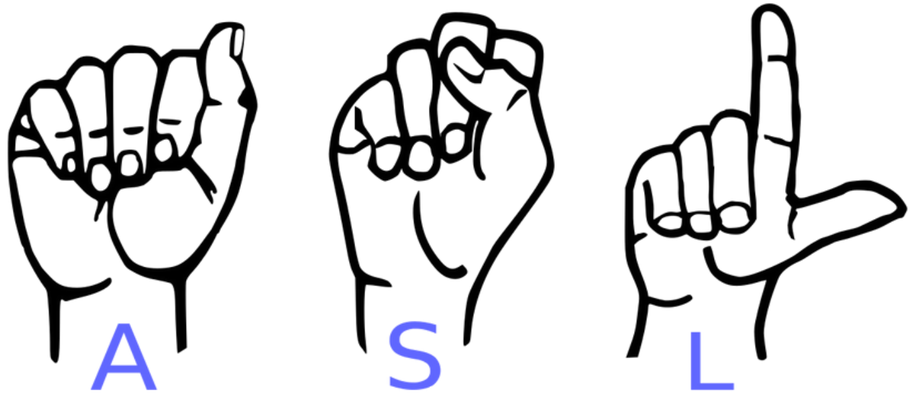
Sign Language Recognition using Convolutional Neural Networks



→

About Our Team

Meet Our Team

Saugat Ghimire
(20087/075)

Sushil Tiwari
(20102/075)

Roshan Basnet
(20075/075)

→

Hello, Welcome!

Introduction

- Sign language is a visual-manual language that is used by many deaf and hard-of-hearing individuals as their primary means of communication.
- It consists of a series of gestures and movements made with the hands, arms, and face, and can include facial expressions and body language.
- The project is to design a system that recognizes a hand gesture representing 26 number of alphabets.

→

Problem Definition

- According to WHO globally there are 1.5 billion deaf people and this number could increase to 2.5 billion by the end of 2050.
- The communication between two peers for exchanging information can be done via linguistic feature or with the help of gestures.
- However, where linguistic feature cannot be used, the realization of gesture into meaningful information leads to development of procedure for communication between two peers.
- Real-time detection with top-level classification and accurate accuracy of model remains challenging.
- This project builds a system capable of recognizing Hand Gestures (ASL) representing alphabets, using CNN.

→

Problem Definition

Here are some of the key problems that deaf people may face in terms of communication:

1. Limited access to sign language interpreters, which can make it difficult to communicate effectively in many settings, including at work, school, and healthcare appointments.
2. Barriers to accessing written or spoken information, as many deaf people rely on sign language as their primary mode of communication and may not be able to fully understand written or spoken language.
3. Difficulty communicating with people who do not know sign language, which can lead to social isolation and frustration.
4. Limited access to technology and communication devices designed for the deaf community, such as video phones and text messaging services.
5. Misunderstandings and misinterpretations due to differences in sign language dialects, regional variations, or limited sign language proficiency among some people.

→

Solution Strategy

The design of a real-time Sign Language Recognition system basically involves:

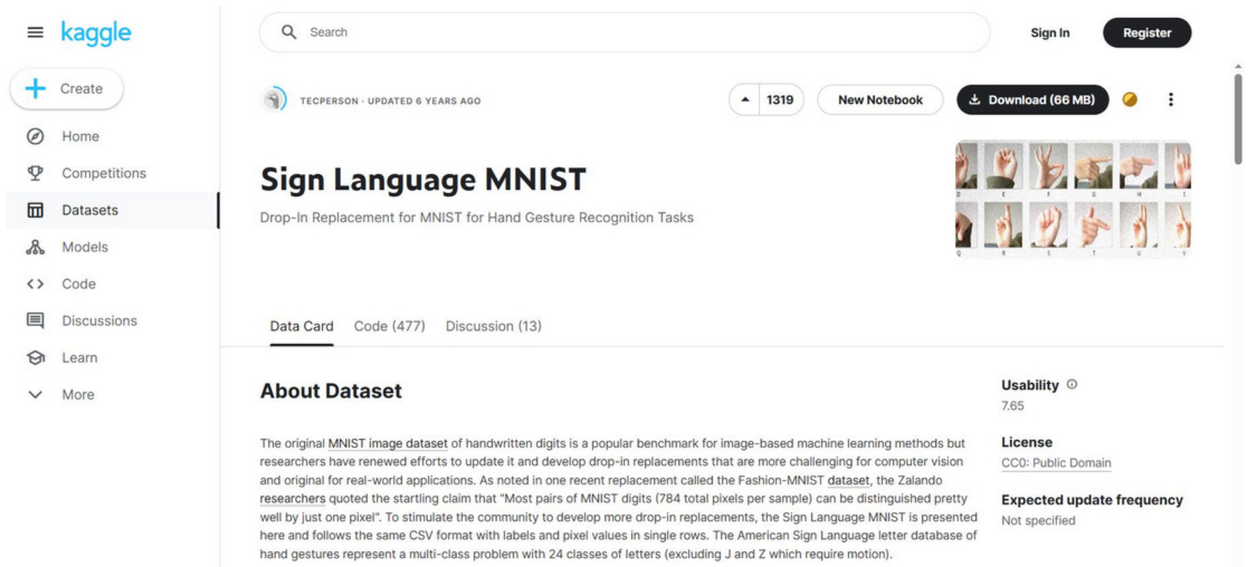
- Collect dataset of every alphabet in varying lighting conditions.
- Build and train the model using CNN architecture considering ASL Dataset.
- Testing involves:
  - Acquiring user’s hand gesture at real-time
  - Classifiying hand gestures using trained CNN Model.
- Text to Speech Conversion
- Word suggestions

→

Workflow

Data Collection and Preprocessing

- Initially, we utilized a dataset sourced from Kaggle and preprocessed it with the VGG16 preprocessing function.
- However, the results obtained under varying lighting conditions were not as anticipated, so we captured our own hand gesture images.



→

Dataset Overview

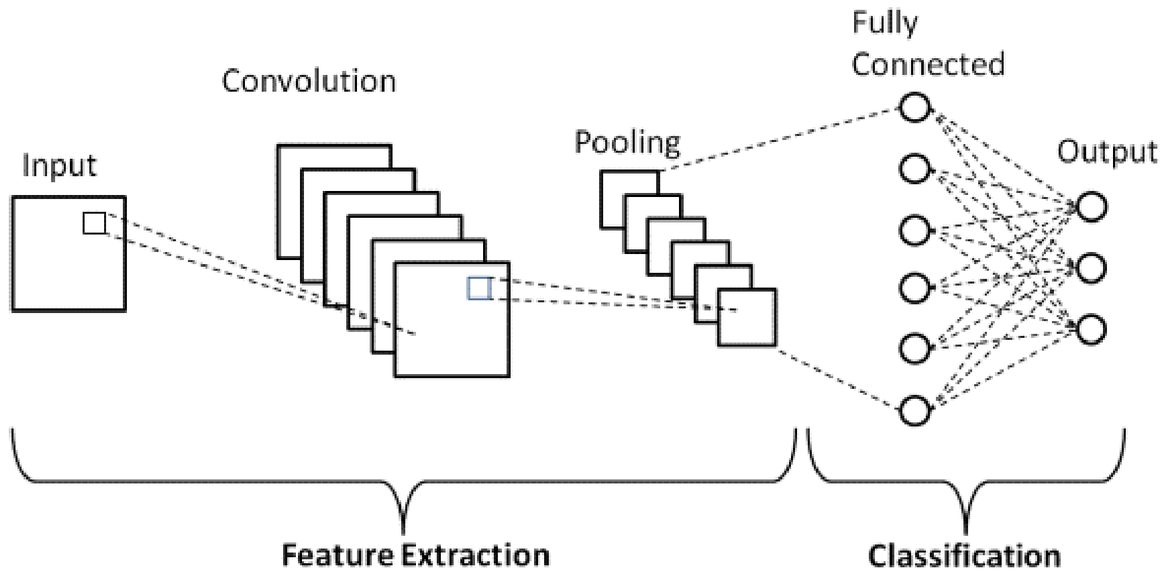
- The training data set contained ~10,000 images.
- Size of each image: 224x224 pixels.
- There are 29 classes (26 are for letters A-Z and 3 classes for space, delete and next)

→

CNN Architecture Selection

We implemented the following layers in our model:

- Conv2D - Convolutional Layer, used for feature extraction
- MaxPooling2D - Calculates largest value in every patch
- Flatten - 2D array to linear vector
- Dense - Deeply connected layer

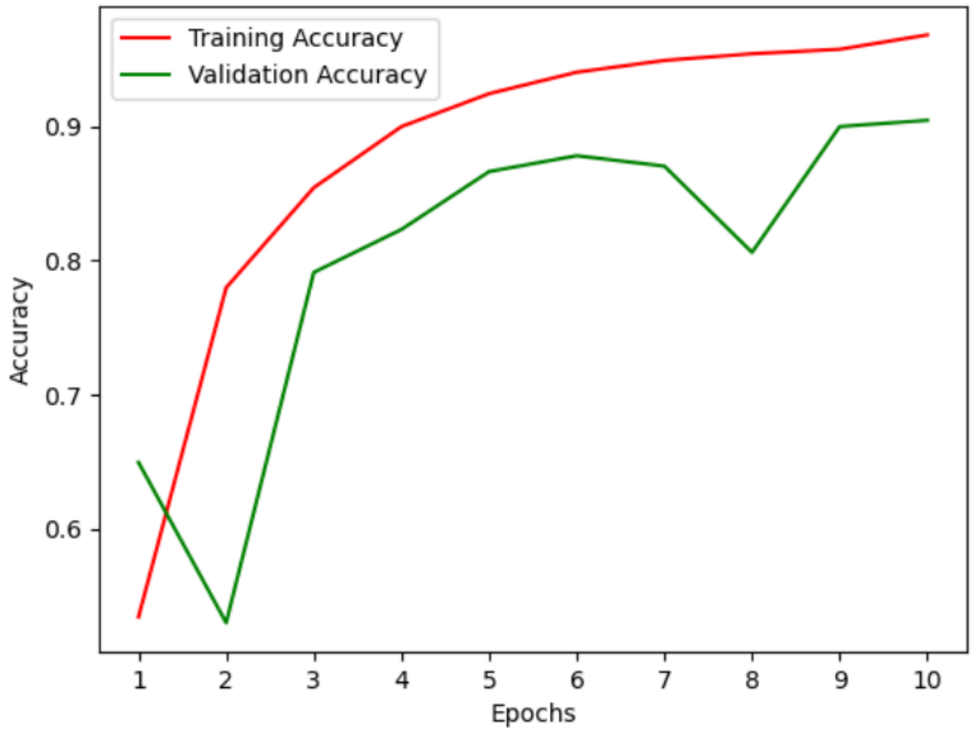


Model Creation and Training

Training of the model is executed fitting the splitted data into training and test set.

```
In [32]: model.summary()

Model: "sequential_1"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 conv2d_2 (Conv2D)           (None, 224, 224, 32)      896

 max_pooling2d_2 (MaxPooling  (None, 112, 112, 32)     0
 2D)

 conv2d_3 (Conv2D)           (None, 112, 112, 64)      18496

 max_pooling2d_3 (MaxPooling  (None, 56, 56, 64)       0
 2D)

 flatten_1 (Flatten)         (None, 200704)            0

 dense_1 (Dense)             (None, 26)                5218330

=================================================================
Total params: 5,237,722
Trainable params: 5,237,722
Non-trainable params: 0
_________________________________________________________________
```

Model Performance and Output



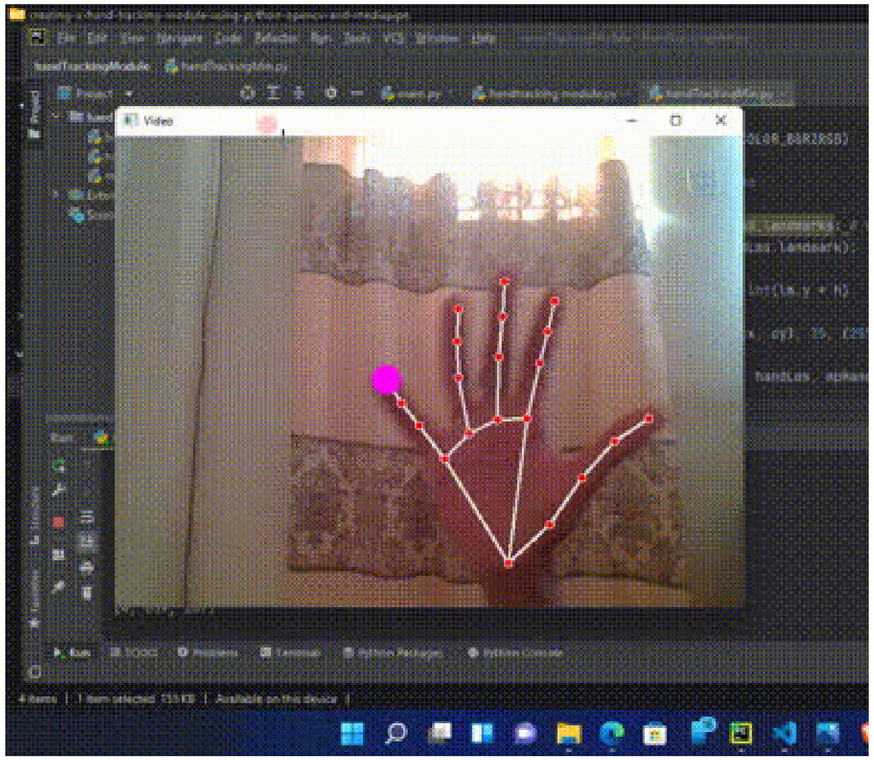
Accuracy Graph

→

How did we overcome this problem?

We used a different approach by using Google's Mediapipe library as a handtracking module to track our hands which we then mapped the skeleton of hands to a white background.

Google's MediaPipe Library

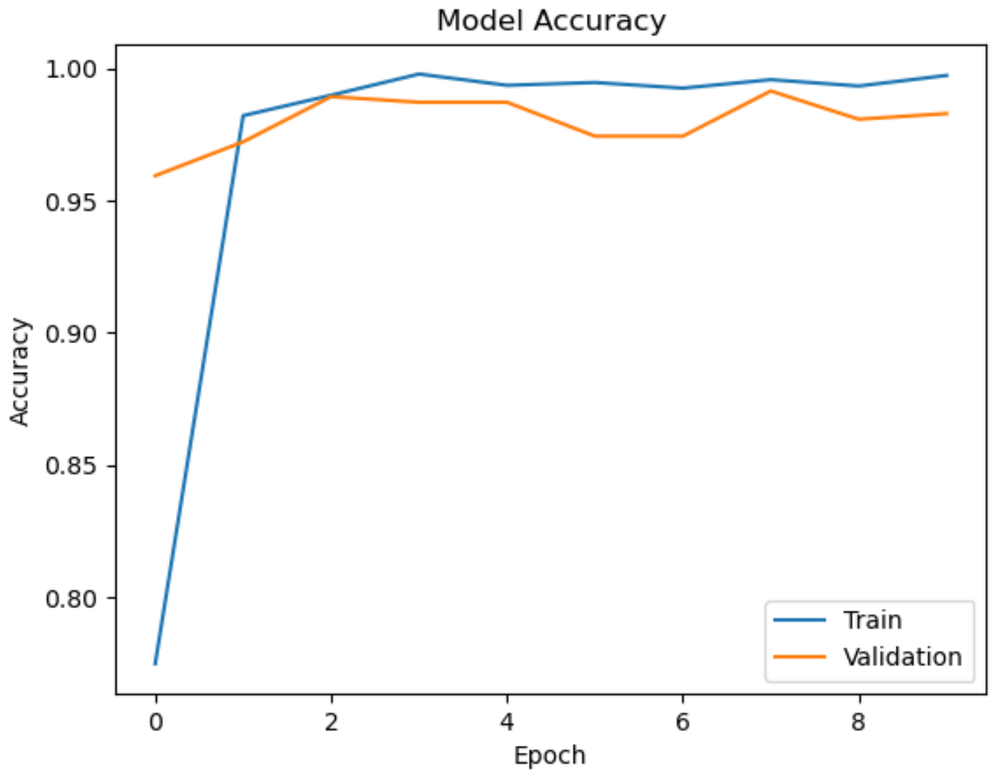


Mediapipe Library

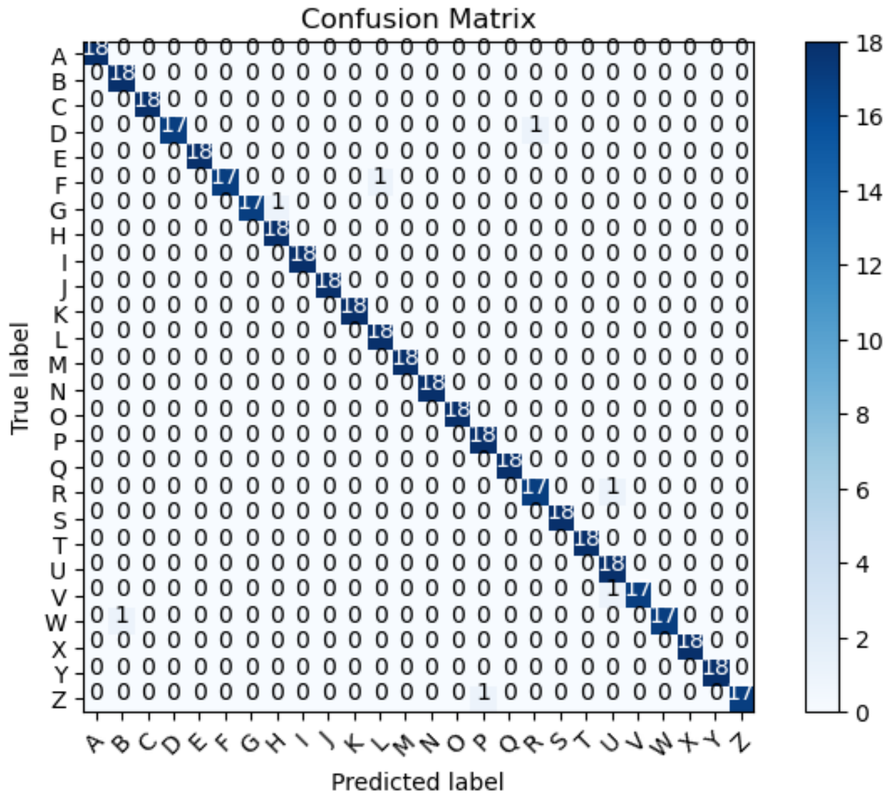
Hand skeleton mapped
to white image

→

Model Performance and Output



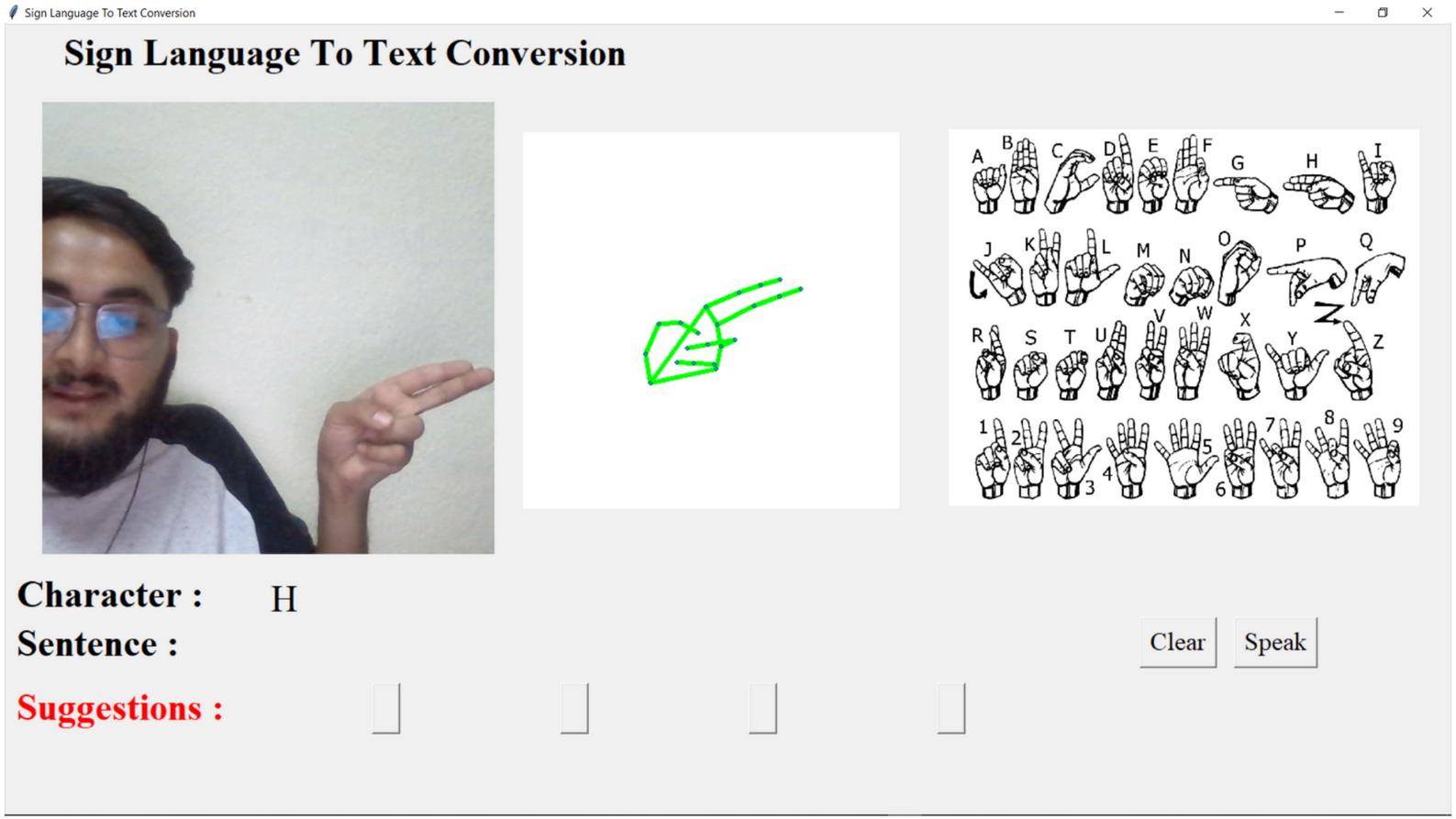
Accuracy Graph



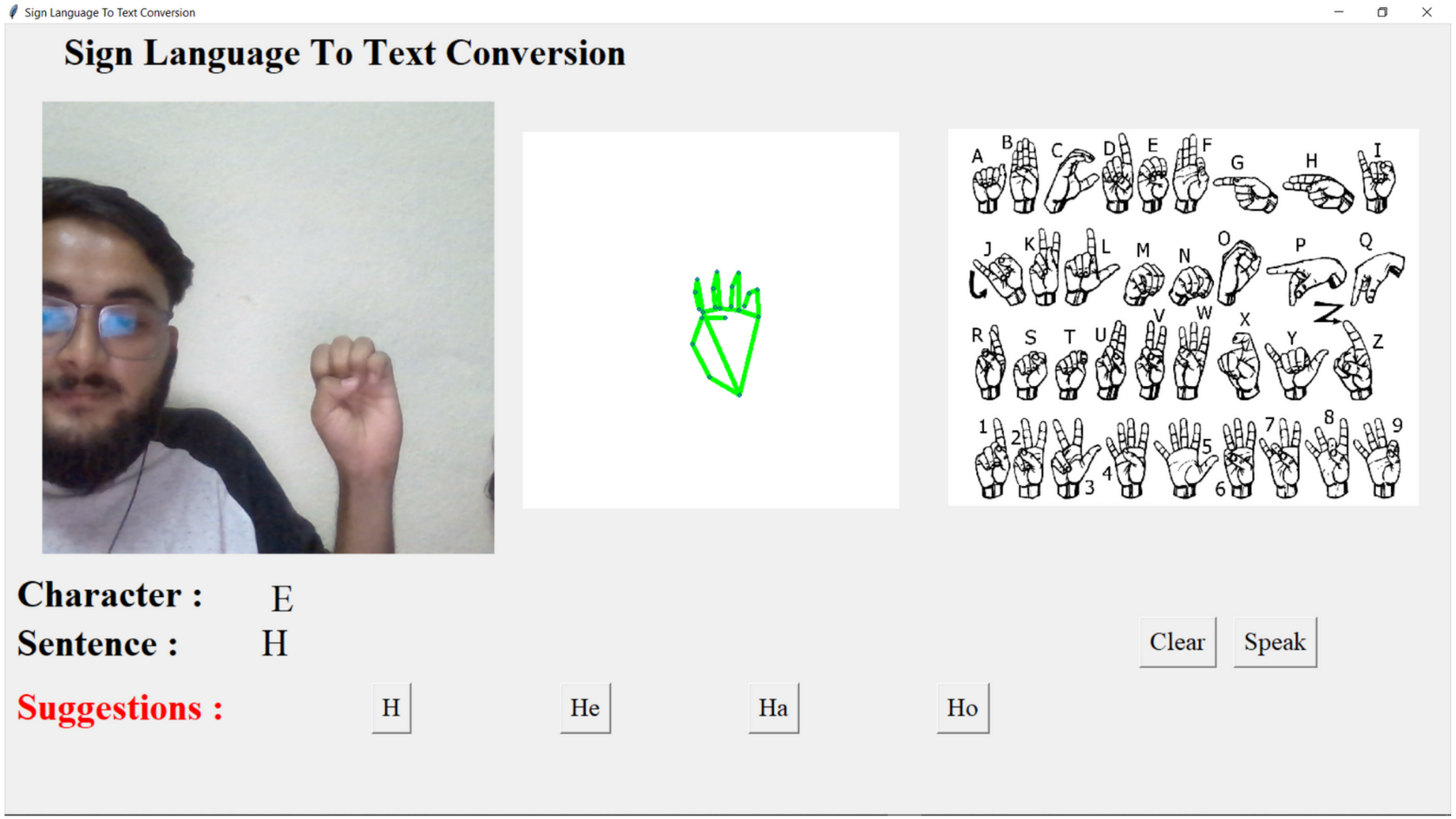
Confusion Matrix

→

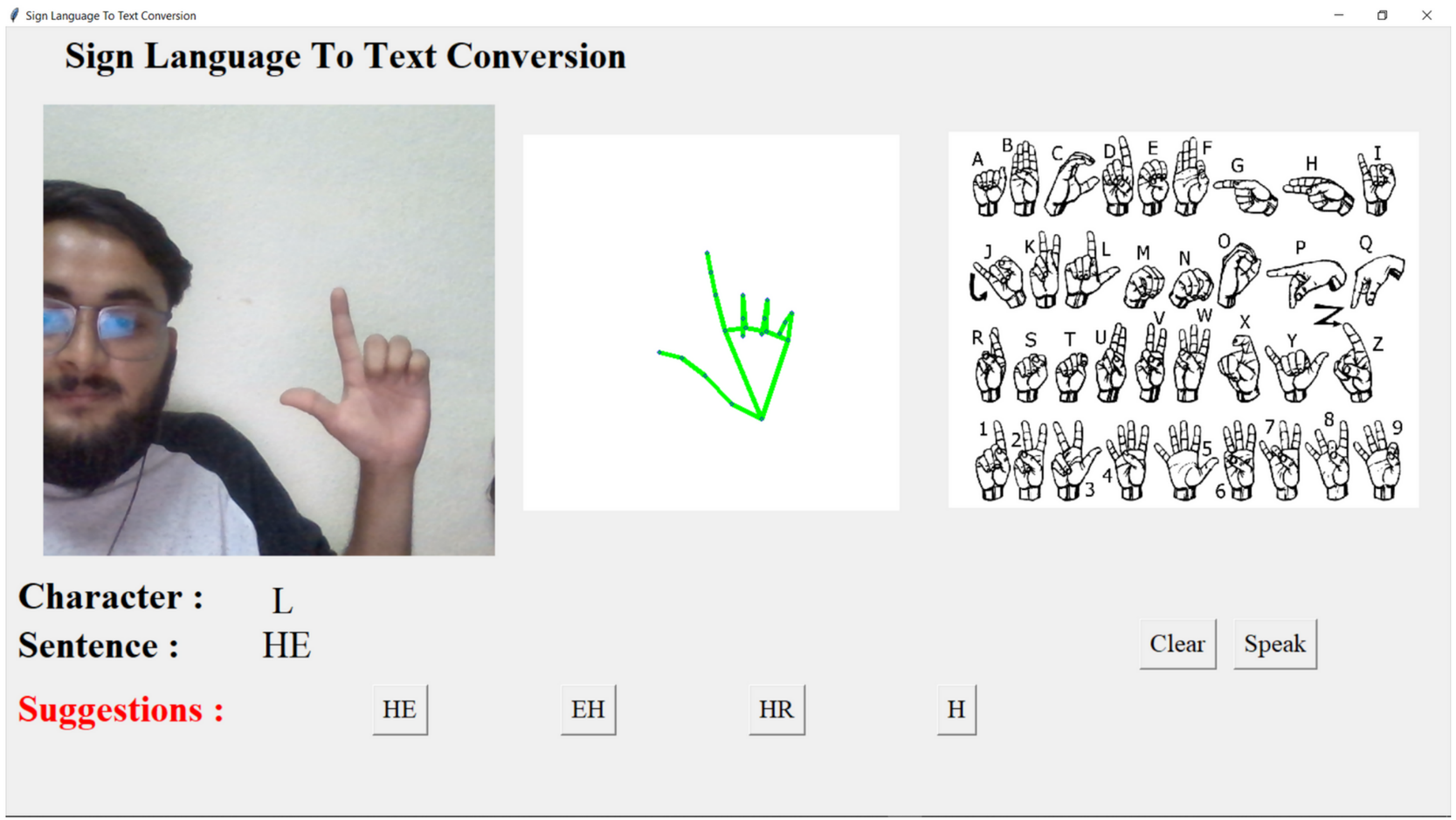
Results


Sign Language To Text Conversion
Sign Language To Text Conversion
Character : H
Sentence :
Clear
Speak
Suggestions :

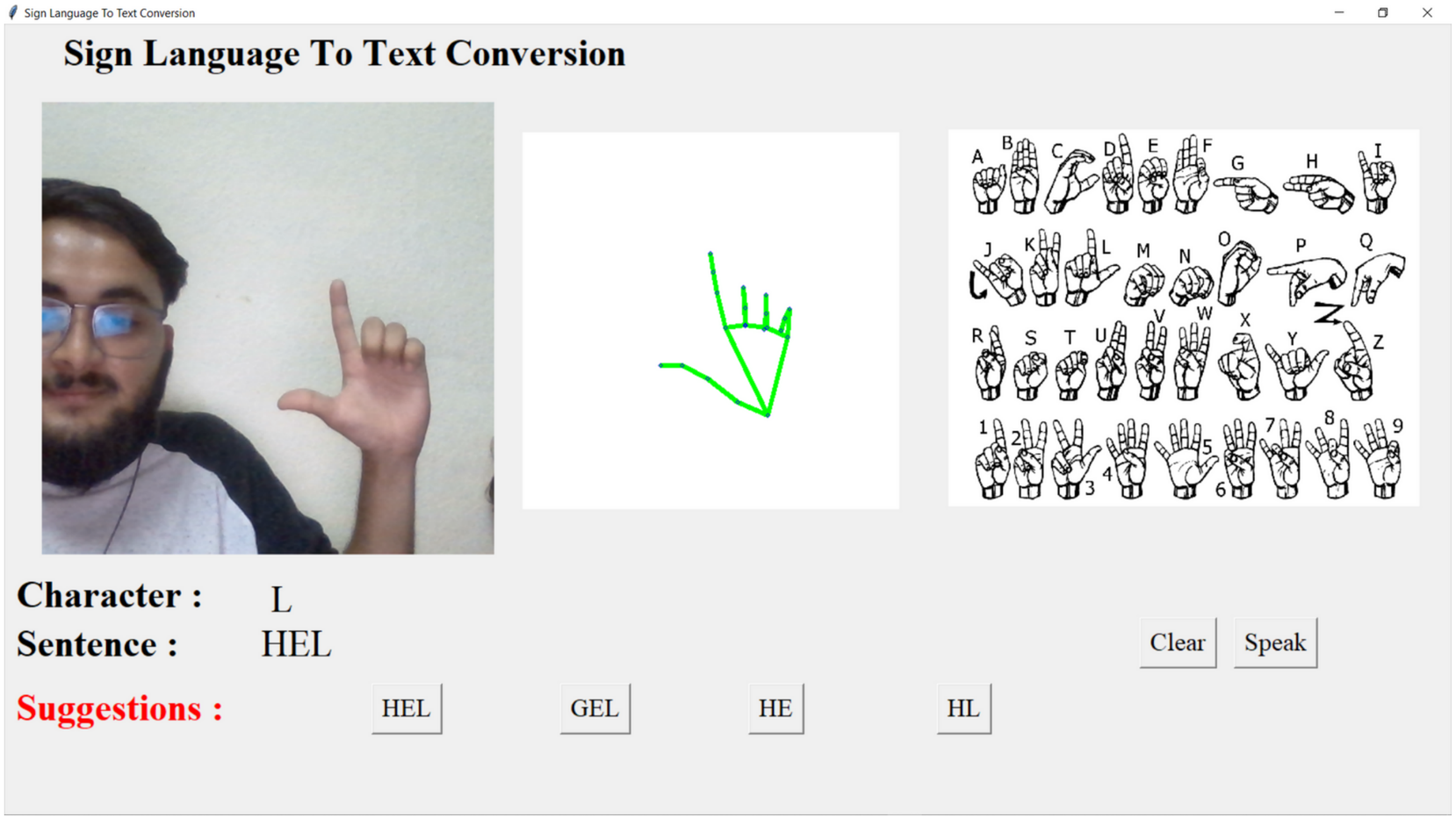
Results


Sign Language To Text Conversion
Sign Language To Text Conversion
Character : E
Sentence : H
Clear
Speak
Suggestions :
H
He
Ha
Ho

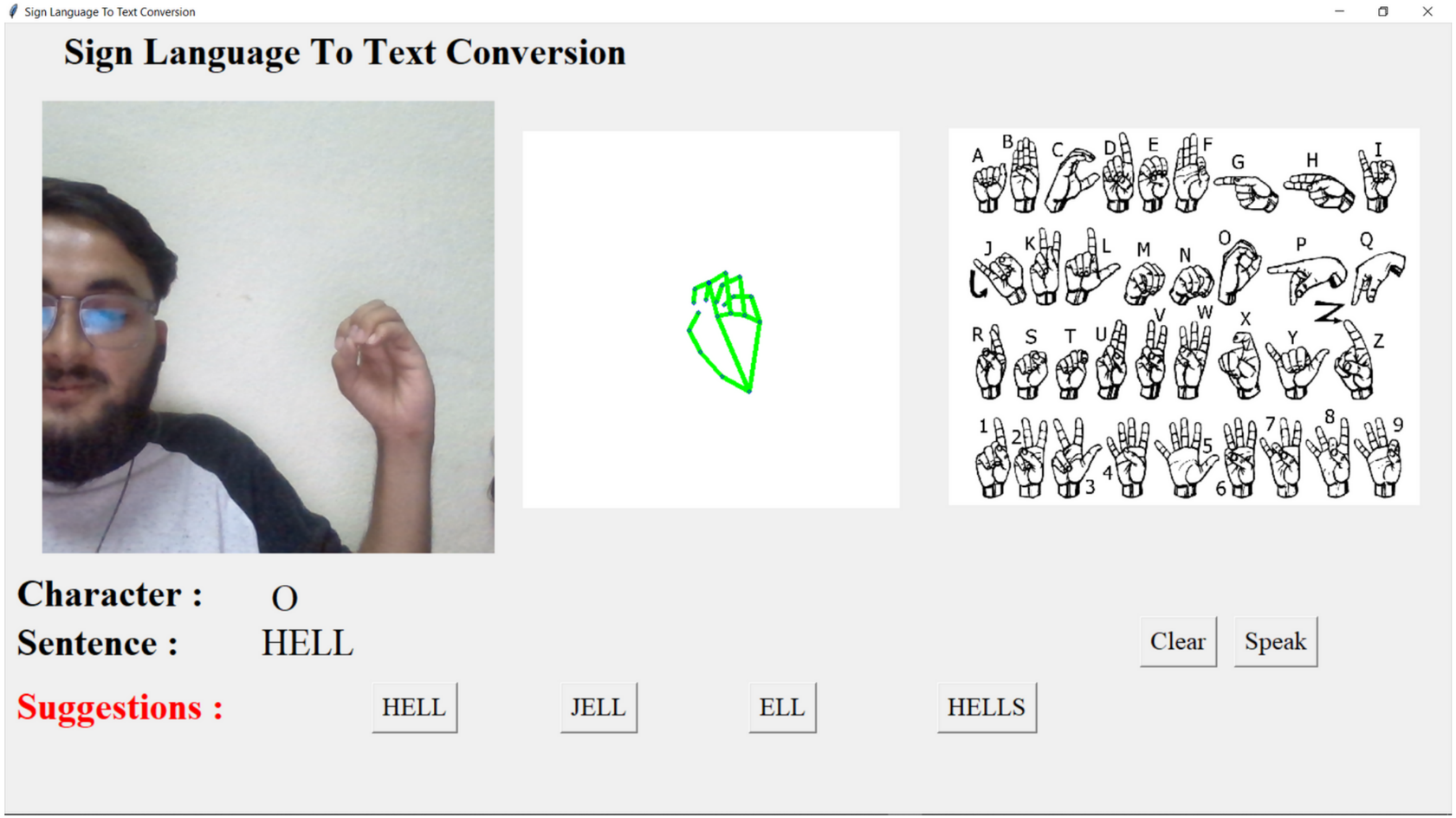
Results


Sign Language To Text Conversion
Sign Language To Text Conversion
Character : L
Sentence : HE
Clear
Speak
Suggestions :
HE
EH
HR
H

Results


Sign Language To Text Conversion
Sign Language To Text Conversion
Character : L
Sentence : HEL
Clear
Speak
Suggestions :
HEL
GEL
HE
HL

Results


Sign Language To Text Conversion
Sign Language To Text Conversion
Character : O
Sentence : HELL
Clear
Speak
Suggestions :
HELL
JELL
ELL
HELLS

CONCLUSION

- The Model performed sign classification with accuracy of ~95% on the observed data.

- A weighted pre-trained model has been obtained after 8 epochs of training as “weights” in this process . Hence , classification of real-time data is possible without need of pre training sessions.

→

LIMITATION & FUTURE SCOPE

- Implementation of translation of letters which requires movement (For ex.Letter Z).
- Increasing accuracy of our model.
- The project can be extended for other type of information as well like numbers, words, etc.

→

Thank You For Listening

THANK YOU