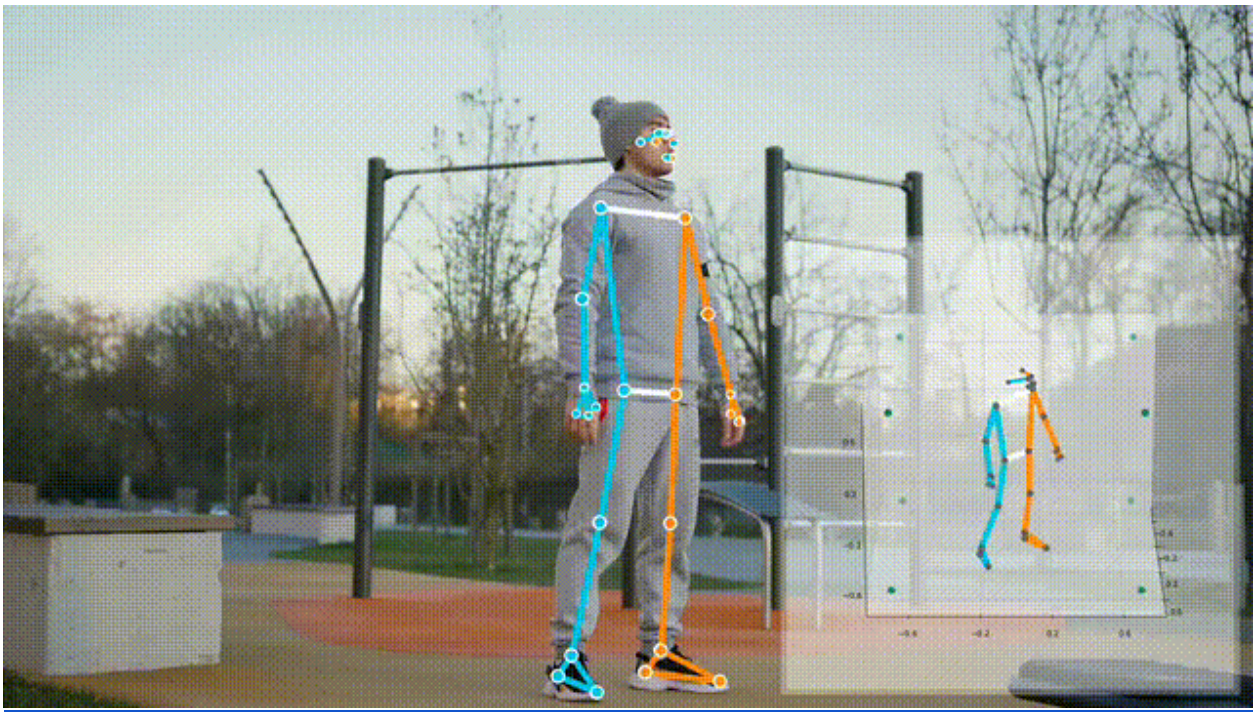


Fig. 4. Pose Landmarks key points Visualization

C. Setup Folders for Collection:

As a part of storing our data, we need a folder to hold our input values. A folder named MP_Data has been created with 3 sub folders within it that will hold our actions(3 actions here) and this may depend on the number of actions that we need to use. Each folder here will contain 30 values(frames) i.e 30 frames for each action.

D. Collect Keypoint Values for Training and Testing(Dataset Generation):

The main device used as an input process in Sign Language Recognition (SLR) is the camera. The SLR input data is in the form of gesture (action) that can be easily captured by camera using OpenCV. We have captured 30 frames per second real-time video, which was then analyzed for dynamic gestures frame by frame.

E. Preprocess Data and Create Labels and Features:

After the data collection process we now need to name the collected input values w.r.t their actions and this is done by labelling. By using Label_map we label the actions generated with their respective names.

F. Build and Train LSTM Neural Network:

Sequential models have been used along with LSTM and Dense layers in the model building part. Relu activation

function has been used in the hidden layers and in the output layer softmax has been applied.

G. Make Predictions

Using the "model.predict" function we can manually predict how the system is working once the model is trained.

H. Save Weights

We can even save our trained model weights so that we can skip the training part everytime we run this system by using the"model.save" function that saves the weights in h5 format. "Model.load_weights" is used to reload the saved weights while re-running the system.

I. Evaluation using Confusion Matrix and Accuracy

A confusion matrix is a table that is used to define the performance of a classification algorithm. A confusion matrix visualizes and summarizes the performance of a classification algorithm and Accuracy score is the most intuitive performance measure of the system. An Accuracy score of 80% and above has been achieved in this system.

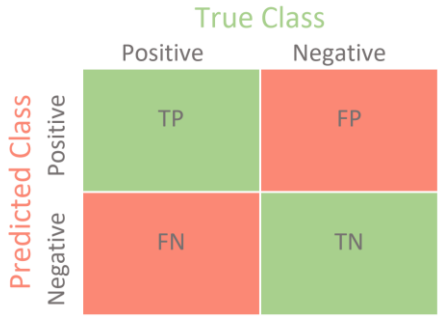


Fig. 5. Confusion Matrix representation

J. Testing in Real Time

This is the final phase where you check the working of the model. The system will now predict the sign language action based on the training provided and display the respective labels or outputs. Once the model is successfully built you can see the outputs along with their respective probability bar and occurrence at the top which helps in a better visualization.

Fig. 6. Detected Action fo YOUR

Fig 7 : Detected Action fo NAME

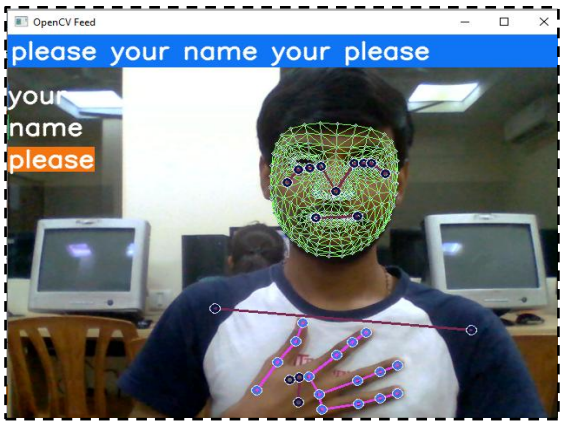
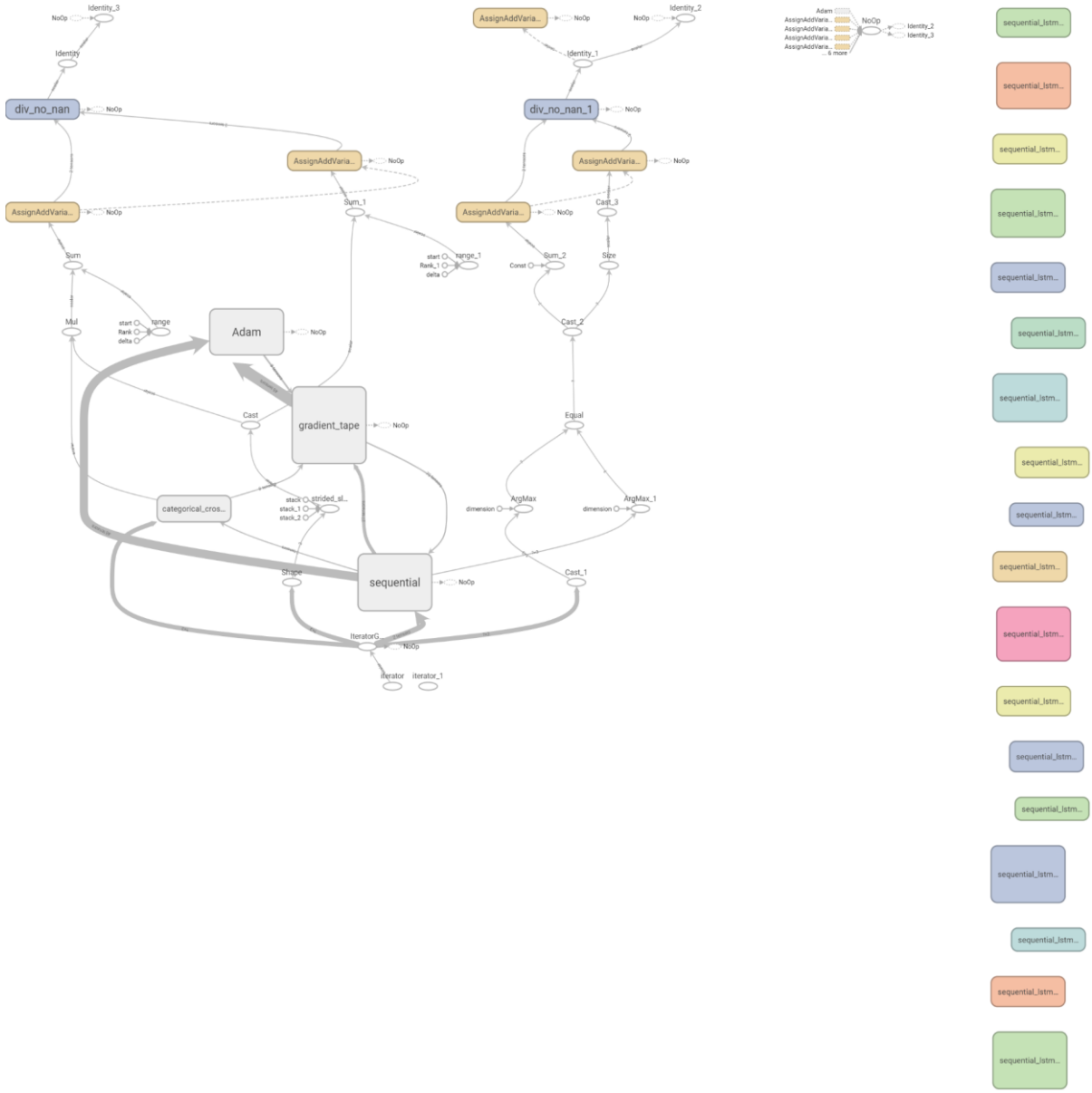


Fig. 8 : Detected Action for PLEASE

IV. NETWORK DIAGRAM



V. RESULTS & CONCLUSION

In this report, a functional real time action recognition for sign language for Deaf and Mute people has been developed for some common actions and words. We achieved a final accuracy of around 80.00%+ on our dataset. The model is accurately predicting the actions and is displaying the respective labelled results that have been fed during the training. As shown in Figures 5-7 the model has accurately predicted 3 actions performed which were "YOUR", "NAME" and "PLEASE" that actually points to the formation of the sentence "YOUR NAME PLEASE". Keypoints are also displayed along with the video feed so that the user can adjust the body part that is being displayed on the screen for generating better results. This system when integrated with existing SL models with advanced vision detection functionalities can be a game changer to detect real time actions and generate their labels accordingly in no time making is much easier to hold a strong conversation with Deaf and Mute people and will definitely act as a language bridge to overcome the existing communication gap among normal people and the deaf-mute.

VI. LIMITATIONS

The methods used in developing Sign Language Recognition are varied between the developers. Each method has its own strengths and limitations compared to other methods. There always is a scope of improvement in any system; and following are the drawbacks of the built model:

1. The implemented solution uses only one camera(laptop), and is based on a set of actions that we have trained and are , hereby defined.
2. The user must be within a defined perimeter area, in front of the camera.

3. The user must be within a defined distance range, due to camera limitations.
4. Limited actions/dataset.
5. A system with good speed, memory and GPU/TPU is required for the development of such models that have a large amount of dataset that has to be trained and tested.

VII. FUTURE SCOPE

The system generated here can be termed as a small scale model due to the system and processing limitations. If computed and processed on a large scale with state of the art technologies this system can do wonders. It can be implemented in any location where there's a need to interpret and communicate with deaf-mute people. The system can be used in public places, schools, hospitals, and various institutes where there's a need for such translations due to the lack of human interpreters. Television broadcasting systems can also use this model to generate real time news and will in turn save time. There are many other places as well where such a system is required for SL translation. May it be any place, this system will never go out of use because there always is a need for some service that has to be used as a communication bridge between the deaf-mute and normal people.

REFERENCES:

[1] Brill R. 1986. The conference of Educational Administrators Serving the Deaf: A History. Washington, DC: Gallaudet University Press.

[2] Banerji J. N. 1928. India International Reports of Schools for the Deaf. Washington City: Volta Bureau. Pp. 18-19

[3] Suryapriya A. K., Sumam S. and Idicula M 2009. Design and Development of a Frame based MT System for English to ISL. World congress on Nature and Biologically Inspired Computing. Pp 1382-1387.