IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798

# INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

## IN COMPUTER & COMMUNICATION ENGINEERING

Volume 10, Issue 5, May 2022

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.165

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |

|| Volume 10, Issue 5, May 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1005249 |

# Sign Language Recognition System Using LSTM Model

**Rapaka Anusha[1], R.Venkata Sai Keerthi[2], Sk.Shajia Farheen[3]**

UG Student, IV Year Dept. of I.T., Vasireddy Venkatadri Institute of Technology, Guntur, India[1]

UG Student, IV Year Dept. of I.T., Vasireddy Venkatadri Institute of Technology, Guntur, India[2]

UG Student, IV Year Dept. of I.T., Vasireddy Venkatadri Institute of Technology, Guntur, India[3]

**ABSTRACT**: Sign language is the preferred method of communication among the hearing and speech impaired people all over the world. Sign language is said to have a structured set of gestures in which each gesture is having a specific meaning.One should learn sign language tointeract with them.Recognition of sign language can have varying degree of success when used in a computer vision. The main objective of our project Sign Language Recognition is to bridge the gap between the hearing/speech impaired people and a normal person by converting the gesture or action enacted by them into text that can be understood by all. To achieve this we are using a camera based computer system that uses openCV and is trained using a neural network in order to recognize and translate the signs. Our project uses Media Pipe Holistic to identify the face, hand and pose movements by the person which form a gesture and use this data to build long short term memory (LSTM) layers which form the neural network that learns these signs and labels them. This neural network is used to identify various gestures and the result is displayed in the form of text on the screen.

**KEYWORDS**: Sign Language Recognition, gesture, camera, MediaPipe, LSTM

## I. INTRODUCTION

In Sign language is the preferred method of communication among the hearing and speech impaired peopleall over the world. One should learn sign language to interact with them. There are very few study materialsavailable for sign learning. Because of this, the process of learning sign language is a very difficult task. Most of the existing tools for sign language learning use external sensors which are costly. Our project aimsat extending a step forward in this field by collecting a dataset and then use various techniques to extract usefulinformation to classify them into sign language gestures.

Recognition of sign language can have varying degree of success when used in a computer vision. Sign language is said to have a structured set of gestures in which each gesture is having a specific meaning. Image classification and deep learning can be used to help computers recognize sign language, which could then be interpreted by other people.

Long Short Term Memory (LSTM) neural networks can be employed to recognize sign language gestures. The video dataset used will consist of dynamic sign language gestures captured on a web camera. Preprocessing will be performed on the captured video, which then serve as the cleaned input. The results are obtained by retraining and testing this dataset on a Long Short Term Memory (LSTM) neural network model. This project aims to narrow the communication gap by developing software which can predict the gestures in real time. Some standard gestures are recognized and are converted into text format for the understanding of normal people.

## II. LITERATURE SURVEY

The researches done in this field are mostly done using a glove based system. In the glove based system, sensors such as potentiometer, accelerometers etc. are attached to each of the finger. Based on their readings the corresponding alphabet is displayed. Christopher Lee and Yangsheng Xu developed a glove-based gesture recognition system that was able to recognize 14 of the letters from the hand alphabet, learn new gestures and able to update the model of each gesture in the system in online mode. Over the years advanced glove devices have been designed such as the Sayre Glove, Dexterous Hand Master and Power Glove.

The main problem faced by this gloved based system is that it has to be recalibrate every time whenever a new user on the finger-tips so that the finger tips are identified by the Image Processing unit. The major disadvantage of these

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |

|| Volume 10, Issue 5, May 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1005249 |

electromagnetic devices is that their price is too high. The users' experience operating these devices, may be unpleasant because it is complex to handle and operate these devices. These devices use sensors which are very sensitive for long term use.

## III. METHODOLOGY

➢ **Dependencies:**

- **tensorflow**: It is an open source artificial intelligence library, using data flow graphs to build models. It allows developers to create large-scale neural networks with many layers. Tensor Flow is mainly used for: Classification, Perception, Understanding, Discovering, Prediction and Creation.
- **Opencv-python**: OpenCV is a great tool for image processing and performing computer vision tasks. It is an open-source library that can be used to perform tasks like face detection, objection tracking, landmark detection, and much more.
- **Mediapipe**: Mediapipe is a cross-platform library developed by Google that provides amazing ready-to-use ML solutions for computer vision tasks.
- **Sklearn:** The sklearn library contains a lot of efficient tools for machine learning and statistical modeling including classification, regression etc.

➢ **VideoCapture:** The gestures of the person are captured by using the web camera and are processed using the openCV.
This processed video is sent for the detection of landmarks by the mediapipe.

➢ **Landmarks Detection by MediaPipe:** The mediapipe uses the mediapipe holistic model which has prebuilt connections like the HAND_CONNECTIONS, POSE_CONNECTIONS AND FACE_CONNECTIONS to connect the various landmarks on the person inorder to interpret the gesture.

➢ **Extract Key Points:** The results from the media pipe detection are iterated and the individual points that give us the pose landmarks, left hand landmarks and the right hand landmarks are processed, concatenated and stored in a numpy array which will be sent as the input to our neural network.

➢ **Build Neural Network:**We are using Long Short Term Memory (LSTM) layers to build our neural network. The first step is to instantiate the model type. We have used Sequential model in our project as the input data is in the form of video which is a continuous data. Using Adam optimizer, categorical loss function and categorical accuracy as the metrics we have compiled the neural network and saved the weights.

➢ **Sign Prediction:**The last layer of the neural network uses the softmax function which adjusts the output in between the values 0 and 1. A value is generated for each of the input gesture. The gesture corresponding to the maximum value is chosen as the sign and is displayed as text on the screen.

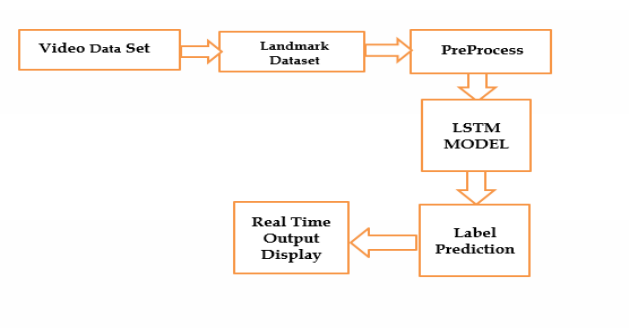



Fig a. Training Architecture

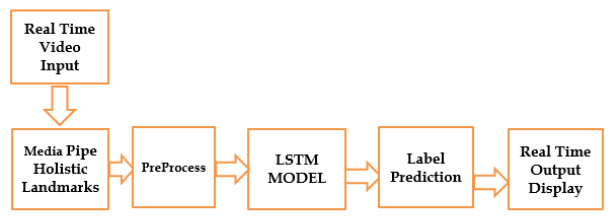



Fig b. Testing Architecture

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |

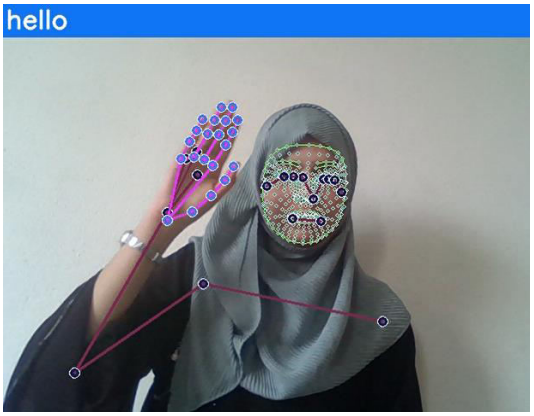
|| Volume 10, Issue 5, May 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1005249 |

## IV. RESULTS & DISCUSSION

We have considered five signs “Hello”, “Thank You”, “Welcome”, “Teacher”, “Indian” of the Indian Sign Language and trained our neural network model using media pipe and LSTM layers. The following are the output screenshots:


Output 1 a. Recognizing “Hello”

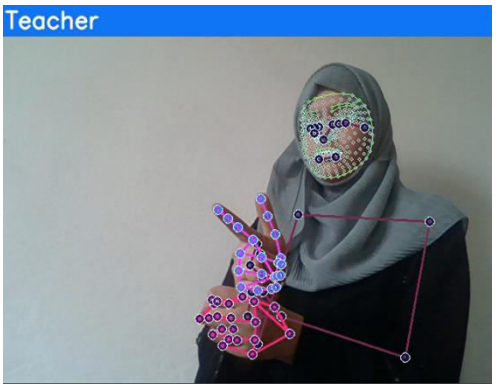



Output 1b. Recognizing “Hello”


Output 2a. Recognizing “Thank You”


Output 2b. Recognizing “Thank You”


Output 3a. Recognizing “Teacher”



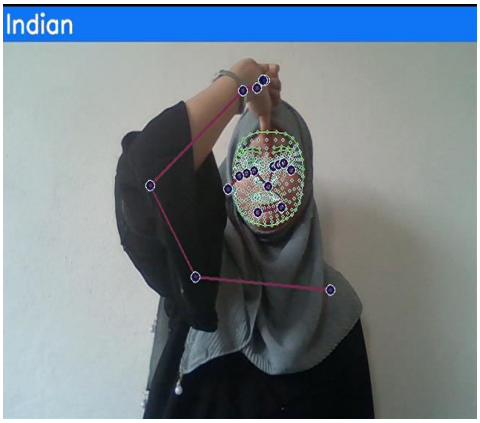

Output 3b. Recognizing “Teacher”

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |

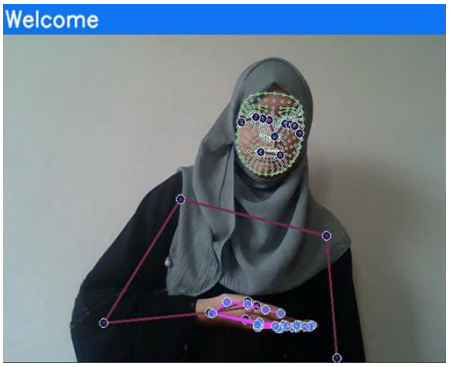
|| Volume 10, Issue 5, May 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1005249 |


Output 4a.Recognising “Indian”




Output 4b. Recognizing “Indian”




Output 5. Recognizing “welcome”

**Performance Measure:**

Model accuracy is a deep learning model performance metric that is defined as the ratio of true positives and true negatives to all positive and negative observations. In other words, accuracy tells us how often we can expect our model will correctly predict an outcome out of the total number of times it made predictions.

**Accuracy Score = (TP + TN)/ (TP + FN + TN + FP)**

Following are the accuracy scores observed for different Train and Test data splits:

| Train Data | Test Data | Accuracy Score |
|---|---|---|
| 60% | 40% | 0.66 |
| 70% | 30% | 0.78 |
| 80% | 20% | 0.87 |

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |

|| Volume 10, Issue 5, May 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1005249 |

**Advantages of using LSTM:**

The recognizer is a neural network that is built over sequential layers of the long short term memory (**LSTM**). When compared to other neural networks that use CNN, using LSTM has the following advantages:

- LSTM layers can be trained with less data. Hence, it does not require millions of neurons to be trained for executing the model.
- They can be used for training dynamic or time series data. Therefore we are using these layers in our project so that they can capture the gesture.
- More accuracy can be achieved with LSTM layers when compared to other layers like CNN which are mostly suitable for training with static data.

## V. CONCLUSIONS & FUTURE WORK

**Conclusion:**

The Sign Language Recognizer aims at bridging the gap between hearing/speech impaired person and a normal person. Our project takes the gestures of a person in the form of a video and displays the recognized sign on the screen. The MediaPipe over LSTM layers gives accurate results even with less data. The neural network build over the LSTM make faster detections when compared to other algorithms that use layers like CNN. Our project is a reliable sign language recognizer, as it uses the dynamic gestures rather than recognizing signs from static images. We have recognized five signs “Hello”, “Thank You”, “Welcome”, “Indian”, “Teacher” in our project. This project can be extended by training the system with more signs and can be developed into an application in future.

**Future Work:**

Standard dataset not available for all countries/sub continents / languages. A need for large vocabulary database is the demand for current scenario. This project can be developed into an application to be used by speechless people in order to ease their lives, and also government offices that should serve all of its citizens equally, private companies that want to reach and serve speechless people as well, co-operations and foundations which aims to help speech-disordered people. We can develop a model for ISL (Indian Sign Language) sentence level recognition. We can also focus on converting the sequence of gestures into the speech that can be heard.A more accurate application that can be reliable can be developed that acts as a perfect translator for sign language.

## REFERENCES

1. JOYEETA SINGH, K. D. Indian sign language recognition using Eigen value weighted euclidean distance based classification technique. International Journal of Advanced Computer Science and Applications 4, 2 (2013). NEHA V. TAVARI, P. A. V. D.
2. Indian sign language recognition based on histograms of oriented gradient. International
3. Journal of Computer Science and Information Technologies 5, 3 (2014), 3657–3660.
4. NMANIVAS. Gesture recognition system https://github.com/nmanivas/ Gesture-Recognition-System
5. PADMAVATHI. S, SAIPREETHY.M.S, V. Indian sign language character recognition using neural networks. IJCA Special Issue on Recent Trends in Pattern Recognition and Image Analysis, RTPRIA (2013).
6. SAKSHI GOYAL, ISHITA SHARMA, S. S. Sign language recognition system for deaf and dumb people. International Journal of Engineering Research Technology 2, 4 (April 2013).
7. SCHAKENBERG Schakenberg code for bag of visual words on github.https://github.com/shackenberg/Minimal Bag-of-Visual-Words-Image- Classifier
8. XIAOLONG TENG, BIANI WU, W. Y., AND LIU, C. A hand gesture recognition system based on local linear embedding, April 2005.

## BIOGRAPHY

**B.Lakshmi Praveena**is an Assistant Professor in the Department of Information Technology, Vasireddy Venkatadri Institute of Technology. She received Btech degree from ANU, Guntur, India in 2010 and Mtech from VVIT, Guntur,India in 2013. She is currently pursuing Ph.D in Data Analytics from SPMVV,Tirupati.

INNO SPACE
SJIF Scientific Journal Impact Factor

Impact Factor: 8.165

doi cross ref

ISSN INTERNATIONAL STANDARD SERIAL NUMBER INDIA

निस्केयर NISCAIR

# INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

## IN COMPUTER & COMMUNICATION ENGINEERING

9940 572 462 6381 907 438 ijircce@gmail.com

www.ijircce.com

Scan to save the contact details