Analysis and Design of Sign Language recognition Model using Machine Learning Techniques

1st Pratham Malhotra
CSE Department, ASET
Amity University Noida, Uttar Pradesh
Noida,India

3rd Shashank Garg
CSE Department, ASET
Amity University Noida, Uttar Pradesh
Noida,India

2nd Naman Jain
CSE Department, ASET
Amity University Noida, Uttar Pradesh
Noida,India

4th Deepak Gaur
CSE Department, ASET
Amity University Noida, Uttar Pradesh
Noida,India

Abstract— **It is quite evident from our routine lives to witness people having disorders related to Hearing and Speaking, respectfully referring to them as deaf and mute. Having these disorders make way to a lot of problems which have to be faced by those suffering from these. As a result of this, one of the most basic forms of survival rather, a form of support that is communication for these people is severely affected. The only mode of communication for these people remains to be the mode of gestures and expressions. Thus, this project on sign language recognition, we develop a sign detector that is easily expandable to include a huge variety of additional signs and hand gestures. This project at it's initial stage makes us of 3 sign namely, "hello", "thanks", and "iloveyou".**

Keywords— *Signs and Expressions, Computer Vision, Python, Machine Learning and Artificial Intelligence, MediaPipe, Feature Extraction, RNN and CNN, Keras, Face Detection*

I. INTRODUCTION

The agenda of this project is to help the people who are deaf and mute, and provide aid to the problem using the technologies of machine learning and computer vision, thereby, develop an interactive hand gesture detection user interface. For the interpreter too, the individual who is in talks with these specially abled person, it is a vital to appropriately recognize the signs conveyed in order to interpret correctly and help the needful. Signs are basically some sort of actions and symbolic representation of figures made with the use of body parts such as hand, fingers, arms, etc. [1]These also include expressions made using face, depicting some emotional response like anger, excitement, sorrow, happiness, etc. Numerous technological developments and extensive research have been made to support the deaf and mute population. The information that mute and deaf persons desire to express may be determined and predicted with the use of various machine learning algorithms. [2]The concepts of AI and ML thus are the basic backbone for such projects. The sign detection system starting from the basic can be implemented to a very advanced level. As it is very difficult for human to human (deaf or mute) interaction, machines with technologies can hereby can play a very vital role. Machines serving the disordered deaf and mute people can be installed at various public sites in order to assist those who are suffering from these disorders. These can eliminate the need of authorities to be present at every place to assist deaf and mute people. This concept should work similar to that of the Braille System which is used for people suffering from blindness. As there are various facilities for people suffering from blindness starting from Braille Assistance, the similar concept can be applied for deaf and mute people with respect to assistance. At places like airports, railway stations, etc. machines and devices running with technological sign detection assistance algorithms can help people who are unable to speak and hear. This also includes visual interpretation assistance as well. his project involves the understanding of sign languages which are the core of the model for recognition, so understanding of sign languages is essential.

II. LITERATURE REVIEW

A. Mediapipe

Mediapipe is a platform independent, open source ML based framework that is utilised to construct compound, multimodal ML pipelines. In addition to that it can be used to construct complicated Object Detection, face detection, multi hand tracking and many other models. The purpose of Mediapipe is to manoeuvre the models implementation on any platform such that the dev centre of attention could be at the model rather than its implementation on the system. Noteworthy features of Mediapipe includes Pose detection and tracking for humans, Hand tracing, Face Tracing, etc.

B. Mediapipe Holistic

MP holistic is complete package of pipelines that contains optimised component for hand face and pose that comes in handy for holistic tracking, thus allowing model to trace body and hand poses along with landmarks for head. The central use case of MP holistic in this iteration is to detect faces, hand to excruciate key point for passing it onto a Computer-Vision model.

C. Models

Landmark models make use of the hand, pose and face land marks model in MP pose, MP face-mesh, and MP hands to form a total of five hundred forty three landmarks in addition to that thirty three landmarks for pose, four hundred sixty eight landmarks for face and twenty one landmarks for

each hand. Hand Recrop Model is used in cases in which the pose model lacks accuracy such that ROI for hand are also inaccurate, there is a provision for running lightweight HRM as it serves the purpose of spatial transformer and fetches less than ten percent of inference time taken by hand model.

D. Tensorflow

This library has the purpose of performing numerical computation using data flow graphs. In which the nodes constitute mathematical operation whereas edges constitute arrays that are multidimensional in nature that is communicated among them.
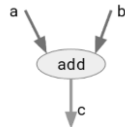

Fig. 1 Node depicting add mathematical operation

The node in the figure above represent the add operation, here the edges a and b are characterised as the input tensors and c can be characterised as the output tensor. It sports a flexible architecture that allows computation on multicore CPU's and GPU's on desktop.

E. Keras

It's python library that is used to progress models of neural networks. It is a High-Level API wrapper. It also has the possibility of executing on top of TensorFlow, Theano and CTNK. The motive of Keras is to simplify and fast forward the development of Neural-Network models. Noteworthy components of Keras include Sequential Model, Dense Layer, optimisers to compile the model etc.

F. Long Short Term Memory Model

Fig2. Depiction of LSTM Model
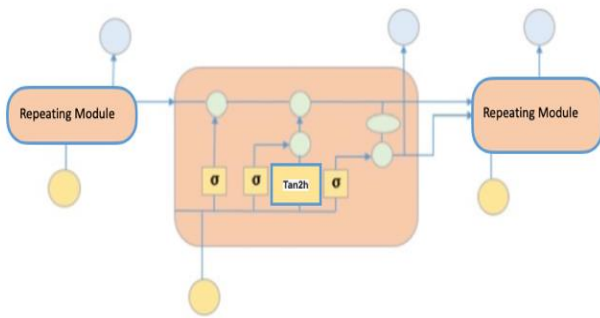
It's a specific class of Recurrent Neural Network that poses the capability of memorizing dependencies in data for long term. This feature is achieved because The module responsible for recurring contains combination of 4 layers that is capable of interacting with each other. The diagram depicted above shows 4 Neural Network layers that are mentioned in boxes that are yellow in colour, in green circles mentioned are the point wise operators. In yellow circles mentioned are the inputs and in blue color mentioned are the state of the cells. The differentiating feature of

LSTM is that it features cell state and 3 gates that gives them capability to learn selectively, unlearn or from each units retain certain information.

G. Open Source Computer VisionLibrary:

Open CV is the acronym for Open Source Computer Vision Library. [2] That is used in Python, C and C++ programs, for giving the computers capability to visualize and process the images to extract useful data. The library comes bundled with over 2000 sophisticated algorithm that possess the capability rendering three dimensional models, tracking human movements, face detection, object detection, rendering HD images by tacking up small images and many more.

H. NumPy Library:

This library is bundled with the Open CV library and is used to perform numerical operations on special arrays termed as NumPy arrays. The computer stores the pixels in the form of two dimensional arrays, that are converted to NumPy arrays for processing.

I. Sign Language

It is a form of language that utilizes visual-manual modality to bear meaning. It is form of a language in which use combination of manual articulation and non manual markers. [3] It possesses its own lexicon and grammar. However these languages are not universal and in most cases it's mutually intelligible. However similarities do exist among different sign language. Sign language is used in majority by people those are deaf, cannot hear properly and are unable to speak physically.

III. METHODOLOGY

A. Data Collection and Preprocessing

To collect the data to build this application we make use of Open CV python library. [4]The sole purpose of it is to capture the live video feed using the built in webcam in the device. To identify the key points present in the human body parts we configure the MediaPipe Holistic model and the drawing utilities from MediaPipe. Which is then followed by configuring a function capable of pre-process the image feed from Open CV to make it compatible with the media pipe model. This specifically involves changing the format of image from BGR TO RGB. Then the MediaPipe holistic model is used to make detection on each frame followed by storing the results of detection onto a variable. Drawing utilities from MediaPipe are utilised to draw the detected key points onto the feed from Open CV. To further the data collection process we define a function to extract the key points for the pose, face and hands from the result of these detections. This is followed by setting up multiple empty folders where data for each actions will be stored. There exist a separate folder for each action which in turn contain 30 subfolders. Each subfolder contains a video of 30 frames in

length for each action. [5]A loop statement is used to collect the data for each action. For each action the system loops through the 30 subfolders and for each subfolder through 30 frames. For each iteration of the loop a frame is captured from the live webcam feed. Detections are made using Mediapipe holistic model. The detected key points are drawn on the feed. Followed by extracting the key points and saving it onto a bumpy array. Which is then saved to a file in the respective subfolder.

B. Model Construction

The pre-processed videos consisting of 1662 key points captured in the data collection phase are used as inputs to train the model in this iteration. [6]This data is further split into testing and training sets. The model is built using the Keras Python Library. The architecture of this model comprises of Three LSTM layers followed by two fully connected layers and a final SoftMax layer that aids in multi-class classification. To compile the model the loss function termed as categorical cross entropy is utilised along with Adam optimiser.

C. Training the Model

The machine learning model is trained by making use of the pre-processed data for 2000 epochs using the fit() function. [7]To monitor the training of the model TensorBoard call-backs are utilised.

D. Model Evaluation

After training is completed, the model is evaluated on the testing set to determine its capability to classify different actions this value is quantised in the form of accuracy. [8]For this iteration a total accuracy of approximately 92% has been achieved.

E. Deployment of Model

The final step is to deploy the trained model such that it can be used in real life applications.

IV. METHOD DIAGRAM
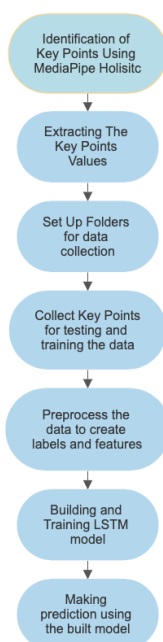
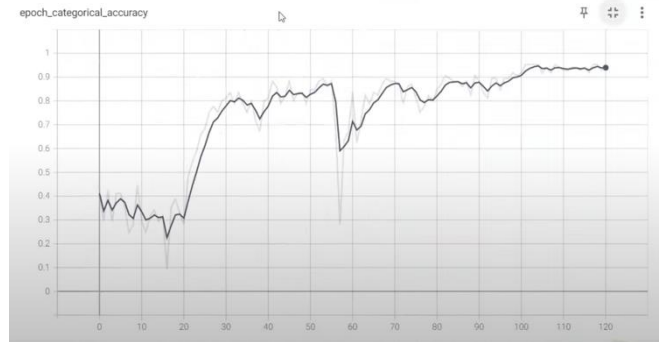
V. RESULTS AND DISCUSSIONS

Fig. 3 Plot of Categorical Accuracy
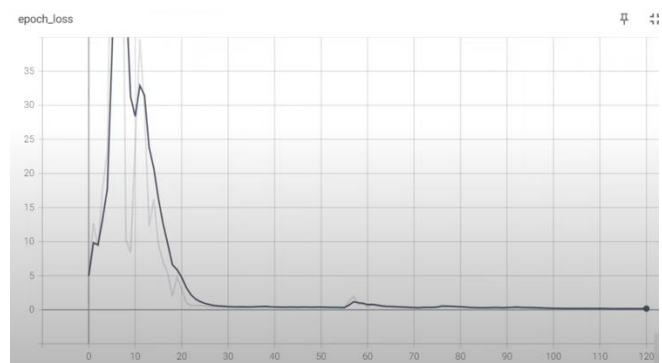X- Axis: Accuracy, Y-Axis: Epochs

Fig. 4 Plot of Epoch Loss
X-Axis: Losses, Y-Axis Epochs

Upon training the model an accuracy of 92% is achieved along with losses close to 8%. Indicating the model has promptly achieved high level of accuracy in classifying different signs accurately. Also, the low loss indicated the model is sufficiently well trained.

[9]The future scope of the project might involve investigating more sophisticated deep learning models to further enhance the accuracy of the model along with reducing the training time. Another potential direction is to further enlarge the dataset to include more diverse sign actions. [10]Finally developing a user friendly interface to easily interact with the model could broaden its potential applications.

VI. CONCLUSION

The main aim for sign language detection system is to give a practical way so that specially abled person can also have a communication with a regular person by using hand gestures. The project which we are building uses the web camera which is installed in the system itself and by using that it will detects the signs for recognition. From the project result, we can came to a conclusion that normal people can talk to specially abled person using this project. In future more gestures can be added to the project so that the specially abled person have more signs to convey to the normal people. As a result of which this project can be further extended on a large scale by adding the required datasets.

VII. CITATIONS

[1] Reddygari Sandhya Rani , R Rumana , R. Prema, 2021, A Review Paper on Sign Language Recognition for The Deaf and Dumb, International Journal Of Engineering Research And Technology(IJERT) Volume 10, Issue 10 (October 2021),

[2] Suharjito, Ricky Anderson, Fanny Wiryana, Meita Chandra Ariesta, Gede Putra Kusuma, Sign Language Recognition Application Systems for Deaf- Mute People: A Review Based on Input-Process-Output, Procedia Computer Science, Volume 116, 2017, Pages 441-448, ISSN 1877-0509,

[3] 3 https://blog.ai-media.tv/blog/7- reasons-sign-language- is-awesome#:~:text=Important%20f or%20Deaf%20people&text =Sign%20languages%20are%20the%20native,hear%20but%2 0can't%20speak.

[4] L. Boppana, R. Ahamed, H. Rane and R. K. Kodali, "Assistive Sign Language Converter for Deaf and Dumb," 2019 International Conference on Interner of Things(iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and social Computing(CPSCom) and IEEE Smart Data, 2019, pp. 302-307,dor10.1109/iThings/GreenCom/CPSCom/SmartData.2019.00071

[5] International Journal of recent Technology and Engineering(IJRTE) ISSN: 2277-3878, Volume-8 Issue-5, January 2020 5592 Published By: Blue Eyes Intelligence Engineering and Sciences Publications Retrieval Number: E6555018520/2020©BEIESP DOI: 10.35940/ijrte.E655.5018520 Translation of Sign Language for Deaf and Dumb People Suthagar S., K. S. Tamilselvan, P. Balakumar, B. Rajalakshmi, C. Roshini

[6] Ade, Shubham & Andurkar, Manas & Bobade, Umakant & Dawalbaje, Somesh & Kapse, Prof. (2023). Hand Gesture Recognition System for Deaf and Dumb People. International Journal of Advanced Research in Science, Communication and Technology. 558-561. 10.48175/IJARSCT-8538.

[7] Ismail, A.P. & Aziz, Farah & Kasim, Nazirah & Daud, Kamarulazhar. (2021). Hand gesture recognition on python and opencv. IOP Conference Series: Materials Science and Engineering. 1045. 012043. 10.1088/1757-899X/1045/1/012043.

[8] Azlin, Azra & Pang, Ying & How, Khoh & Ooi, Shih Yin. (2023). Hand Gesture Signature Recognition with Machine Learning Algorithms. 10.1007/978-981-19-8406-8_30.

[9] John, Jogi & Deshpande, Shrinivas. (2023). A Comparative Study on Challenges and Solutions on Hand Gesture Recognition. 10.1007/978-981-19-8493-8_18.

[10] Kotavenuka, Swetha & Kodakandla, Harshitha & Krishna, Nimmakayala & Rao, Dr. (2023). Hand Gesture Recognition. International Journal for Research in Applied Science and Engineering Technology. 11. 331-335. 10.22214/ijraset.2023.48557.