iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 11 **Issue:** V **Month of publication:** May 2023

DOI: https://doi.org/10.22214/ijraset.2023.53085

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue V May 2023- Available at www.ijraset.com

Two Way Indian Sign Language Translator using LSTM and NLP

Ch. Rakesh[1], G. Madhumitha[2], S. Meghana[3], T. Sahithi Niharika[4], M. Rohith[5], Ajay Ram K[6]
[1]*Assistant Professor(c),* [2,3,4,5,6]*Students, Department of Computer Science, University College of Engineering Narasaraopet*

Abstract: ***Sign language is the efficacious medium that connects silent people and the world, and there are many significant existing sign languages across the globe. A lot of research is done to compromise the borders of difficulty in the communication between silent and normal person and most of them are based on ASL (American Sign Language). The aspiration for bridging the gap between silent people and normal people in terms of communication using ISL (Indian Sign Language) led to unfold this project. It functions as a two-way sign translator that is the conversion of sign to text and contrariwise. It recognizes various poses as well as gestures and returns appropriate results. The designed translator predicts the sign with an accuracy of 88 percent in real-time and was trained to recognize 15 actions using LSTM and MediaPipe. The text to sign translator works up to paragraph level using NLP.***

Keywords: ISL (Indian Sign Language), Long Short Term Memory (LSTM) , Lemmatization, OpenCV, MediaPipe , NLP , Silent people (Speech and Hearing disabled)

I. INTRODUCTION

There are approximately 70 million silent people across the world and most of them are illiterate and communicate using signs. The nonverbal communication used by silent people is called sign language, and it is comprised of various gestures, each gesture has its meaning. These gestures are used to express their thoughts and feelings. There is a wide range of sign languages each having their own importance and significance. In Indian sign language both hands along with facial expressions are used to express a gesture. These gestures are to be understood by vision. The performance of the recognition system is a significant factor that cannot be ignored. And the performance mostly depends on the classifier as also the feature extraction method. So, a careful combination of classifier and extraction methods is to be considered to obtain optimal solutions.

The proposed model is a two-way translator and has two modules, one each for the translation of sign to text and the translation of text to sign. In the text to sign module, the text is taken from the user, it is processed using NLP, and the respective sign video is returned. And in the sign to text module, the input can be either uploading a file or enacting before the webcam, the appropriate text for the sign is displayed using LSTM, MediaPipe and OpenCV. In order to communicate signers mostly use dynamic signs. To translate the dynamic gestures with time-series video we have chosen LSTM as it is good at processing sequences of data. This proposed model can translate up-to word level in the sign to text module and up-to paragraph level in the text to sign module.

Our work is predominantly on Indian sign language as it is mostly used sign language in our nation and research to a great extent is done and yet there is no proper data set or fixed stature for the Indian sign language. There are many factors that influence the performance and accuracy of the model like hand movement, illumination, pose, occlusion, background information and noise. These also effects in the building of a robust system that could give high accuracy in the real time.

II. LITERATURE REVIEW

This literature survey states various existing systems on the translation of the sign to text and vice versa using various methods. These works are broadly classified into two kinds one is hardware based and the other is vision based. In hardware-based model a hardware is used as a medium. In vision-based camera is used as a medium and respective model is built.

A. Hardware Based

In [1]. Ebey Abraham and team worked on a wireless Data Gloves that are designed for sign language. This glove uses a combination of flex sensors and gyroscope and an accelerometer to read data. Bluetooth is used to transmit the data to the respective device. The main drawback of this work is that one always must wear a glove and if there are any hardware failures it would give wrong predictions

B. Vision Based

In [2] Britta and team described relevant features for video recognition which is based on continuous density Hidden Markov Models (HMM). Limitations of their work is that the sign is affected by the preceding and the subsequent sign, Movements of the finger, like shifting in one direction or rotating around the body axis, must be considered which is difficult to consider and sign boundaries are automatically deleted. Wuyang Qin and his team developed a model for sign language in [3]. This model is a recognition framework based on VTN. This model takes an isolated word as an input and trains the model and saves it for later use. The continuous extraction module is the sign language module that stores the relevant data in the model. Limitations of this work are shortcomings in the recognition of similar signs or actions and the extraction of long-sequence key frames needs to be improved. In [4]. Ankita Wadhawan and his team built a model based on CNN architecture for communication between speech-impaired and normal people. This model has a total of 4 phases which are data acquisition, image processing, training, and testing the CNN classifier. In the data acquisition phase, the data is collected using a camera and in the next phase the collected data is processed in the later phase, the processed data is trained using the CNN classifier lastly the results are obtained. The limitations of this work are that it is confined to recognise only static images.

In [5]. Poonam Yerpude and her team worked on a sign detector model for the detection of ISL alphabets and numbers (from 1 to 10) . CNN algorithm and OpenCV are used throughout this system. The main drawbacks of this work are that it is trained under certain conditions like camera apparatus and lighting conditions, and the model is only limited to recognizing static alphabet signs and number signs. Advaith Sridhar and his team worked on large datasets in [6]. A single-color video camera is used for image recording. The size and quality of the dataset enable the exploration of deep models for Sign Language Recognition on ISL. They also presented a comparison of multiple deep-learning models and identified a model that achieves high accuracy. The limitations are that they have used an open pose that has about 135 key points for the extraction of data. Because of low key points, this cannot give more information about facial expressions, but facial expressions also play a key role in Sign Language. All the existing systems are limited to recognize the alphabet, number and static image level signs and have shortcomings like background conditions , using hardware all the time.

III. METHODOLOGY

This system has two modules one can translate text to sign and the other sign to text , the research is mainly focused on facilitating communication between normal and silent person.

A. Text to Sign

Fig. 1 Workflow of Text to Sign Translation

The text to sign conversion is done using Lemmatization in NLTK library where the given text is processed and base words are fetched. Initially, the given text is split into words by using the inbuilt tokenizer in NLTK library. Then the words are processed using the Lemmatization procedure to fetch the base words. The database is searched for respective sign videos. These fetched sign videos are concatenated using the inbuilt function in python MoviePy library. Finally, the concatenated video is displayed as output. The process is depicted in Fig. 1.

C. Sign to Text

1) *Dataset Collection:* We have made dataset for 15 sign actions by collecting 50 videos per each sign in real time by using OpenCV to capture the video. A video is divided into 20 frames and in each frame MediaPipe holistic view is used to detect 33 keypoints in pose, 468 keypoints on face and 22 keypoints on each hand. These extracted keypoints are stored in the form of NumPy arrays one array for each frame of a video. As ISL uses both hands to sign and facial expressions also play important role in the meaning of a sign MediaPipe holistic view helps to differentiate between signs in Indian Sign Language more efficiently. The 15 action level words and Phrases are :(absent, day, Good morning, Green, hearing, How are you, I don't understand, maths, Maximum, sign, Take a photo, Talk, Thank you very much, time, up).

2) *Pre-processing and creating labels for the features extracted:* The data features extracted using Mediapipe are pre-processed initially and refined. Then dataset was labelled with the sign meaning where each folder consists of 50 extracted features one for each video further 20 NumPy arrays in each one for each frame of the video.
3) *Building the LSTM model and Training the model :* After dataset collection and pre-processing the data, LSTM is used to build the model by adding 4 LSTM and 4 dense layers , it is observed that adding the necessary number of layers as per the size of the dataset makes the model robust. In this system we have increased the number of fully connected LSTM and Dense layers in the model which made the model to learn the complex and abstract features of video sequences and give accurate predictions. By running epochs and passing the dataset to the LSTM model is trained, As shown in Fig. 2 the model has 732,815 parameters in total which are trainable. The Fig. 2 depicts the summary of the LSTM model , number of layers in it.

```
Model: "sequential"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 lstm (LSTM)                 (None, 20, 64)            442112

 lstm_1 (LSTM)               (None, 20, 128)           98816

 lstm_2 (LSTM)               (None, 20, 128)           131584

 lstm_3 (LSTM)               (None, 64)                49408

 dense (Dense)               (None, 64)                4160

 dense_1 (Dense)             (None, 64)                4160

 dense_2 (Dense)             (None, 32)                2080

 dense_3 (Dense)             (None, 15)                495

=================================================================
Total params: 732,815
Trainable params: 732,815
Non-trainable params: 0
_________________________________________________________________
```

Fig. 2 LSTM Model Summary

4) *LSTM Working:* LSTM is a neural network algorithm that is classified as an advanced RNN. When training the model the extracted features are given as input to the model, LSTM architecture consists of feedback connections which makes it capable of processing the entire sequence of data rather than single or static data points such as images. As shown in Fig. 3, when the features are passed to the model the first layer forgets the irrelevant data captured by comparing it with the data that it got in previous frames. This extracted new data is now added to the existing and later passed on to continue with the next frame. In this way LSTM works in processing the long sequences of data.

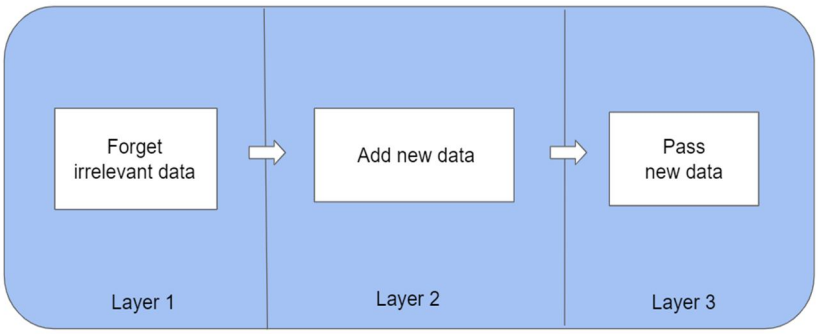


Fig. 3 LSTM Working and Architecture

IV. RESULTS

The outcome of our system is a website that has options to convert text to sign and sign to text. In text to sign, the input can be given either by uploading a file or entering the text, the given text is processed and the respective concatenated video is displayed as output. In Fig. 4 we have given a paragraph as input by uploading a file and corresponding sign videos are concatenated and displayed. The sample output is shown in Fig. 5.

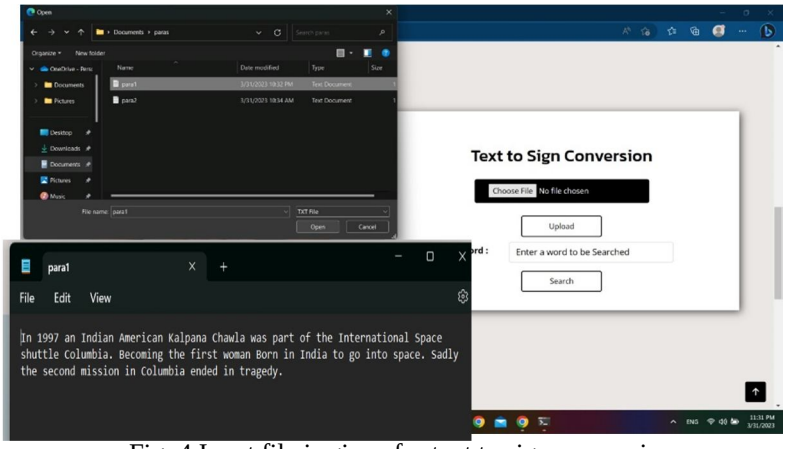


Fig. 4 Input file is given for text to sign conversion

Fig. 5 Corresponding sign language video is displayed

In sign to text conversion, the input can be given as a video or in real time user can sign before the system using the webcam. The system uses OpenCV to capture the video frames, continuously 20 frames are collected and, in each frame, key points are extracted using MediaPipe holistic pipeline. These values are used to predict the sign in real time. The sign to text model predicts the sign with an accuracy of 88 percent. Once, the sign is detected, the predicted phrase is displayed on the screen as in Fig. 6.

Fig. 6 Sign is detected as Good Morning in real-time

The results are also visualized using the confusion matrix in Fig. 7 as true labels versus predicted labels where the accurate prediction rate for the 15 actions is displayed.

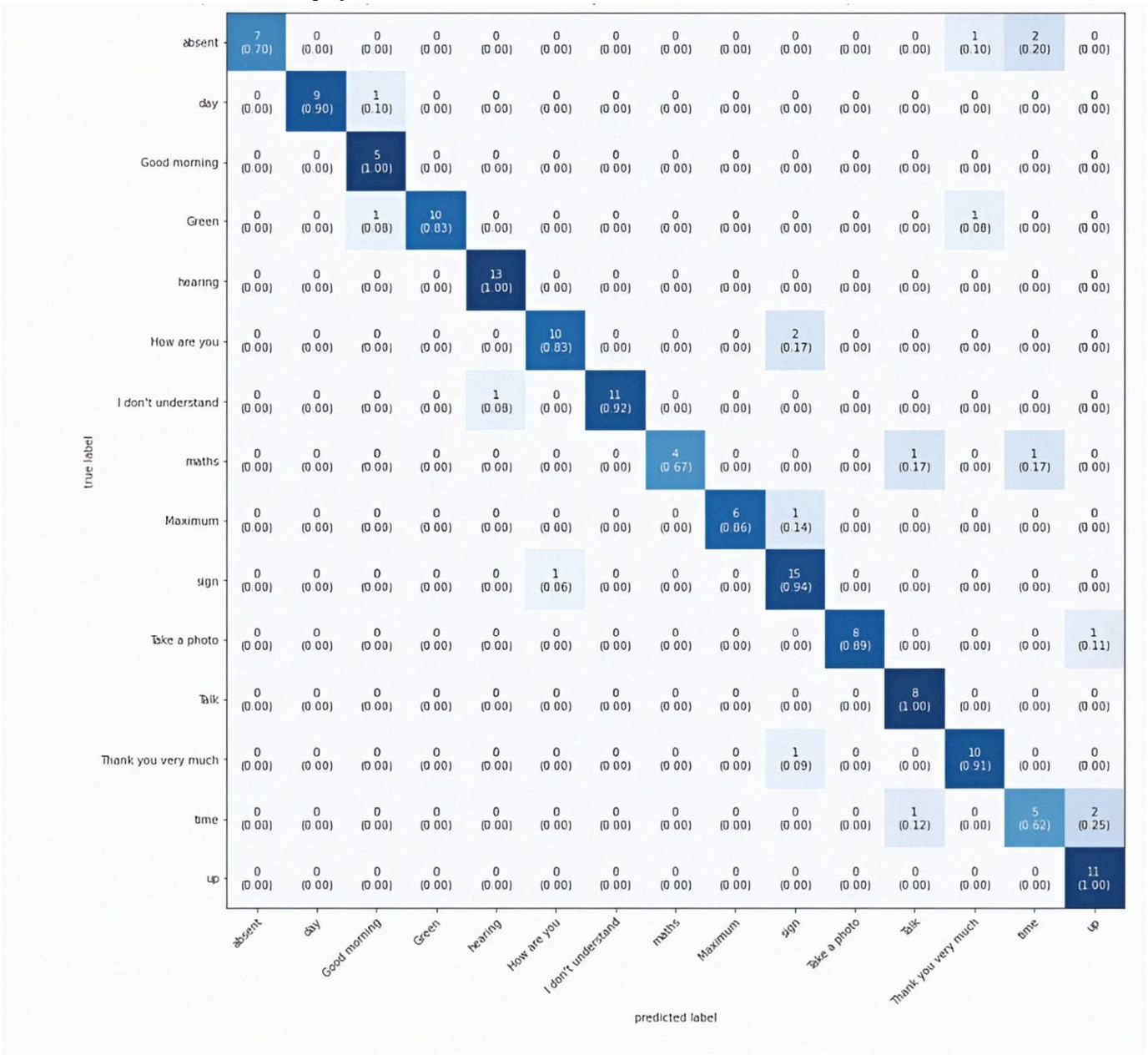


Fig. 7 Resulting Confusion Matrix for sign to text real time model

V. CONCLUSION AND FUTUREWORK

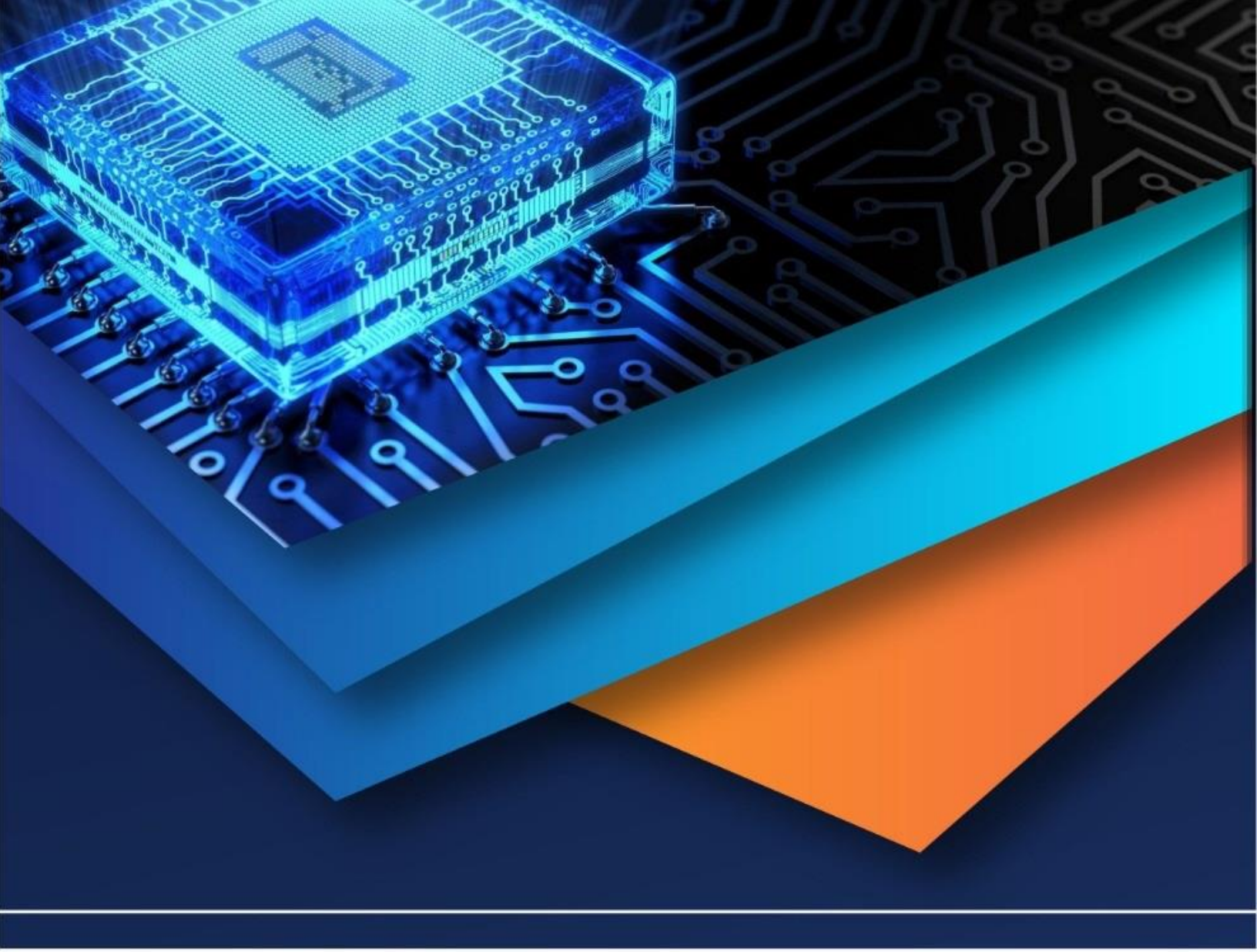
The proposed model is a two-way Indian Sign Language translator and has two modules, one each for the translation of sign to text and text to sign. The overall system can bridge the communication gap between silent and normal people. The web interface is user friendly and interactive. In the text to sign module, the text is taken from the user and undergoes lemmatization process and returns the appropriate sign for the text given. And in the sign to text module, we have made a custom dataset using OpenCV and MediaPipe and trained LSTM model to predict the sign meaning when sign video is given which predicts the sign with 88 percent accuracy. The proposed system can translate up-to word level i.e., can process video sequences in any background conditions in the sign to text module and up-to paragraph level in the text to sign module. Due to the limited size of the custom dataset, there are a few wrong predictions for similar signs while translating sign to text, the proposed model can be further enhanced by increasing the size of the dataset which will help the model to learn more efficiently.

VI. ACKNOWLEDGMENT

We extend our gratitude towards Ch. Rakesh sir, without whose motivation and guidance, this project would not have been possible. His constant efforts to enhance this system have proved to be invaluable throughout the project. We are also thankful to Dr. T. Siva Rama Krishna sir, HOD (Department of Computer Science and Engineering) for their valuable guidance.

REFERENCES

[1] Ebey Abraham, Akshatha Nayak, Ashna Iqbal “Real-Time Translation of Indian Sign Language using LSTM” in 2019 Global Conference for Advancement in Technology (GCAT) Bangalore, India. Oct 18-20, 2019.
[2] Britta Bauer; Hermann Hienz “Relevant Features for Video-Based Continuous Sign Language Recognition” , in Proceedings Fourth IEEE International Conference on Automatic Face and Gesture Recognition (Cat. No. PR00580) , Grenoble, France , 2002.
[3] Wuyang Qin; Yuming Chen; Yanyin Yao; Xue Mei; Qihang Zhang; Shi Hu “ Sign Language Recognition and Translation Method based on VTN” in 2021 International Conference on Digital Society and Intelligent Systems (DSInS) IEEE , Chengdu, China , 2021.
[4] Ankita Wadhawan ; Parteek Kumar “Deep learning-based sign language recognition system for static signs” Springer in 2020.R. E. Sorace, V. S. Reinhardt, and S. A. Vaughn, “High-speed digital-to-RF converter,” U.S. Patent 5 668 842, Sept. 16, 1997.
[5] Poonam Yerpude ; Pratiksha Jagat; Reema Sahu; Palash Dubey “Non-Verbal (Sign Language) To Verbal Language Translator Using Convolutional Neural Network” in IJRASET 2022.
[6] Advaith Sridhar, Rohith Gandhi Ganesan Pratyush Kumar Mitesh Khapra “A Large-Scale Dataset for Indian Sign Language Recognition” in Proceedings of the 28th ACM International Conference on Multimedia , New York, United States , October 2020.
[7] https://youtu.be/doDUihpj6ro
[8] https://zulko.github.io/moviepy/
[9] https://www.tensorflow.org/api_docs/python/tf/keras



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY