iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL
FOR RESEARCH
IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 11 Issue: IX Month of publication: September 2023

DOI: https://doi.org/10.22214/ijraset.2023.55621

www.ijraset.com

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

Real Time Sign Language Recognition Using Deep Learning

K. Mahimanvitha[1], Dr. M. Arathi[2]

[1]*Student of Department of Information Technology, Jawaharlal Nehru Technological University Hyderabad, University College of Engineering, Science & Technology Hyderabad*

[2]*Professor of Department of Information Technology, Jawaharlal Nehru Technological University Hyderabad, University College of Engineering, Science & Technology Hyderabad*

Abstract: Sign language is the only tool of communication for the person who is not able to speak and hear anything. Sign language is a boon for verbally challenged people to express their thoughts and emotion. Using this Sign Language Recognition system, the communication gap between people with hearing impairments and the general public can be cleared. In this work, a scheme of sign language recognition has been proposed for identifying the gestures in sign language. With the help of computer vision and neural networks, system can detect the signs and give the respective text as output. The major aim of this work is to build a neural network using a Long Short-Term Memory (LSTM) deep learning model using the video frames which offer the translation of gestures into text. The model is trained with the dataset that is collected using MediaPipe holistic key points from the video of the person which detects the pose, face and hand landmarks. after building the neural network, real-time sign language recognition is performed using OpenCV and a user interface is developed using Streamlit where the gestures are recognized and displayed as text within the highlighted section on the screen

Keywords: Long short-term Memory, OpenCV, MediaPipe, Streamlit.

I. INTRODUCTION

The necessity for communication has always been there right from the early ages of human existence. Language plays a vital role for communication. In order to express emotions, feelings, thoughts, and ideas language is used. The language can be expressed as verbal or non-verbal communication in the form of speech, symbols and gestures, expressions respectively. Although, Language has paved the way for communication, it does have challenges too. The first one is, at least one common language is needed for effective communication, so that everyone can understand and respond properly. And the other problem is that, human beings born with or developed with disabilities limit them from expressing and sharing their thoughts and ideas. The people who are hearing impaired and those who cannot speak are restricted to use gestures and sign language to communicate with others.

In general, need for communication between sign and no sign speakers is tried to satisfy using either text or translator. But both of them have their own problems, first takes a long time to convey and the latter one eliminates privacy. There is need to remove the barrier between the speech impaired people and others members of the society. In general speech or hear disabled people use gestures to communicate but those gestures may not be understood by others and it also takes lot of time to convey a small and simple message. An AI vision-based system that converts sign language to text or speech can be quite useful to solve the problem of communication between specially abled people and common people. This research is to develop a proficient deep learning model which is able to detect and understand the actions given by the user in British sign language. The developed model is then tested in real time to evaluate its performance in reality.

II. LITERATURE REVIEW

A. Overview

Sign language is a visual way of communicating using body gestures, facial expressions, hand signals and visual communication. It’s the mostly used type of communication used by the people with disability of hearing or speech. People with disabilities like autism spectrum disorder may also find sign language beneficial for communicating. The system will realize British Sign Language using a keypoint detection model. Sequence of keypoints are extracted. These keypoints can then be passed to an action detection model. The proposed system will be predicting British Sign Language signs using several frames and predicting what action is being demonstrated. The system will use MediaPipe Holistic to extract the keypoints. It helps in extracting keypoints from the user’s hands, from the user’s body, and the user’s face.

Then use Tensorflow and Keras to build an LSTM model which is able to recognize and predict the action from the live video. A deep learning model using LSTM layers which predict action from several frames from the live video at user end is developed. Then, put it all together, the MediaPipe Holistic and trained LSTM model using OpenCV is tested in real-time using the webcam.

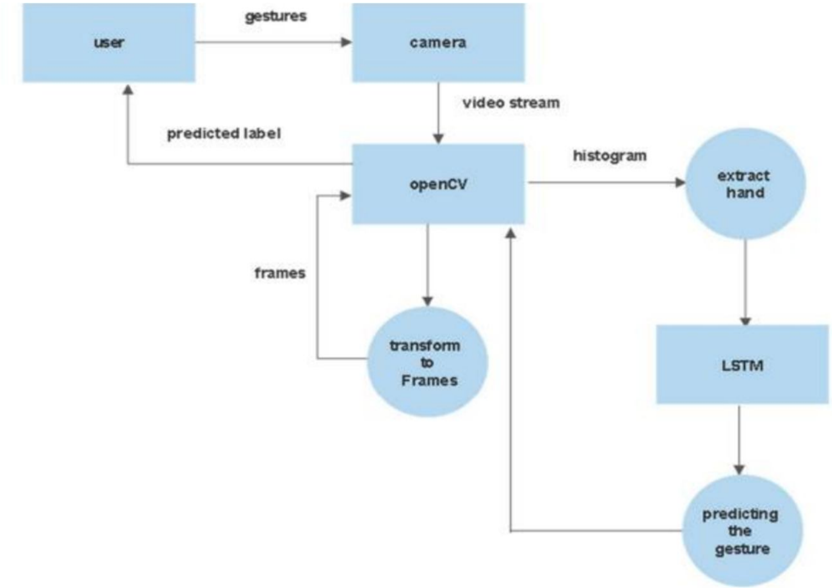


Fig 1. System Architecture.

B. *Existing System*

The existing system can recognize English alphabets based on the sign language symbols. Each English alphabet can be shown using hand symbols in sign language. These hand symbols are recognized by the system and predicts the corresponding alphabet through sign language classification. The classification of sign language symbols here is done using e Convolutional Neural Network (CNN). The model is trained using Convolutional Neural Network (CNN) after which it successfully predicts the corresponding alphabet of sign language. To train the model a customized dataset along with MediaPipe, OpenCV. After developing the model, it can be evaluated using classification accuracy, normalized and non-normalized confusion matrices and so on. The problem with the existing system is that, it recognizes only alphabet where the hand movement is restricted, due to which gestures or actions which represent words cannot be recognized. And sign language differs from country to country in gestures, facial expressions, and body language. Therefore, a system which recognizes actions instead of symbols is required.

C. *Proposed System*

Earlier developed models included detections of symbols, static signs using simple deep learning algorithms like Convolution Neural Networks. But detection of dynamic actions given by the user is necessary to develop a proficient model for sign language detection. Therefore, the proposed system is based on detecting user actions in real time by capturing the frames from the video of live stream. The model is trained over a Long Short Term Memory neural network after extracting keypoints from face, pose, and hand with the help of MediaPipe holistic. Initially the keypoints are extracted from live stream and are saved for training the model. Pre-processing of data, training, testing, evaluating the model are carried out sequentially. After building the model with good accuracy, sign language recognition is performed in real time through webcam where the user given actions are recognized by the system and displayed as text in highlighted section.

III. PROBLEM STATEMENT

Sign language which is used by speech-impaired people is not known to common people. When speech impaired people need to get services at public institutions from a public servant, communication barrier occurs as the public servant may not understand their need as they communicate through sign language. Due to this access or offer of public services to speech-impaired people are limited. So, in order to break the barrier, there is a need to develop a system where both can understand and promote effective communication.

IV. METHODOLOGY

A. Importing the Necessary Dependencies.

Firstly, we need to install and import all the required dependencies, the required modules include OpenCV, TensorFlow, Keras, MediaPipe, Sklearn.

B. Extracting Keypoints using MediaPipe Holistic.

MediaPipe already offers fast and accurate, yet separate, solutions for these tasks. Combining them all in real-time into a semantically consistent end-to-end solution is a uniquely difficult problem requiring simultaneous inference of multiple, dependent neural networks. MediaPipe Holistic pipelines combines individual models for face, hand and pose components. The model takes fixed resolution of (256X256) as input for given video frame. MediaPipe Holistic treats different regions. using appropriate image resolution according to the need. The MediaPipe Holistic Landmarker task lets you combine components of the pose, face, and hand landmarkers to create a complete landmarker for the whole human body. This can be used to analyze poses, actions, gestures. Therefore, in order to detect and recognize the action from user MediaPipe Holistic Landmarker is the most important one. Using a machine learning model on continuous sequential or stream of images, the task outputs 33 pose landmarks, 468 face landmarks, 21 hand landmarks per hand that is a total of 543 landmarks. Define a function to extract key points from the frames captured by webcam using MediaPipe Holistic.

.

C. Collecting and Pre-Processing Data

Create folder “action” which contains sub folders for each action hello, thanks, and I love you. Each action sub-folder further contains 30 sub-directories to store 30 videos from the user for training. So, there will be total of 30 multiplied by 3 i,e 90 folders to collect data , furthermore each video has 30 frames, each frame containing data of 1662 landmarks, which will be collected from the user during the run time for training. This data is sufficient to develop a good model.

The user poses for three different actions each thirty times that is hello action for thirty times, thanks action for thirty times and I love you for thirty times. During each action from user the keypoints are extracted and stored in their respective folders which is further used for training of the model. It is necessary to covert the available raw data into structured and clean data for machine learning algorithms to work efficiently.

Actions are mapped to numerical values using the above code. Hello: 0, Thanks: 1, I love you: 2 After labelling the data, next step is, to split the dataset into train set and test set. This Train Test Split is among the vital steps in Machine Learning as its important to evaluate the model before deploying it. The train set consists of training data and training labels and test set consists of testing data and testing labels. Training is done through the train set and the developed model is tested on the test data. In general, the test set value is about 20 percentage of the whole data. As the next incoming data is new and unseen, initially its tested-on test data and evaluate its performance using certain metrics.

D. Build and train LSTM model

Long Short-Term Memory is a type of Recurrent Neural Network that is especially designed to handle sequential data. The LSTM model addresses the issue of vanishing gradients in traditional Recurrent Neural Networks by introducing memory cells and gates to control the flow of information and a unique architecture. The neural network architecture consists of a visible layer with one input, a hidden layer with LSTM blocks (neurons), and an output layer that predicts a single value. Long Short-Term Memory neural networks utilize a 36 series of gates to regulate information flow in a data sequence. Using the Keras library makes easy to build and train deep learning Cell models. To build a model of three layers and three dense layers, import it from keras.models.

1) Initialize a Sequential model.
2) Add three LSTM layer to model.
3) Add three dense layers to the model.
4) Compile the model.
5) Fit the model.
6) Evaluate the model on test data.

E. Evaluation

After training the model, evaluate its performance on the training and test datasets to establish a baseline for future models. The following metrics are used to evaluate its performance.

1) Accuracy
2) Confusion Matrix
3) Precision
4) Recall

F. Test in Real Time

The built model is then used to predict the gestures. real-time sign language recognition is performed using OpenCV where the gestures are recognized and displayed as text within the highlighted section on the screen.

V. TESTING

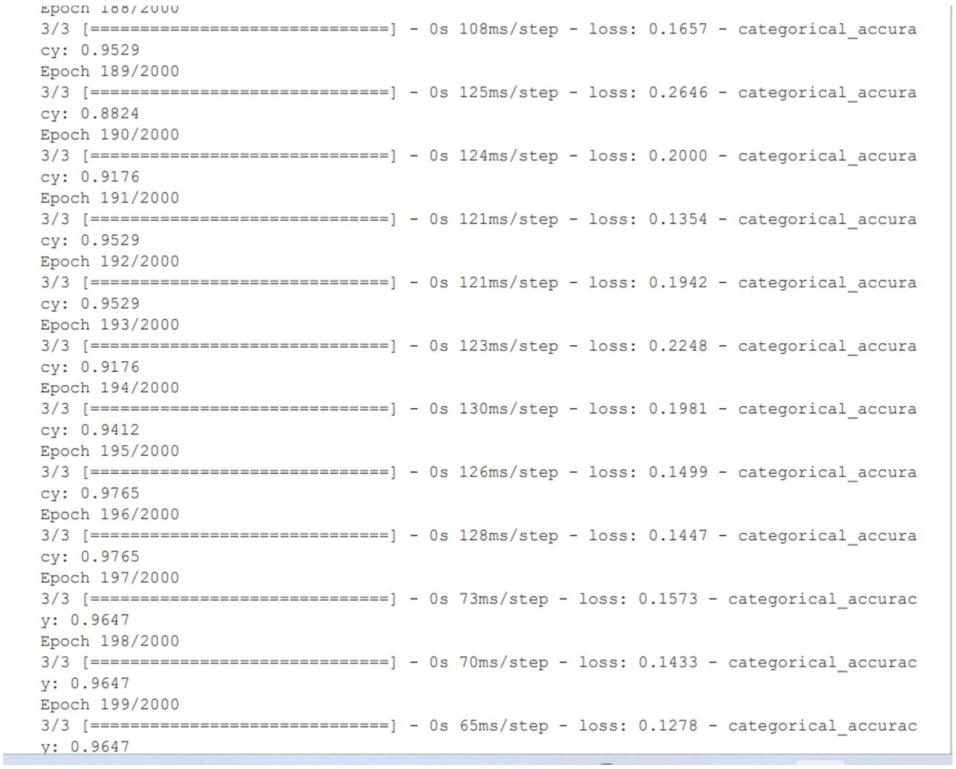
If testing is done carefully, it enhances the overall quality of the software and ensures that the software is error free. In this paper, a sign language recognition system which detects actions given by user through web cam and predict the action as text is developed. Testing is done on three major test cases. Passing all these test cases would ensue the software quality and decrease risk of failure.

A. Test Cases

1) Converting Videos into frames: To check if the captured videos are converted into frames in the form of numpy arrays. If it saves the frames of captured video then the test case is passed.
2) Loading Model: Firstly save the developed model using model.save and try loading the model using keras.models.load model. If it successfully loads without any errors then the test case is passed.
3) Recognize Gesture: Perform the Real Time Sign Language Recognition if the gestures performed by the user are correctly labelled on the output screen, then the test case is passed.

VI. RESULTS AND DISCUSSIONS

The model is trained for 2000 epochs, after 199 epochs, a pretty good accuracy at about 96.47%. which is enough for a good model. Therefore, stop the training, cause training it after sufficient accuracy may lead to overfitting of the model.



```
Epoch 188/2000
3/3 [==============================] - 0s 108ms/step - loss: 0.1657 - categorical_accura
cy: 0.9529
Epoch 189/2000
3/3 [==============================] - 0s 125ms/step - loss: 0.2646 - categorical_accura
cy: 0.8824
Epoch 190/2000
3/3 [==============================] - 0s 124ms/step - loss: 0.2000 - categorical_accura
cy: 0.9176
Epoch 191/2000
3/3 [==============================] - 0s 121ms/step - loss: 0.1354 - categorical_accura
cy: 0.9529
Epoch 192/2000
3/3 [==============================] - 0s 121ms/step - loss: 0.1942 - categorical_accura
cy: 0.9529
Epoch 193/2000
3/3 [==============================] - 0s 123ms/step - loss: 0.2248 - categorical_accura
cy: 0.9176
Epoch 194/2000
3/3 [==============================] - 0s 130ms/step - loss: 0.1981 - categorical_accura
cy: 0.9412
Epoch 195/2000
3/3 [==============================] - 0s 126ms/step - loss: 0.1499 - categorical_accura
cy: 0.9765
Epoch 196/2000
3/3 [==============================] - 0s 128ms/step - loss: 0.1447 - categorical_accura
cy: 0.9765
Epoch 197/2000
3/3 [==============================] - 0s 73ms/step - loss: 0.1573 - categorical_accurac
y: 0.9647
Epoch 198/2000
3/3 [==============================] - 0s 70ms/step - loss: 0.1433 - categorical_accurac
y: 0.9647
Epoch 199/2000
3/3 [==============================] - 0s 65ms/step - loss: 0.1278 - categorical_accurac
y: 0.9647
```

Fig 2. Accuracy

A loss function is a mathematical function that quantifies the difference between predicted and actual values in a machine learning model. The model is trained to minimize the loss using Adam Optimizer. The model epoch loss graph can be seen in the fig 3.

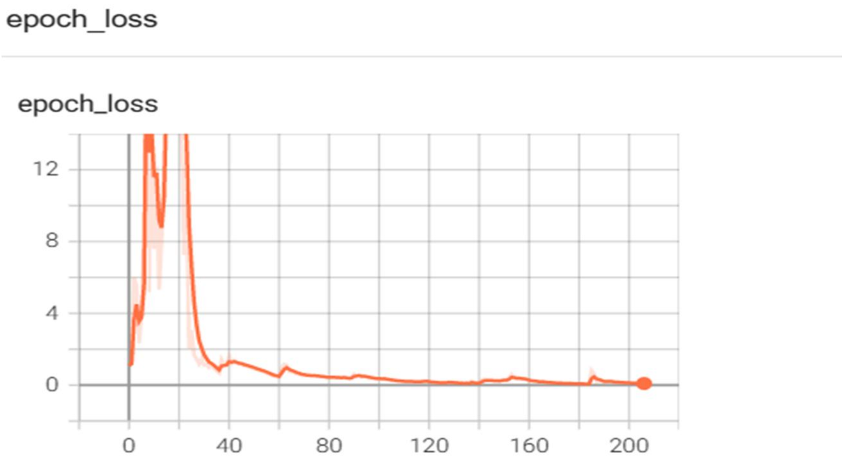


Fig 3. Epoch Loss graphs

The model when tested in real time recognizes the gesture and displays it on the screen. When the user performs thanks action, the developed model correctly predicts and outputs it on the screen as shown in the Fig 4.

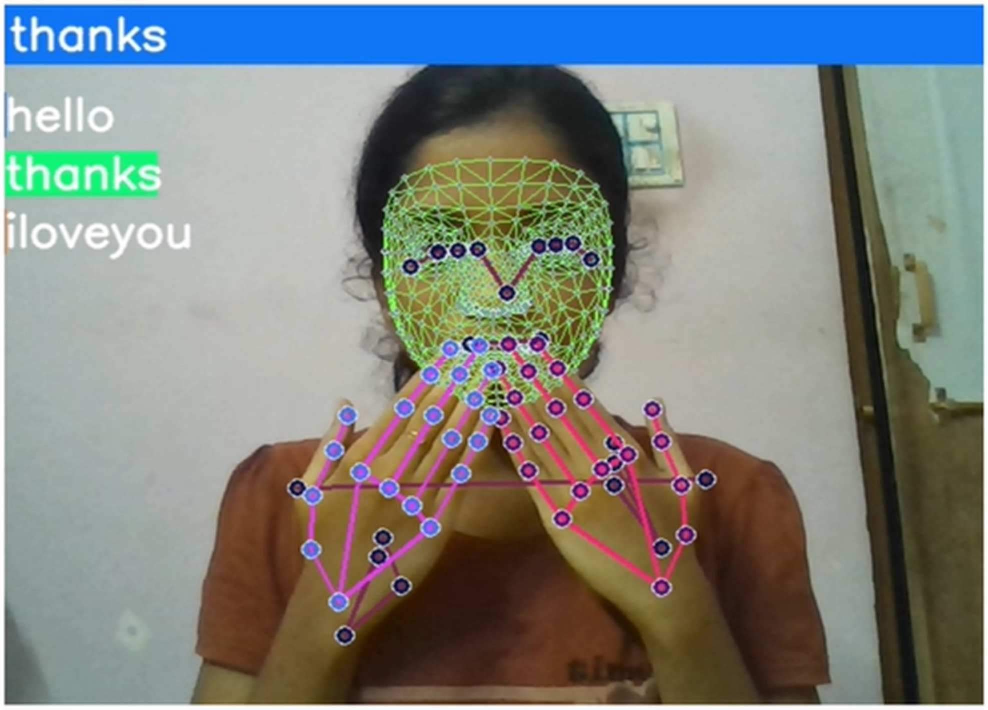
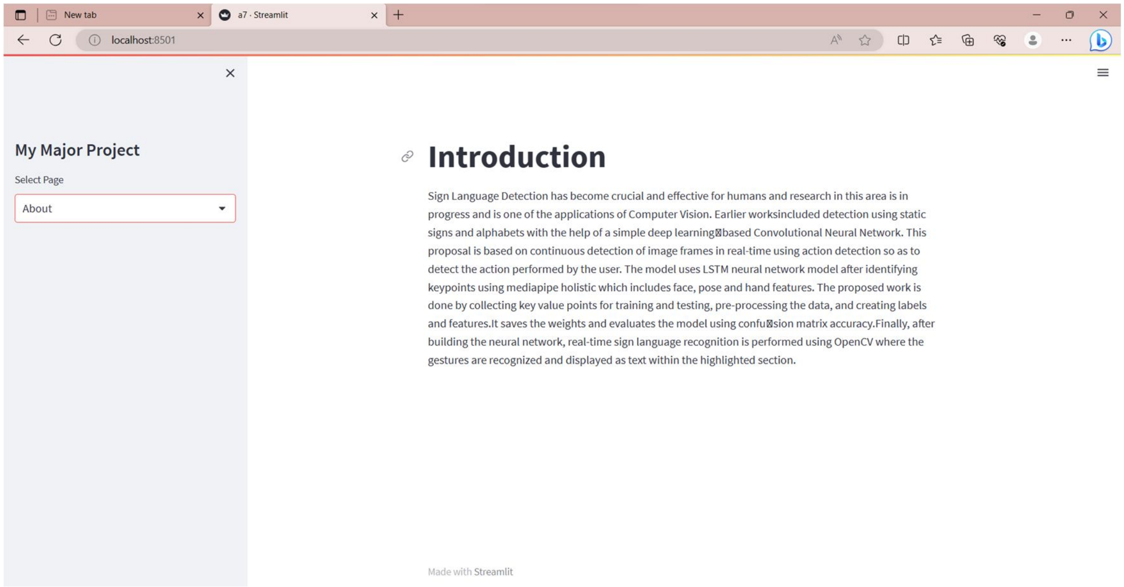


Fig 4.Model recognizes Thanks Action

After Successfully building the model and testing it in real time, GUI is developed using streamlit where a sidebar with two options, Introduction and Demo, where introduction page gives description of the Sign Language Recognition System as shown in Fig 5. The Demo page has a start button to record the video, where the user can pose for three different actions, “hello”, “Thanks”, “I Love you” , and the system will predict the corresponding action on the screen as shown in Fig 6.



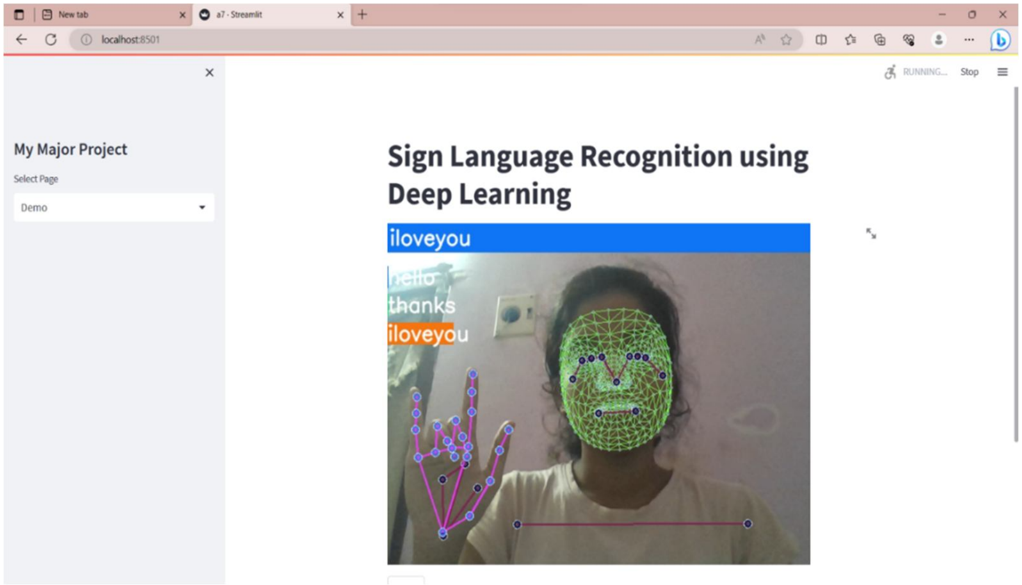
My Major Project

Select Page: About

Introduction

Sign Language Detection has become crucial and effective for humans and research in this area is in progress and is one of the applications of Computer Vision. Earlier worksincluded detection using static signs and alphabets with the help of a simple deep learning⊠based Convolutional Neural Network. This proposal is based on continuous detection of image frames in real-time using action detection so as to detect the action performed by the user. The model uses LSTM neural network model after identifying keypoints using mediapipe holistic which includes face, pose and hand features. The proposed work is done by collecting key value points for training and testing, pre-processing the data, and creating labels and features.It saves the weights and evaluates the model using confu⊠sion matrix accuracy.Finally, after building the neural network, real-time sign language recognition is performed using OpenCV where the gestures are recognized and displayed as text within the highlighted section.

Fig 5. Introduction Page



My Major Project

Select Page: Demo

Sign Language Recognition using Deep Learning

iloveyou

hello
thanks
iloveyou

Fig 6. Demo page for SLR System

VII. CONCLUSION

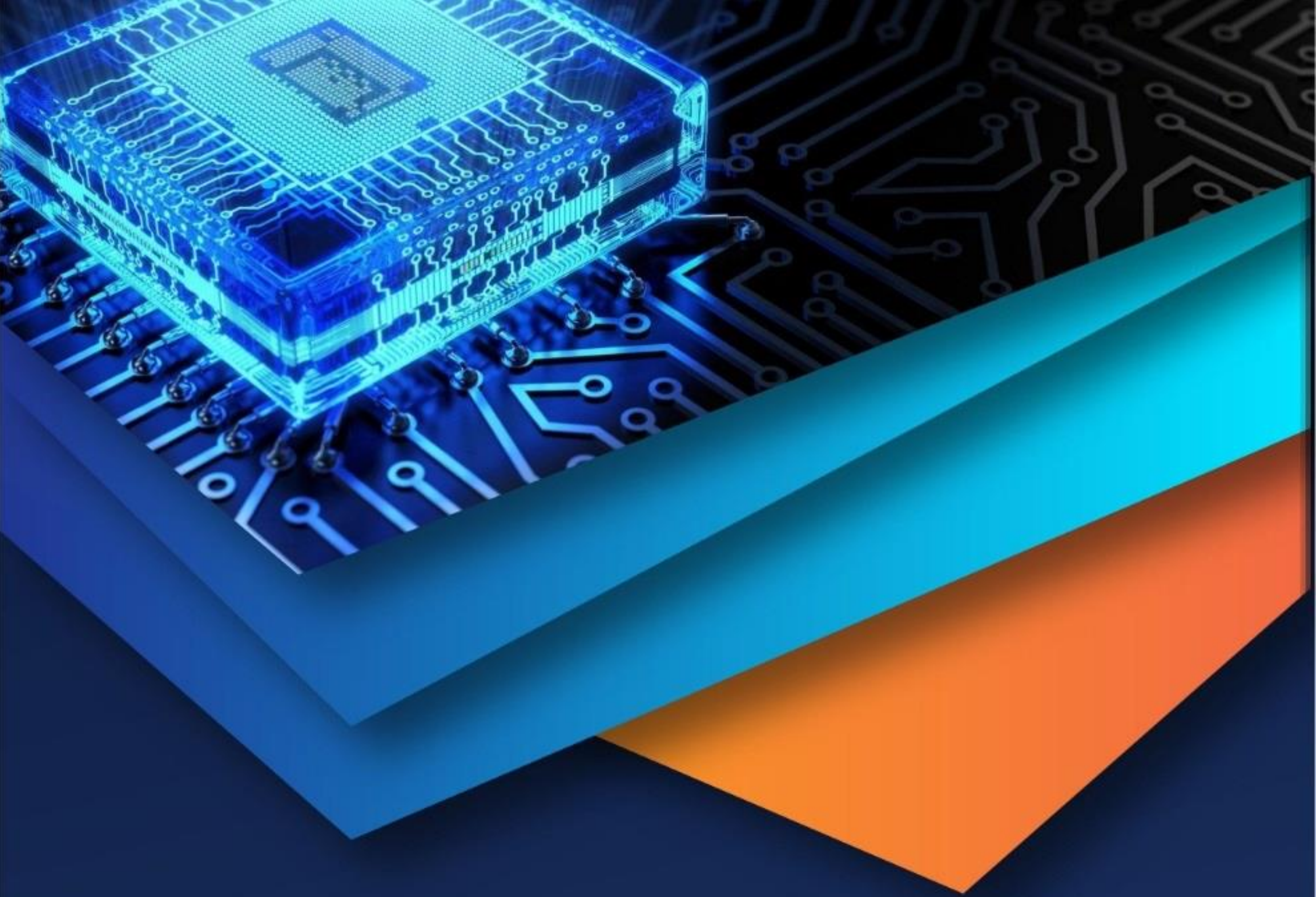
This project aimed to develop a deep learning model for British Sign Language Recognition from gestures to written text. The approach was to build a LSTM model which is trained using dynamic data given by the user, who pose for different actions multiple times. Here the model is trained for 2000 epochs using 30 video sequences with each sequence containing 30 frames for each gesture “hello”, “Thanks”, “I love you” which is collected from the user. After building the model with the collected data, the accuracy on the test set was 96.47%. Long Short-Term Memory Network has given an outstanding performance in the detection of sign language hand gestures through video sequences. One issue that was raised is gestures were a bit hard to categorize in our live demo such as “Hello” vs. “I love you” as they differ by a very small change of palm and fingers. The accuracy may further be improved by increasing the database by increasing the number of video sequences, instead of 30 video sequences, to improve the accuracy furthermore. Therefore, this research concludes deep learning models perform very well recognizing and predicting British sign language in attempting to narrow the gap between speech-impaired people and the general public.

VIII. FUTURE ENHANCEMENTS

The further enhancements that could be made to this existing project include, instead of simply displaying the text, the recognized action can be converted to speech. Furthermore, a greater number of training examples can be added to improve the overall accuracy of the system. In addition to this further training of system to recognize facial expressions, alphabets, symbols can be done. By doing these further enhancements a proficient system that accommodates the entire set of British Sign Language and using it to help speech-impaired people to communicate more easily in public offices or facilities.

REFERENCES

[1] https://www.mygreatlearning.com/blog/opencv-tutorial -in-python/
[2] https://www.researchgate.net/publication/342331104 Sign Language Recognition Using Deep Learning and Computer Vision
[3] https://google.github.io/mediapipe/getting started/python
[4] https://www.geeksforgeeks.org/understanding-of-lstm-networks/
[5] https://ai.googleblog.com/2019/08/on-device-real-time-handtracking-with.html
[6] https://docs.streamlit.io/.
[7] https://developers.google.com/mediapipe
[8] https://www.datacamp.com/tutorial/streamlit



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)