BASAVARAJESWARI GROUP OF INSTITUTIONS

BALLARI INSTITUTE OF TECHNOLOGY & MANAGEMENT

Autonomous Institute under VTU, Belagavi | Approved by AICTE,New Delhi Recognized by Govt. of Karnataka

NACC Accredited Institution*

(Recognized by Govt. of Karnataka, approved by AICTE, New Delhi & Affiliated to Visvesvaraya Technological University, Belgavi)

"JnanaGangotri" Campus, No.873/2, Ballari-Hospet Road, Allipur, Ballar1-583 104 (Karnataka) (India)

Ph: 08392 – 237100 / 237190, Fax: 08392 – 237197

DEPARTMENT OF ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING

A Mini Project Report On

"Sign-Speak"

Project Associates:

Bhavana Joshi	3BR22AI025
Bhavya Jain	3BR22AI027
G Akash Kulkarni	3BR22AI042
G Sahethi	3BR22AI045

Under the Guidance of

P SAHANA PRASAD
Dept of AIML,
BITM, Ballari.

Visvesvaraya Technological University

Belagavi, Karnataka

2024-2025

BASAVARAJESWARI GROUP OF INSTITUTIONS

BALLARI INSTITUTE OF TECHNOLOGY & MANAGEMENT

Autonomous Institute under VTU, Belagavi | Approved by AICTE,New Delhi Recognized by Govt. of Karnataka

NACC Accredited Institution*

(Recognized by Govt. of Karnataka, approved by AICTE, New Delhi & Affiliated to Visvesvaraya Technological University, Belgavi)

"JnanaGangotri" Campus, No.873/2, Ballari-Hospet Road, Allipur, Ballar1-583 104 (Karnataka) (India)

Ph: 08392 – 237100 / 237190, Fax: 08392 – 237197

DEPARTMENT OF ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING

CERTIFICATE

This is to certify that the project work entitled **"Sign-Speak"** is a bonafide work carried out by **Bhavana Joshi(3BR22AI025), Bhavya Jain(3BR22AI027), G Akash Kulkarni(3BR22AI042), G Sahethi(3BR22AI045)}** in partial fulfillment for the award of degree of **Bachelor Degree in AIML** in the VISVESVARAYA TECHNOLOGICAL UNIVERSITY, Belagavi during the academic year 2024-2025. It is certified that all corrections and suggestions indicated for internal assessment have been incorporated in the report deposited in the library. The project has been approved as it satisfies the academic requirements in respect of mini project work prescribed for a Bachelor of Engineering Degree.

Signature of project guide	Signature of Coordinator	Signature of HoD
P SAHANA PRASAD	**P. ASHA JYOTHI**	**Dr. B M Vidyavathi**
ASSISTANT PROFESSOR	**ASSISTANT PROFESSOR**	

Abstract

Sign language serves as a vital means of communication for individuals with hearing or speech impairments, bridging the gap in interaction with the hearing population. This project focuses on the development of a robust and efficient sign language detection system that translates hand gestures into textual and audio outputs. Using a Convolutional Neural Network (CNN) model trained on image datasets, the system is designed to recognize individual letters, construct words, and form coherent sentences. Key elements of the project include the use of OpenCV for image preprocessing and TensorFlow with Keras for implementing and training the CNN model. The training dataset comprises labeled images of hand gestures, ensuring high accuracy in detection. The system integrates a text-to-speech (TTS) component implemented in JavaScript, enabling real-time voice output for detected sentences, further enhancing accessibility. The proposed model aims to bridge the communication gap, providing a seamless, user-friendly interface for individuals reliant on sign language. By leveraging advanced image processing and machine learning techniques, this project contributes to the domain of assistive technologies, facilitating effective communication and fostering inclusivity. The system's capability to decode gestures into meaningful sentences with audio output highlights its potential for real-world applications.

Acknowledgement

I would like to express my heartfelt gratitude to all those who supported and guided me throughout the development of this project on sign language detection.

First and foremost, I would like to thank my institution, BITM College, for providing the resources and environment necessary to carry out this project. I am particularly grateful to the faculty of the AIML branch for their encouragement and technical guidance, which proved invaluable during the various stages of development.

A special thanks to my project guide “P Sahana Prasad mam”, whose expertise and feedback have been instrumental in refining this work. Their consistent support and constructive criticism helped me overcome challenges and achieve the desired results.

I am also deeply thankful to my project coordinator “P Asha Jyothi mam” for their dedication and assistance in managing the project's milestones and for ensuring that the project remained on track. Their guidance and support were vital in achieving the project’s objectives.

I also extend my gratitude to my teammates, who provided valuable suggestions and moral support during this journey.

This project is dedicated to improving inclusivity and accessibility, and it would not have been possible without the combined efforts and encouragement of everyone involved. Thank you for your unwavering support!

Thank you all!

Name	USN
Bhavana Joshi	3BR22AI025
Bhavya Jain	3BR22AI027
G Akash Kulkarni	3BR22AI042
G Sahethi	3BR22AI045

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction

American Sign Language is a predominant sign language Since the only disability D&M people have been communication related and they cannot use spoken languages hence the only way for them to communicate is through sign language. Communication is the process of exchange of thoughts and messages in various ways such as speech, signals, behavior and visuals. Deaf and Dumb (D&M) people make use of their hands to express different gestures to express their ideas with other people. Gestures are the nonverbally exchanged messages and these gestures are understood with vision. This nonverbal communication of deaf and dumb people is called sign language. In our project we basically focus on producing a model which can recognise Fingerspelling based hand gestures in order to form a complete word by combining each gesture. The gestures we aim to train are as given in the image below.

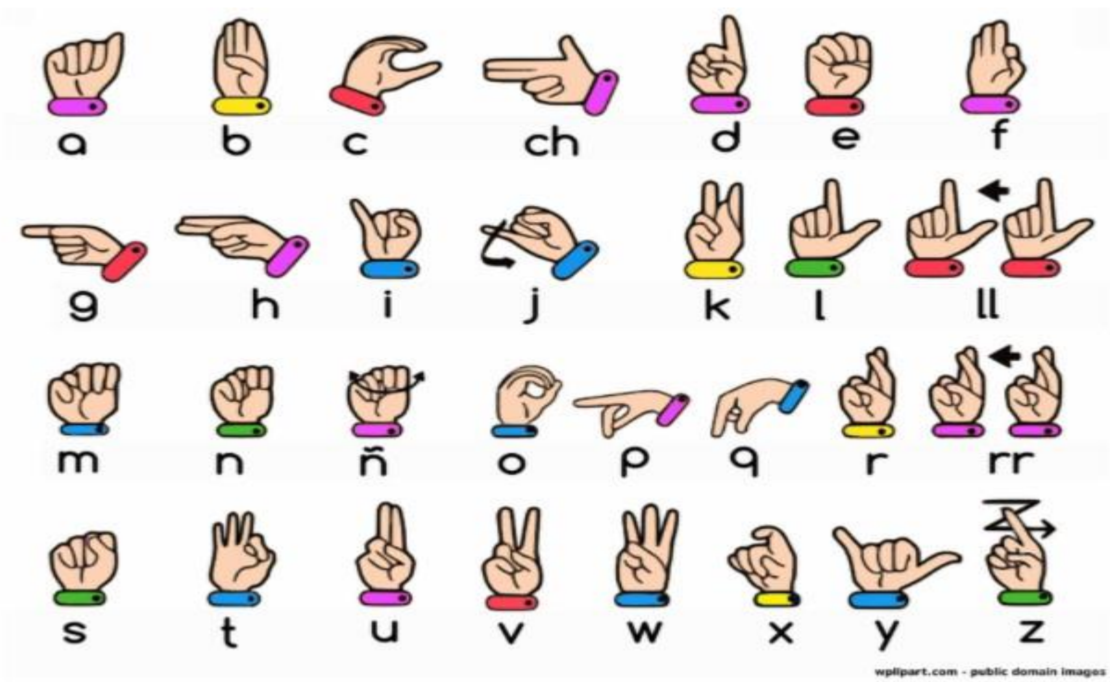

Figure 1 ASL gestures.

1.1 Objectives

More than 70 million deaf people around the world use sign languages to communicate. Sign language allows them to learn, work, access services, and be included in the communities. It is hard to make everybody learn the use of sign language with the goal of ensuring that people with disabilities can enjoy their rights on an equal basis with others. So, the aim is to develop a user-friendly human computer interface (HCI) where the computer understands the American sign language This Project will help the dumb and deaf people by making their life easy. To create a computer software and train a model using CNN which takes an image of hand gesture of American Sign Language and shows the output of the particular sign language in text format converts it into audio format.

1.2 Scope of the project

This System will be Beneficial for Both Dumb/Deaf People and the People Who do not understands the Sign Language. They just need to do that with sign Language gestures and this system will identify what he/she is trying to say after identification it gives the output in the form of Text as well as Speech format.

Chapter 2: System Analysis

2.1 Comparison Table

Table 1 Comparision Table

Author name	Mahesh Kumar	Krishna Modi	Bikash K. Yadav	Ayush Pandey	Victorial Adebimpe Akano	Rakesh Kumar
Algorithm	LDA	Blob Analysis	CNN	CNN	KNN	contour measure ment
Accuracy	80%	93%	95.8%	95%	92%	86%
Year	2018	2013	2020	2020	2018	2021

2.2 Project Requirements

2.2.1 Hardware Requirement

- Webcam
- GPU: NVIDIA GPU with CUDA support (GTX 1660, RTX 2060)
- RAM: Minimum 8GB (16GB+ recommended)
- Storage: SSD with 256GB+
- Monitor: Full HD display
- Network: Stable internet connection

2.2.2 Software Requirement

- Operating System: Windows 10 and Above
- IDE: Tkinter
- Programming Language: Python 3.10.0
- Python libraries: OpenCV, NumPy, Keras, mediapipe, Tensorflow, pyttsx3, pandas

2.3 Project Feasibility Study

2.3.1 Operational feasibility

The whole purpose of this system is to handle the work much more accurately and efficiently with less time consumption. This app is very user-friendly to use. They only require knowledge about American Sign Language. The system is operationally feasible as it is very easy for the End users to operate it. It only needs basic information about windows application.

2.3.2 Technical feasibility

The technical needs of the system may include: Front-end and back-end selection an important issue for the development of a project is the selection of suitable front-end and back-end. When we decided to develop the project, we went through an extensive study to determine the most suitable platform that suits the needs of the organization as well as helps in development of the project. The aspects of our study included the following factors.

- Front-end selection:
    - It must have a graphical user interface that assists users that are not from IT background. So we have made front-end using Python Tkinter Gui.
    - Features:
        1. Scalability and extensibility.
        2. Flexibility.
        3. Easy to debug and maintain.
- Back-end Selection:
    - We have used Python as our Back-end Language which has the most widest library collections. The technical feasibility is frequently the most difficult area encountered at this stage. Our app will fit perfectly for technical feasibility

2.4 Literature Survey

2.4.1 Mahesh Kumar N B1 Assistant Professor (Senior Grade),

Bannari Amman Institute of Technology, Sathyamangalam, Erode, India (2018):

This paper shows the sign language recognizing of 26 hand gestures in Indian sign language using MATLAB. The proposed system contains four modules such as: pre-processing and hand segmentation, feature extraction, sign recognition and sign to text. The Linear Discriminant Analysis (LDA) algorithm was used for gesture recognition and recognized gestures are converted into text and voice format. The proposed system helps to dimensionality.

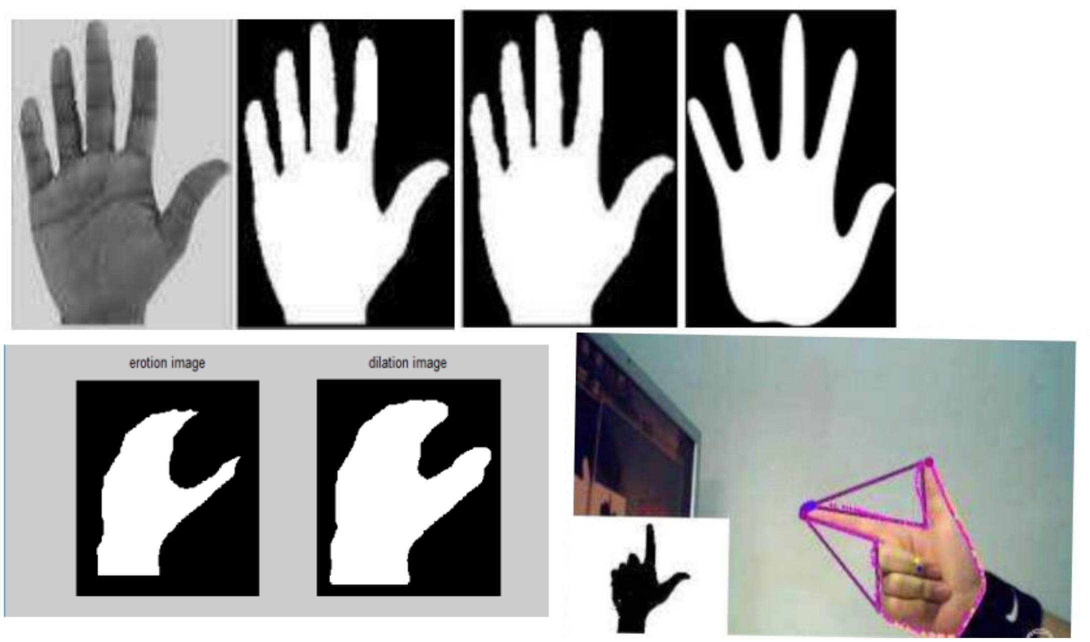

Figure 2 Literature survey pic-1

2.4.2 Sign Language to Text and Speech Conversion by Bikash K. Yadav Sinhgad College of Engineering, Pune, Maharashtra (2020)

Sign language is one of the oldest and most natural forms of language for communication. Since most people do not know sign language and interpreters are very difficult to come by, They have come up with a real-time method using Convolution Neural Network (CNN) for fingerspelling based American Sign Language (ASL). In Their method, the hand is first passed through a filter and after the filter has applied the hand is passed through a classifier that predicts the class of the hand gestures. Using Their approach, They are able to reach a model accuracy of 95.8%.

Figure 3Literature Survey pic-2

2.4.3 Sign Language to Text and Speech Translation in Real Time Using Convolutional Neural Network-by Ankit Ojha Dept. of ISE JSSATE Bangalore, India.

Creating a desktop application that uses a computer's webcam to capture a person signing gestures for American sign language (ASL), and translate it into corresponding text and speech in real time. The translated sign language gesture will be acquired in text which is farther converted into audio. In this manner they are implementing a finger spelling sign language translator. To enable the detection of gestures, they are making use of a Convolutional neural network (CNN). A CNN is highly efficient in tackling computer vision problems and is capable of detecting the desired features with a high degree of accuracy upon sufficient training. The modules are image acquisition, hand region segmentation and hand detection and tracking hand posture recognition and display as text/speech. A finger spelling sign language translator is obtained which has an accuracy of 95%.

2.4.4 CONVERSION OF SIGN LANGUAGE TO TEXT AND SPEECH USING MACHINE LEARNING TECHNIQUES, Author: Victorial Adebimpe Akano (2018)

Communication with the hearing impaired (deaf/mute) people is a great challenge in our society today; To convert ASL signed hand gestures into text as well as speech using unsupervised feature learning to eliminate communication barrier with the hearing impaired and as well provide teaching aid for sign language.

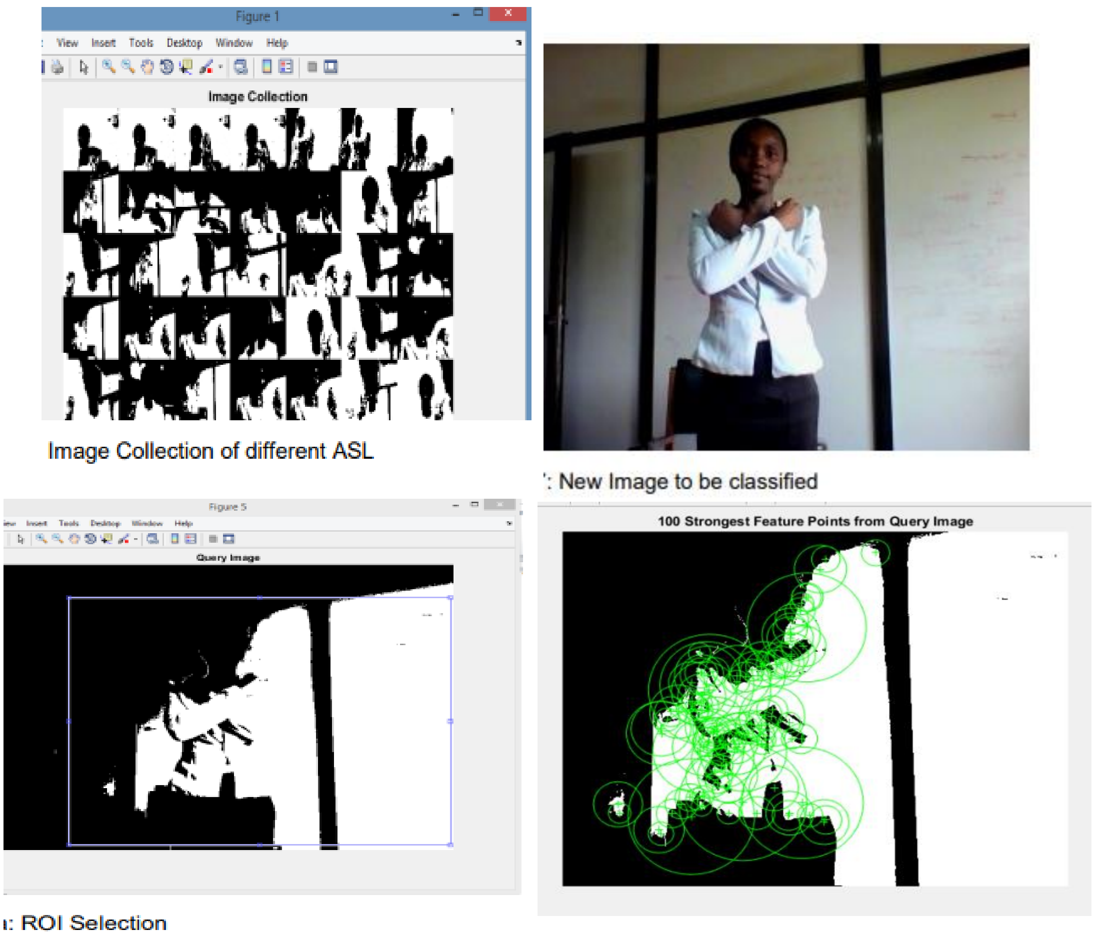

Figure 4 Literature Survey pic-3

Chapter3 System Design

3.1 Architectural Design

3.1.1 Block Diagram

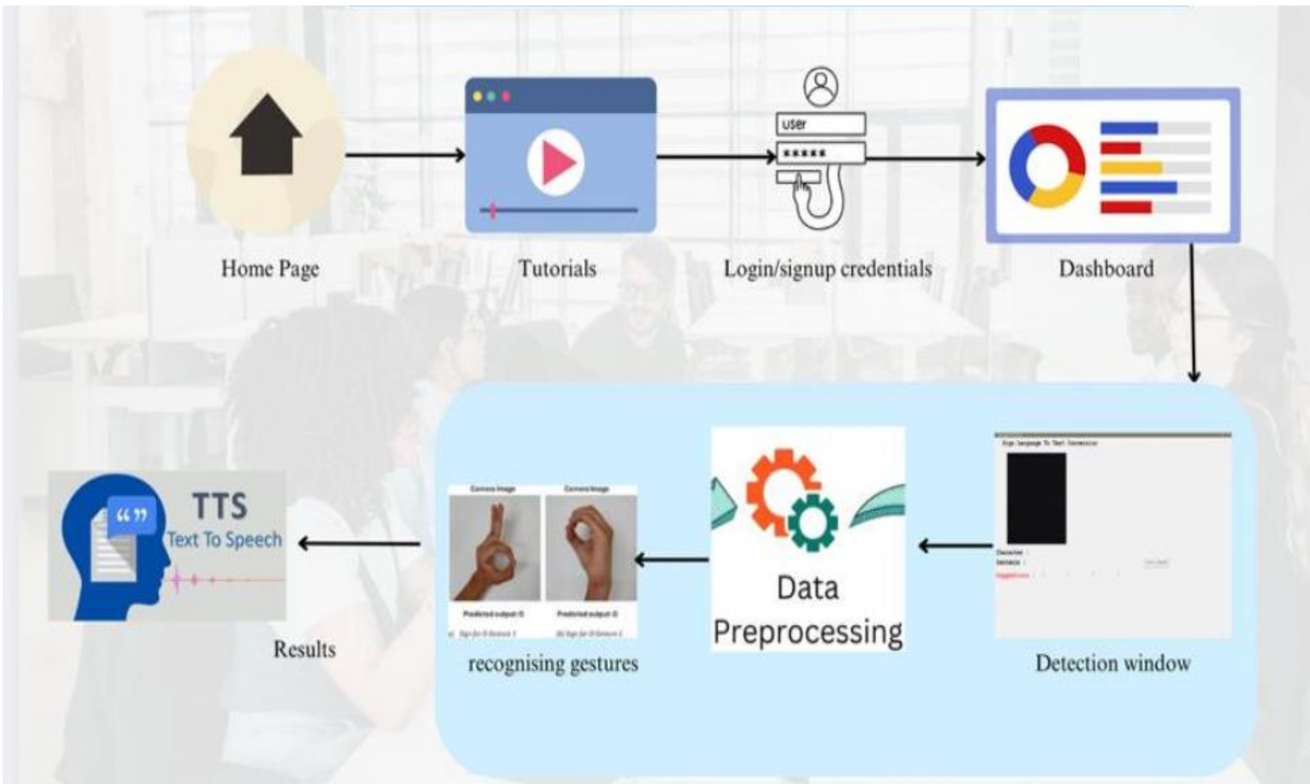

Figure 5 Architectural Design

3.2 High-Level Design

3.2.1 Use-Case Diagram

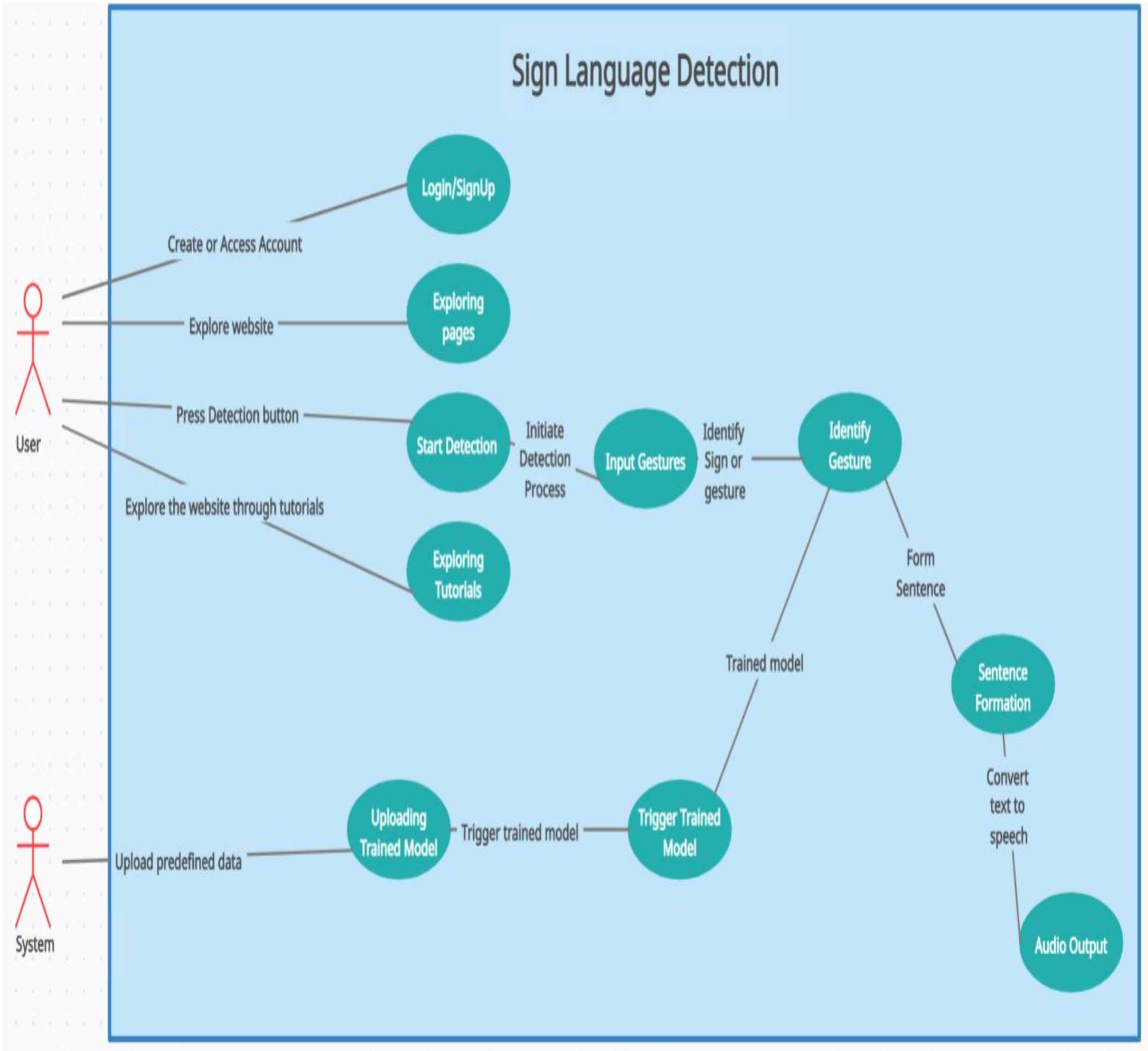

Figure 6 Use-Case diagram

3.2.2 Data Flow Diagram (Level-1)

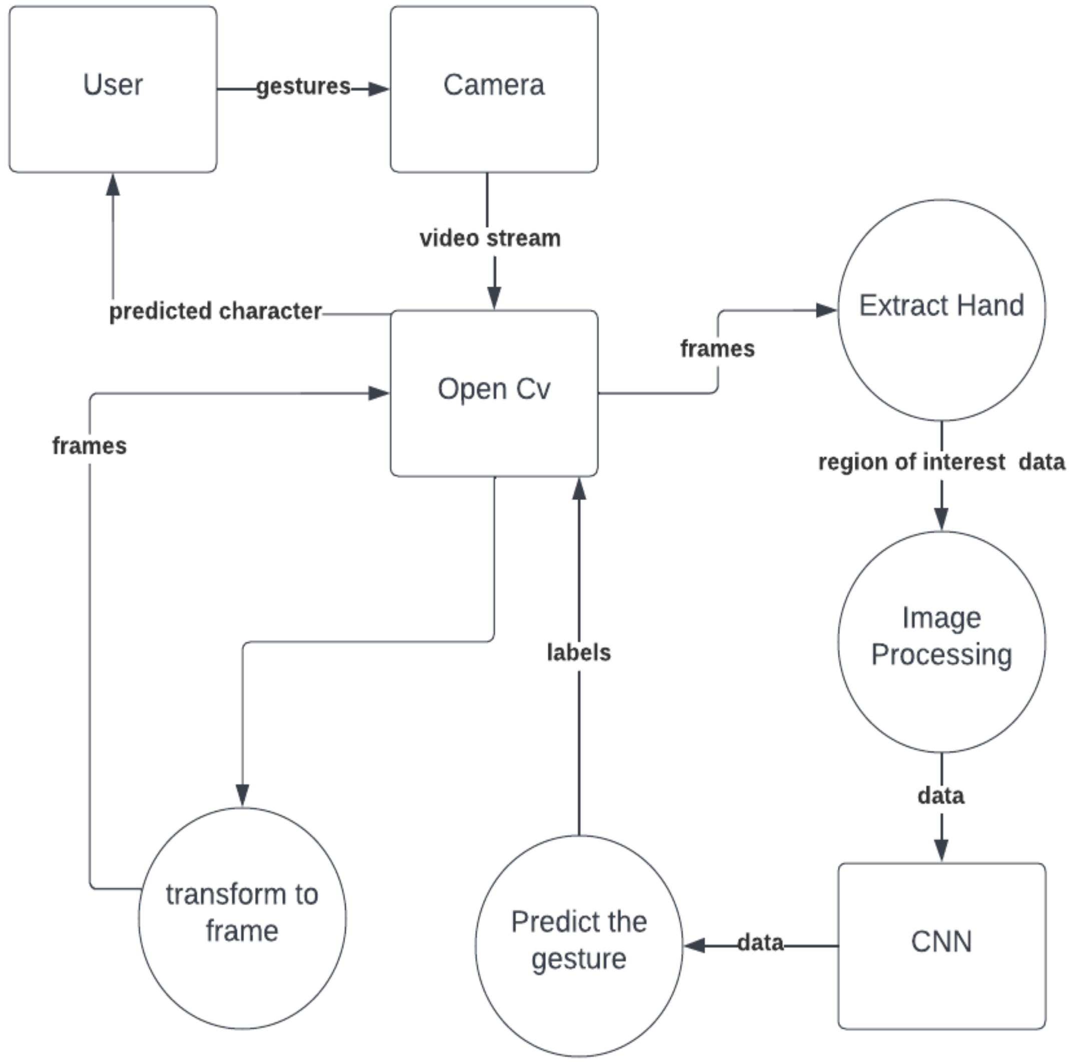

Figure 7 - DFD(Level-1) Diagram

3.3 Detailed Design

3.3.1 Sequence Diagram

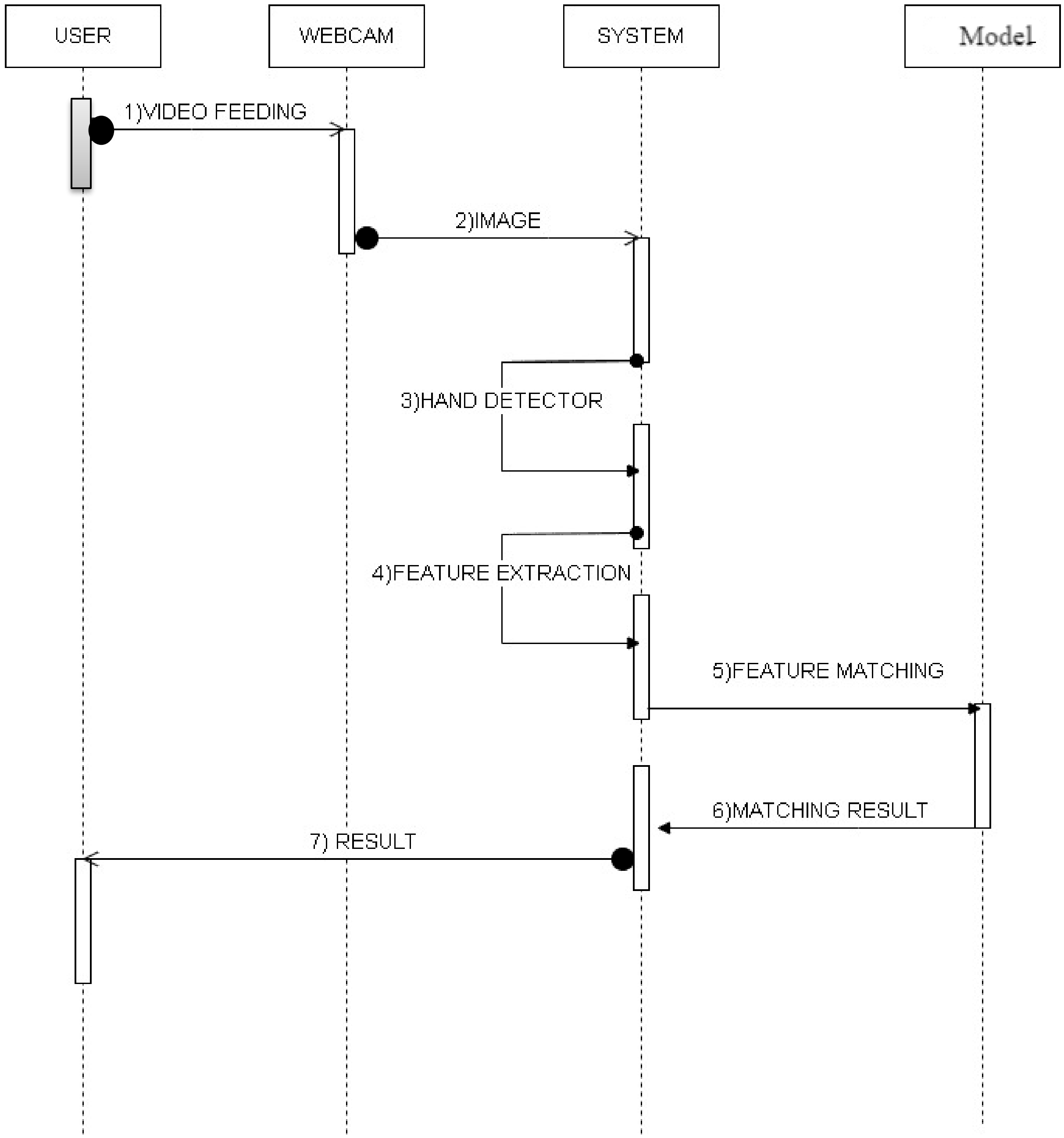

Figure 8 - Sequence Diagram

3.3.2 Activity Diagram

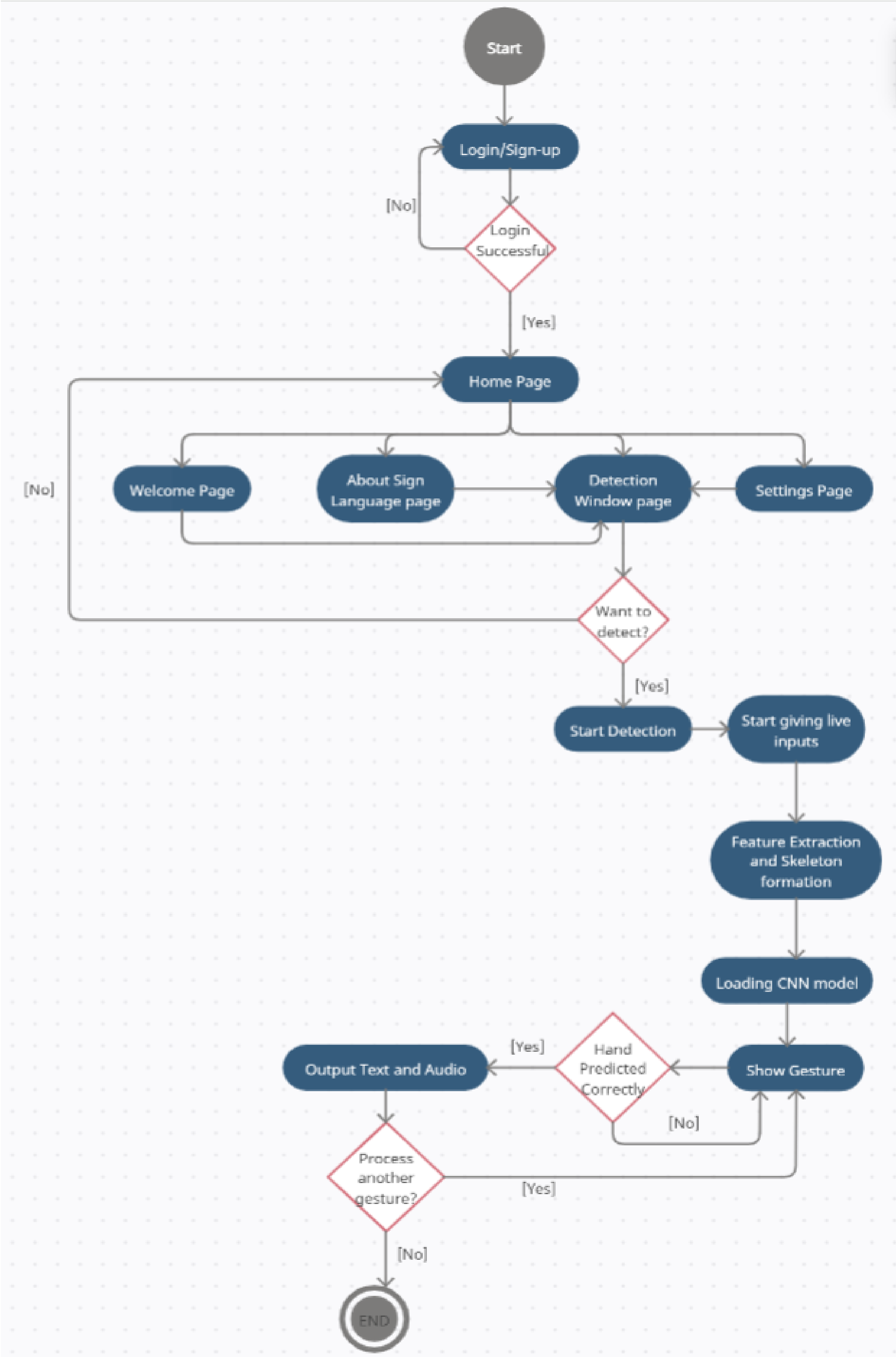

Figure 9 Activity Diagram

3.4 Algorithmic Design

3.4.1 Flowchart

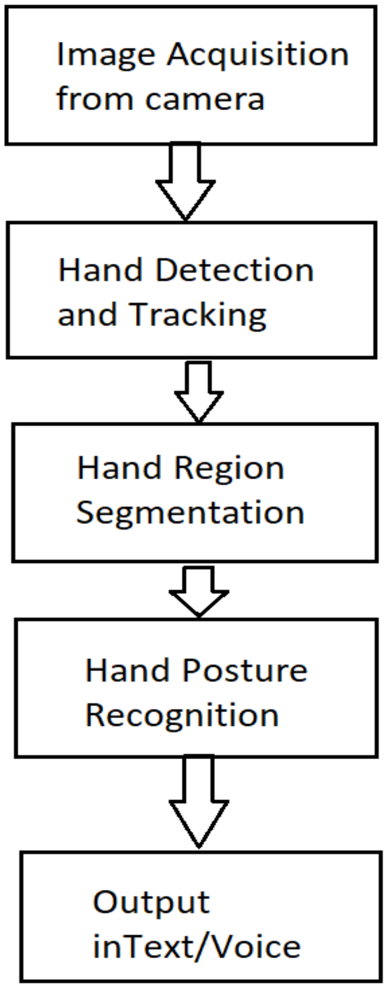

Figure 10 Flow-Chart

Chapter-4 Implementation

4.1 Data Acquisition

The different approaches to acquire data about the hand gesture can be done in the following ways:

It uses electromechanical devices to provide exact hand configuration, and position. Different glove-based approaches can be used to extract information. But it is expensive and not user friendly. In vision-based methods, the computer webcam is the input device for observing the information of hands and/or fingers. The Vision Based methods require only a camera, thus realizing a natural interaction between humans and computers without the use of any extra devices, thereby reducing costs

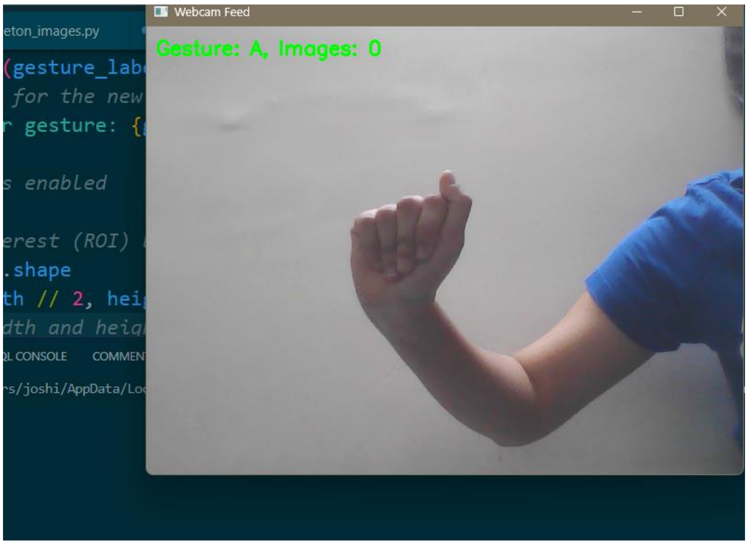

Figure 11 Vision Based method

4.2 Data pre-processing and Feature extraction

In this approach for hand detection, firstly we detect hand from image that is acquired by webcam and for detecting a hand we used media pipe library which is used for image processing. So, after finding the hand from image we get the region of interest (Roi) then we cropped that image using OpenCV library. We have collected images of different signs of different angles for sign letter A to Z.

Figure 12 Captured gesture

- In this method there are many loop holes like your hand must be ahead of clean soft background and that is in proper lightning condition then only this method will give good accurate results but in real world we don't get good background everywhere and we don't get good lightning conditions too.
- So to overcome this situation we try different approaches then we reached at one interesting solution in which firstly we detect hand from frame using mediapipe and get the hand landmarks of hand present in that image then we draw and connect those landmarks in simple white image

Mediapipe Landmark System:

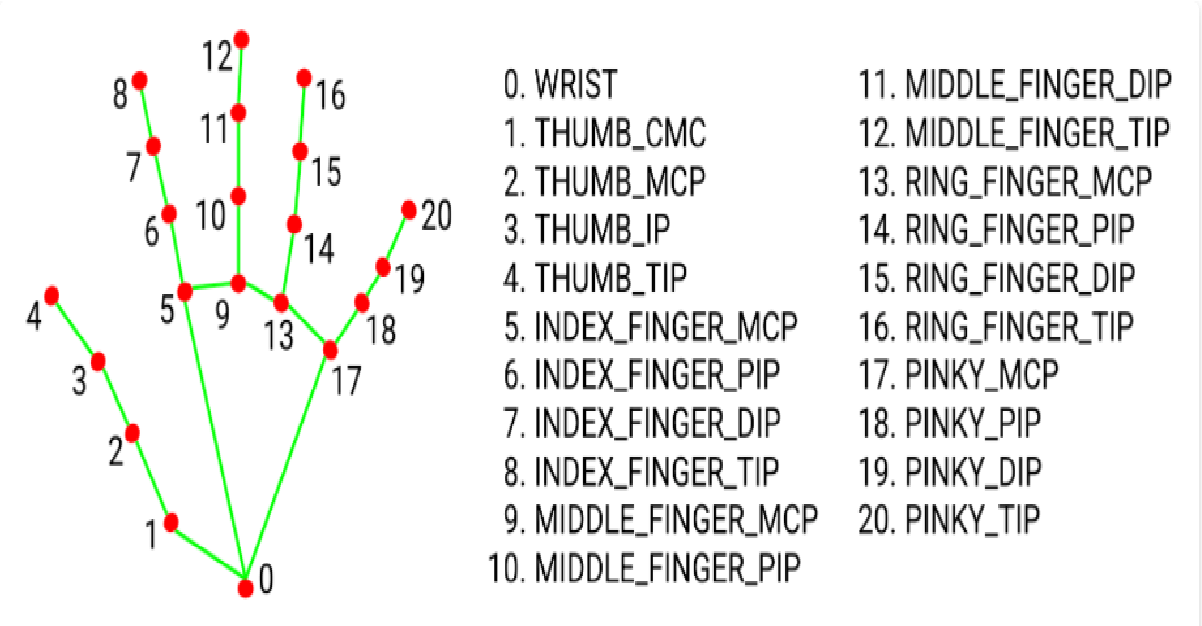

Figure 13 Mediappe Landmarking System

Now we will get this landmark points and draw it in plain white background using opencv library.

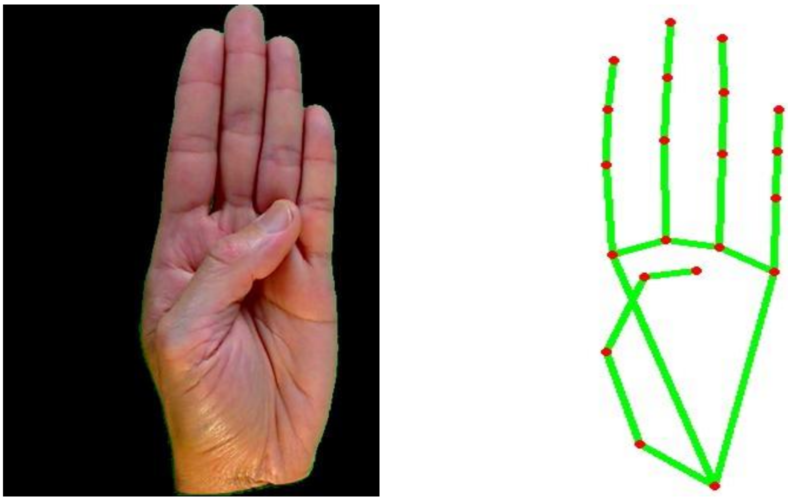

Figure 14 Drawing utils-1

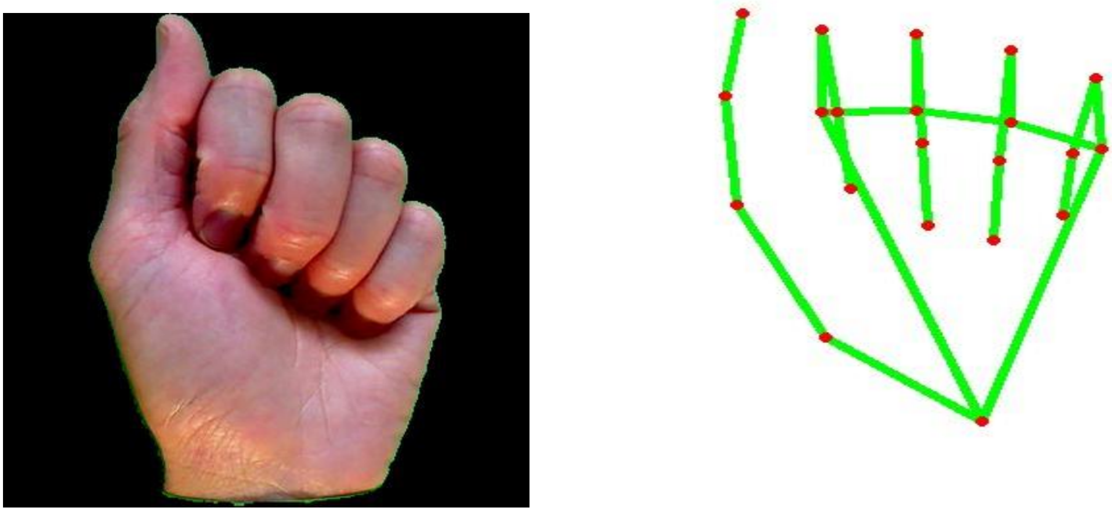

Figure 15 Drawing Utils-2

4.3 Gesture Classification

4.3.1 Convolutional Neural Network (CNN)

CNN is a class of neural networks that are highly useful in solving computer vision problems. They found inspiration from the actual perception of vision that takes place in the visual cortex of our brain. They make use of a filter/kernel to scan through the entire pixel values of the image and make computations by setting appropriate weights to enable detection of a specific feature. CNN is equipped with layers like convolution layer, max pooling layer, flatten layer, dense layer, dropout layer and a fully connected neural network layer. Unlike regular Neural Networks, in the layers of CNN, the neurons are arranged in 3 dimensions: width, height, depth. The neurons in a layer will only be connected to a small region of the layer (window size) before it, instead of all of the neurons in a fully-connected manner. Moreover, the final output layer would have dimensions(number of classes), because by the end of the CNN architecture we will reduce the full image into a single vector of class scores.

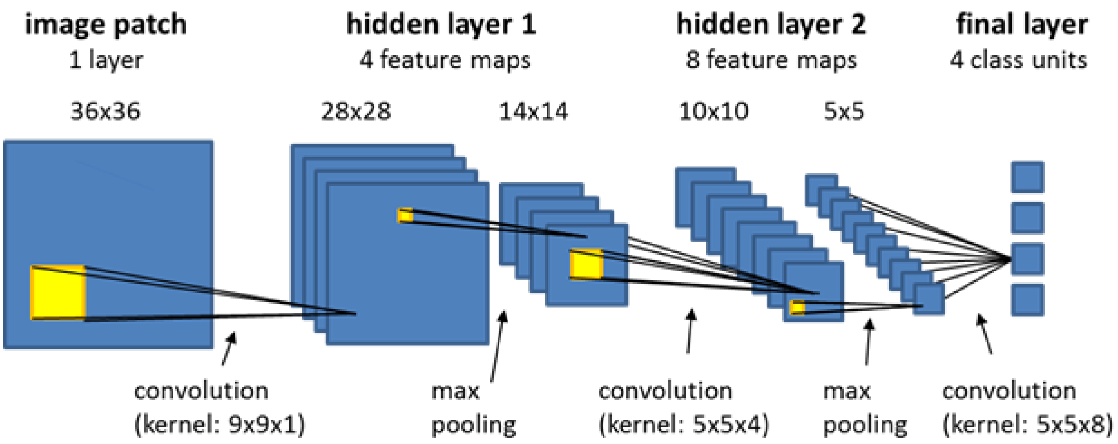

Figure 16 CNN Architecture

4.3.2 Convolutional Layer:

- In convolution layer I have taken a small window size [typically of length 5*5] that extends to the depth of the input matrix.
- The layer consists of learnable filters of window size. During every iteration I slid the window by stride size [typically 1], and compute the dot product of filter entries and input values at a given position.
- As I continue this process well create a 2-Dimensional activation matrix that gives the response of that matrix at every spatial position.
- That is, the network will learn filters that activate when they see some type of visual feature such as an edge of some orientation or a blotch of some colour.

4.3.3 Pooling Layer:

- We use pooling layer to decrease the size of activation matrix and ultimately reduce the learnable parameters.
- There are two types of pooling
- **Max Pooling:**
  - In max pooling we take a window size [for example window of size 2*2], and only taken the maximum of 4 values.
  - Well lid this window and continue this process, so well finally get an activation matrix half of its original Size.
- **Average Pooling:**
  - In average pooling we take average of all Values in a window.

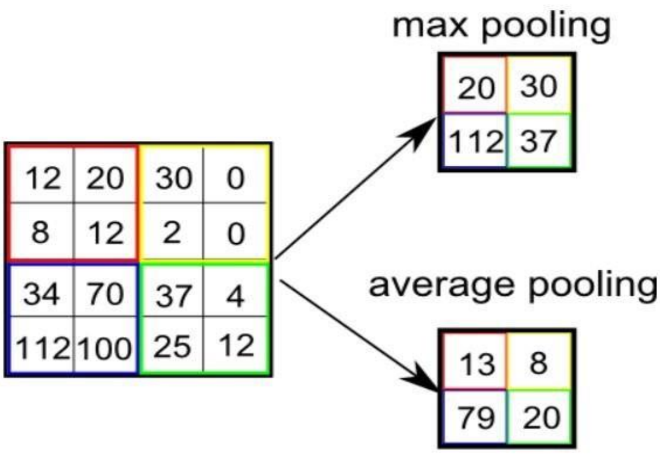

Figure 17 Pooling

4.3.4 Fully Connected Layer:

In convolution layer neurons are connected only to a local region, while in a fully connected region, well connect the all the inputs to neurons.

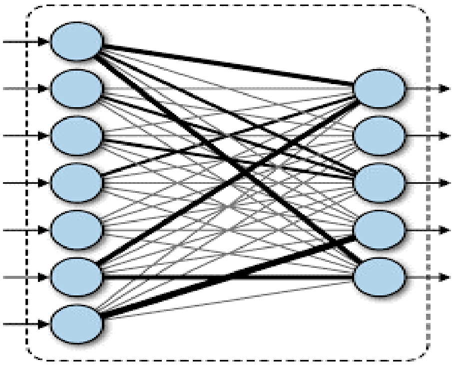

Figure 18 Fully Connected Layer

The preprocessed 180 images/alphabet will feed the keras CNN model.

Because we got bad accuracy in 26 different classes thus, We divided whole 26 different alphabets into 8 classes in which every class contains similar alphabets:

[y,j]

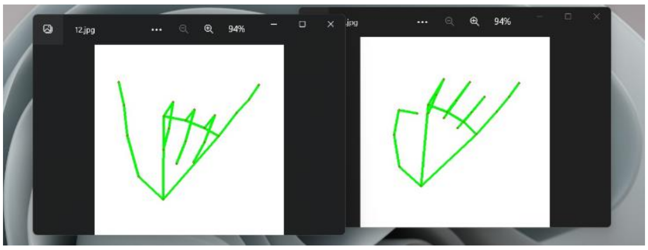

Figure 19 Classification 1

[c,o]

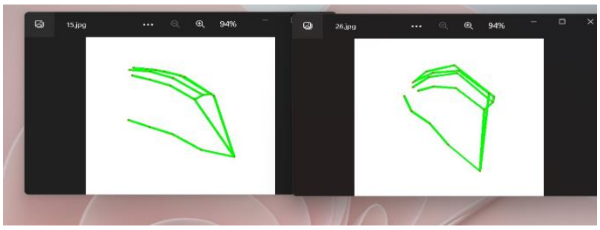

Figure 20 Classification 2

[g,h]

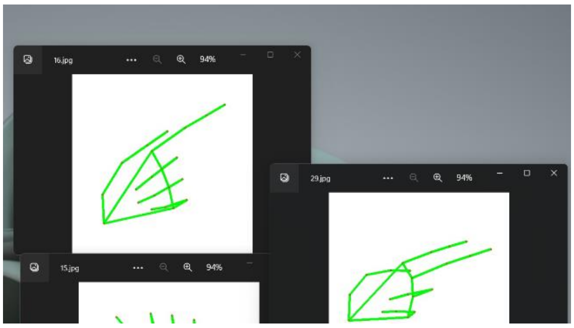

Figure 21 Classification 3

[b,d,f,I,u,v,k,r,w]

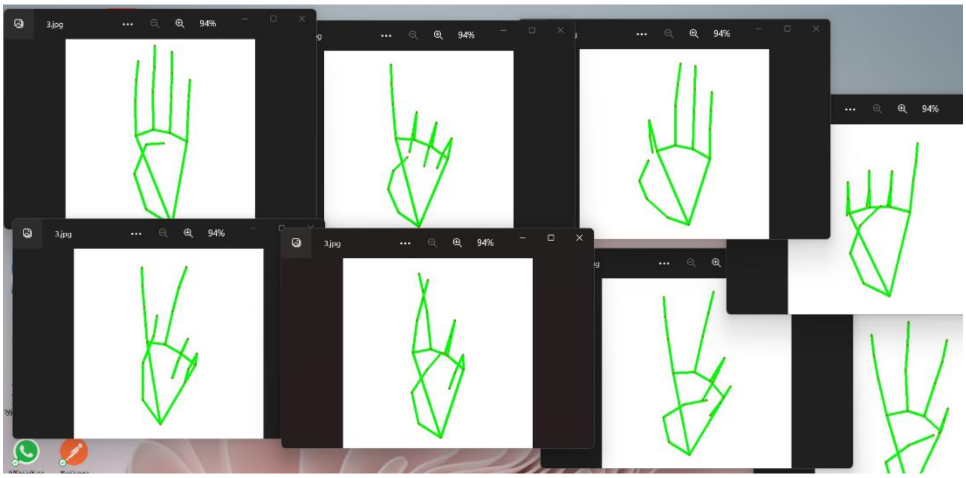

Figure 22 Classification 4

[p,q,z]

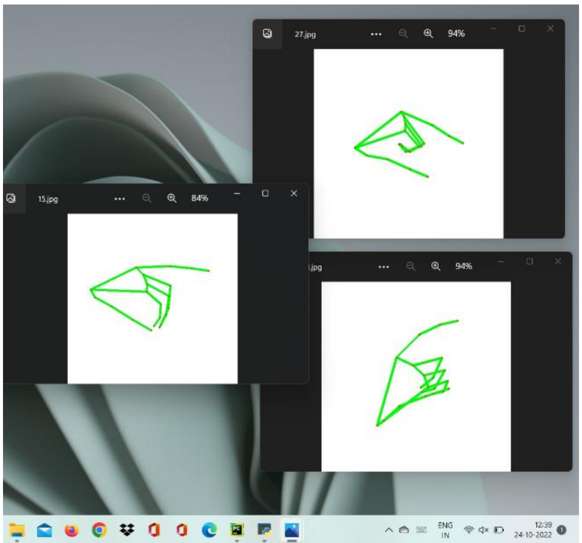

Figure 23 Classification 5

[a,e,m,n,s,t]

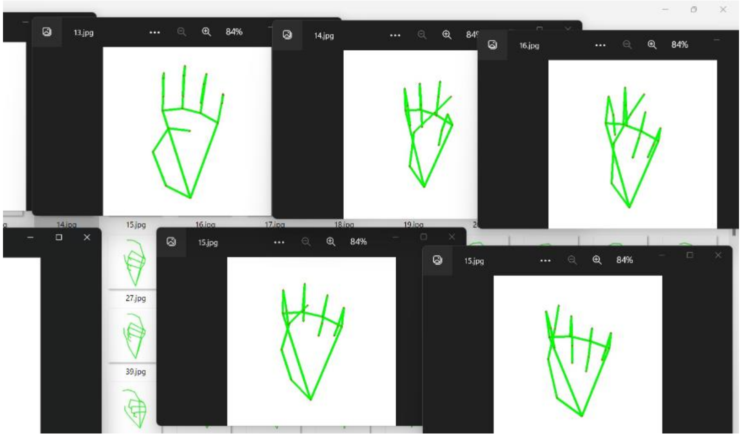

Figure 24 Classification 6

All the gesture labels will be assigned with a probability. The label with the highest probability will treated to be the predicted label. So when model will classify [aemnst] in one single class using mathematical operation on hand landmarks we will classify further into single alphabet a or e or m or n or s or t.

4.4 Text and Speech Translation

The model translates known gestures into words. we have used pyttsx3 library to convert the recognized words into the appropriate speech. The text-to-speech output is a simple workaround, but it's a useful feature because it simulates a real-life dialogue.

Chapter 5: Testing

5.1 Test Cases

Table 2 Test Cases

TC ID	TC Description	Input/ Action	Expected Result	Pass/Fail
01	Verify navigation to all pages	Click on navigation links (e.g., Home, Detect, Collect Data)	Verify navigatn	Pass
02	Check unexpected actions	Random clicks, or rapid repeated clicks	Website doesn’t crash; appropriate messages are shown	Pass
03	Check the representation of results	use real-time input	Detected letters/words/sentences displayed accurately on the screen	Pass
04	Test TTS output clarity	Input detected sentence for TTS	Voice output matches displayed text with proper clarity	Pass
05	Login with valid credential	Valid username and password	User is redirected to the dashboard	Pass
06	Login with invalid credential	Invalid username or password	Error message "Invalid credentials"	Pass
07	Sign up with valid details	Valid username, email, password	Account is created, redirected to login page	Pass

5.2 Outputs

After Implementing the cnn algorithm we made gui using python Tkinter and added Suggestions also to make the process smooth for user.

Here are some snapshots of the results:

5.2.1 Home Page

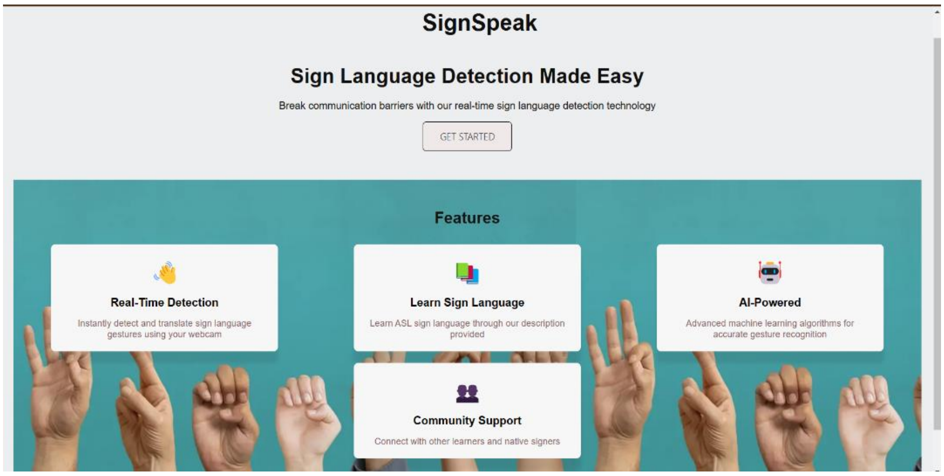

Figure 25 Website img 1

5.2.2 Tutorials Page

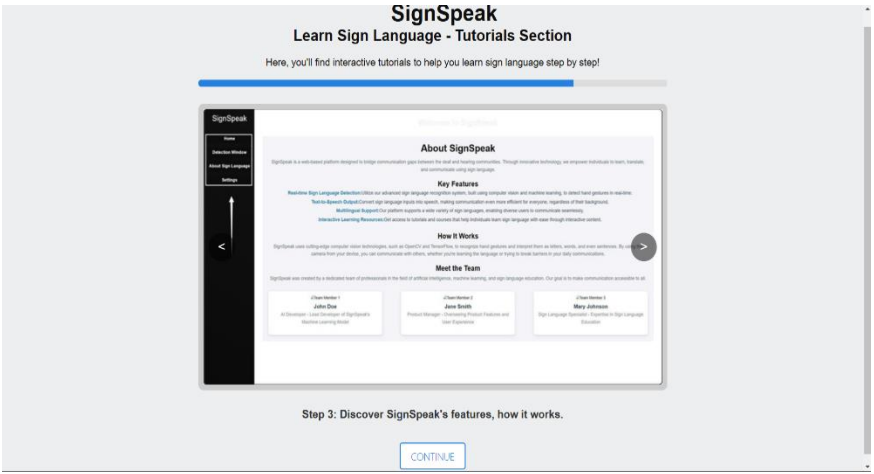

Figure 26 Website img 2

5.2.3 Login Page

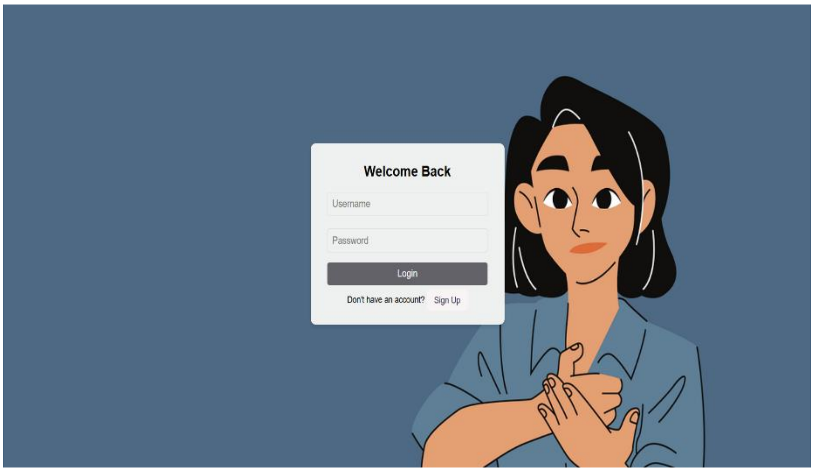

Figure 27 Website img 3

5.2.4 Sign Up page

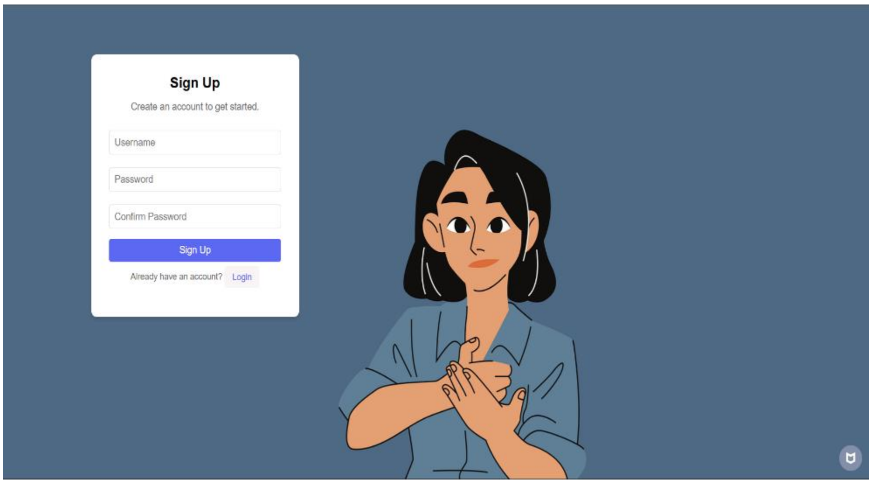

Figure 28 Website img 4

5.2.5 Dashboard Home Page

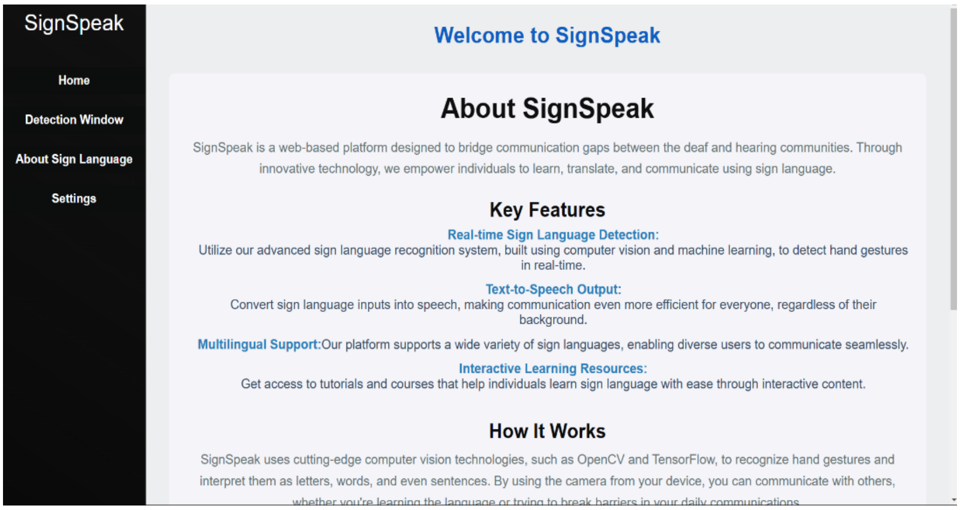

Figure 29 Website img 5

5.2.6 Detection Window Page

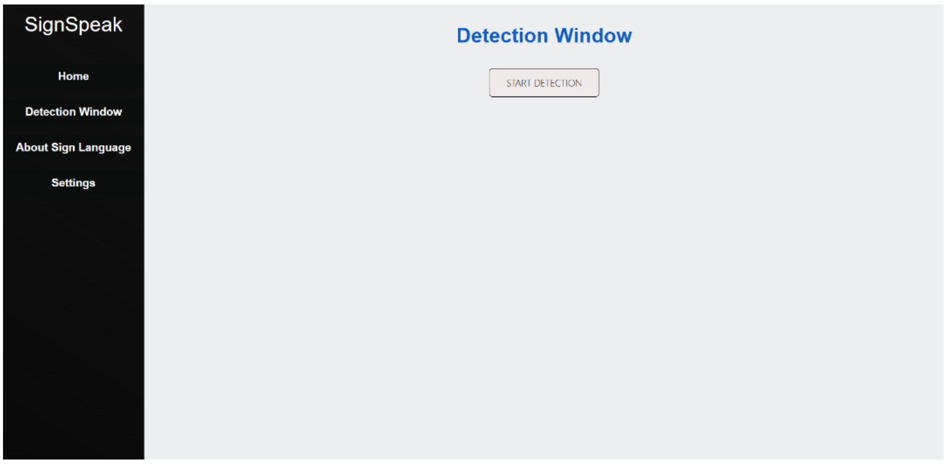

Figure 30 Website img 6

5.2.7 Detection Window

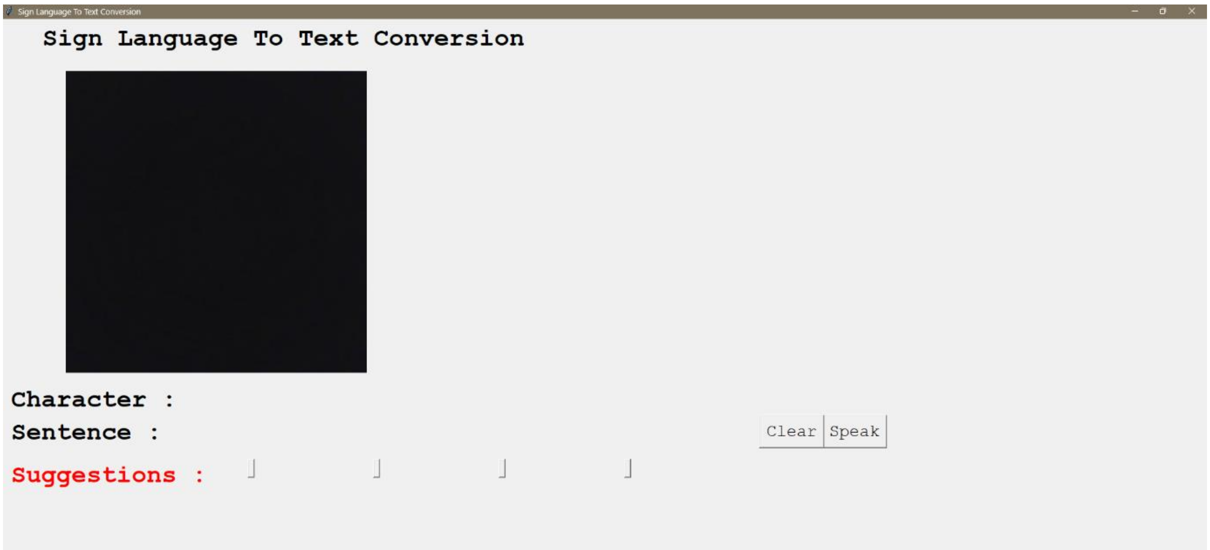

Figure 31 Website img 7

5.2.8 About Page

Figure 32 Website img 8

5.2.8 Settings Page

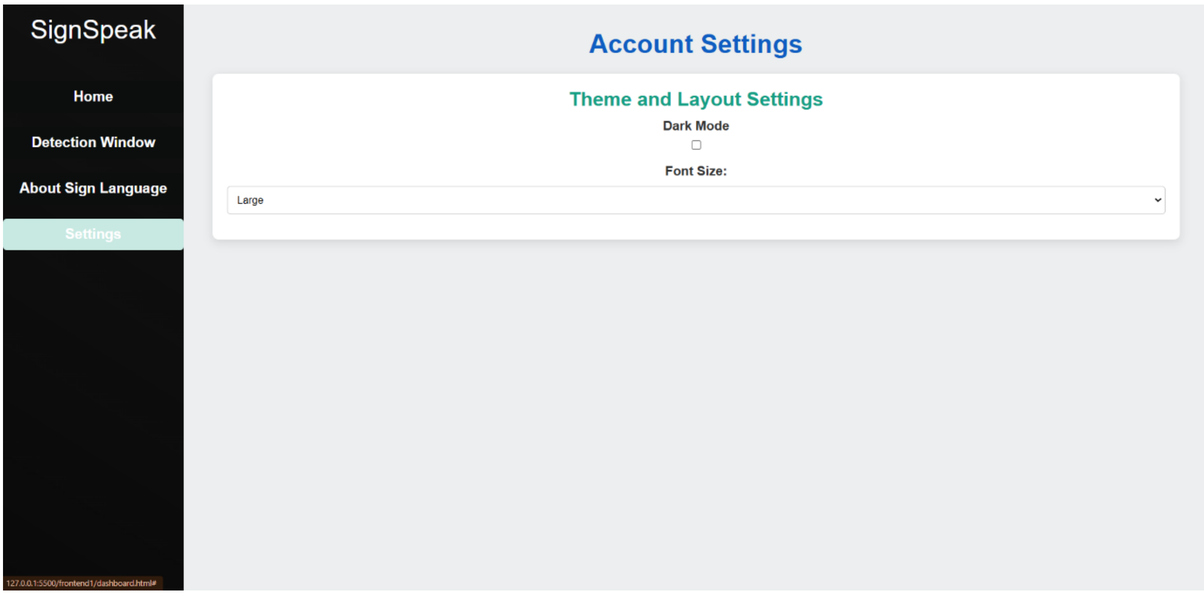

Figure 33 Website img 9

5.3 Accuracy Score and Graphs

After implementing the model and predicting values we have got a very good accuracy score of 88%.

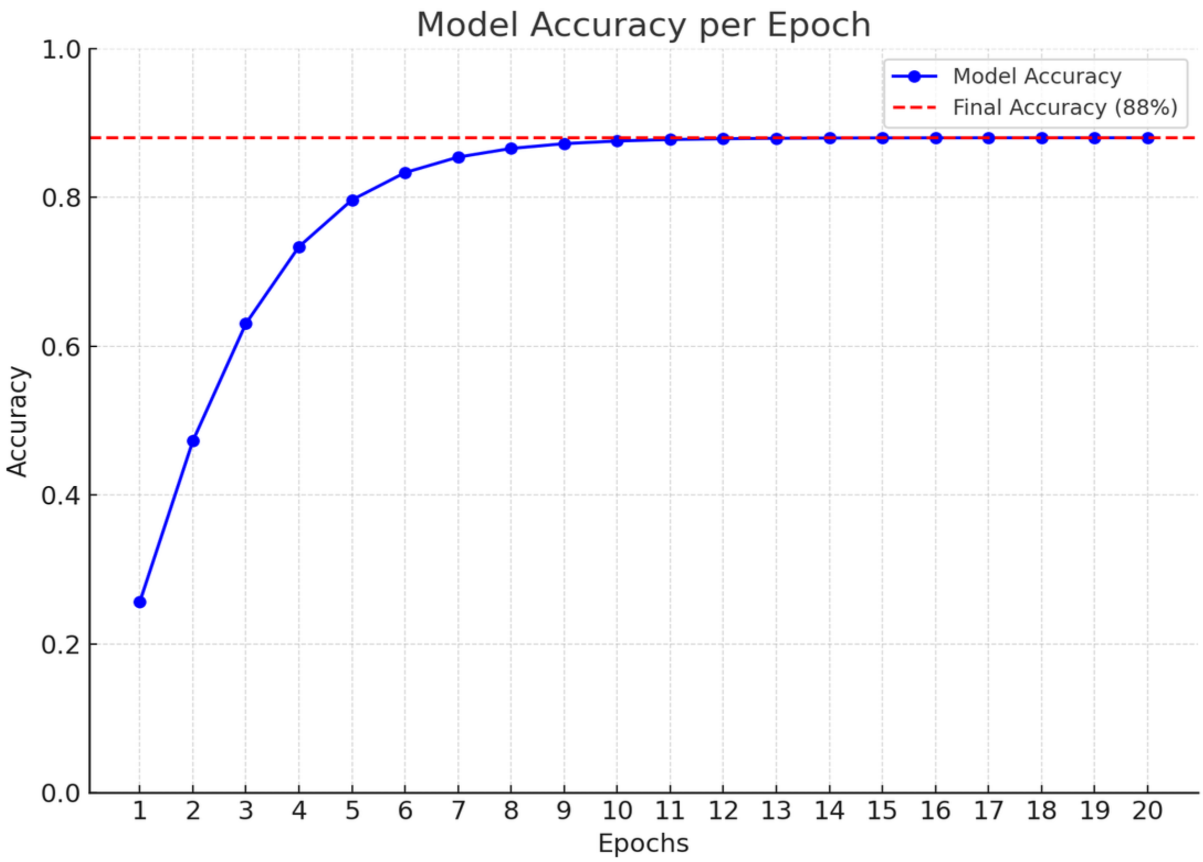

Figure 34 Accuracy Score

Conclusion

Finally, we are able to predict any alphabet[a-z] with **88%** Accuracy (with and without clean background and proper lightning conditions) through our method. And if the background is clear and there is good lightning condition then we got even above **90%** accurate results. In Future work we will make one android application in which we implement this algorithm for gesture predictions.

References

[1] Convolutional Neural Networks,
https://www.geeksforgeeks.org/introduction-convolution-neural-network/

[2] Sign Language to Text and Speech Conversion,
https://www.ijariit.com/manuscript/sign-language-to-text-and-speech-conversion/

[3] Conversion of Sign Language To Text And Speech Using Machine Learning Techniques,
https://www.researchgate.net/publication/335433017_Conversion_of_Sign_Language_To_Text_And_Speech_Using_Machine_Learning_Techniques

[4] Conversion of Sign Language into Text,
https://www.ripublication.com/ijaer18/ijaerv13n9_90.pdf