**SIGN LANGUAGE RECOGNITION TO**

**TEXT & SPEECH**

A PROJECT REPORT

BACHELOR OF TECHNOLOGY

in

Computer Science & Engineering

# By

(22L31A05C4)

MALLIPUDI AASHIK SAGAR

Under the Esteemed Guidance of

Dr. CH V V Ramana

# Associate Professor Department of Mechanical Engineering



VIGNAN’S INSTITUTE OF INFORMATION TECHNOLOGY

VISAKHAPATNAM

(Autonomous)

(Approved by AICTE, New Delhi, Accredited by NBA,

Affiliated to Jawaharlal Nehru Technological University, Vizianagaram)

Besides VSEZ, Duvvada, Vadlapudi Post, Gajuwaka Visakhapatnam - 530049, A.P., India

2024-2025



VIGNAN’S INSTITUTE OF INFORMATION TECHNOLOGY

VISAKHAPATNAM

# CERTIFICATE

This is to certify that the project report entitled –

**SIGN LANGUAGE RECOGNITION (TO TEXT & SPEECH)**

is a Bonafide record of the work done by **MALLIPUDI AASHIK SAGAR(22L31A05C4), CHINDADA DAVID LIVING STON (23L31A0543), BOLLIPOGU SHYAM (23L31A0531),**

**BV THRIDHAR (23L31A0537),** **BHUKYA PAVAN KUMAR NAIK (23L31A0524)**

## Of COMPUTER SCIENCE & ENGINEERING Department ,

## Vignan’s Institute of Information Technology, Visakhapatnam.

Head of the Department Project Guide

Dr. S. Dinesh Reddy Dr. CH V V Ramana

Professor and Head of Department Associate Professor

Department of CSE Department of Mechanical Engineering

Vignan’s Institute of Information Technology

EXTERNAL EXAMINER

### **ACKNOWLEDGEMENT**

It is privilege for me to present on “**SIGN LANGUAGE RECOGNITION TO TEXT & SPEECH**” submitted to engineering section, Vignan’s institute of information technology (autonomous) Duvvada, in partial fulfilment of requirement for award of “B-TECH IN COMPUTER SCIENCE ENGINEERING”.

I take this opportunity to express my deep and sincere indebtedness to our guide. Sri Dr. Ch. V V Ramana, Associate Professor in Computer Science Engineering for his guidance, support and encouragement in completion of this project work.

I intent to express my thanks with obedience to our Head of Computer Science Engineering. Mr. B. Dinesh Reddy, for his valuable suggestions and instant motivation in bringing out this project work.

I feel happy to convey my sincere thanks to our principal for his kind co-operation in providing all provisions for the successful completion of the project.

I am thankful to all other staff members of Computer Science Engineering department and workshop staff who helped me directly and indirectly in the completion of this project.

I am also thankful to my friends and my batch co-students for their interest, assistance and helpful suggestion. Lastly, I am grateful to one and all that help me in the completion of the project work.

**ABSTRACT**

Communication plays a crucial role in human interaction, yet for individuals with hearing and speech impairments, expressing thoughts becomes a major challenge. Sign language is their primary mode of communication, but understanding it requires specialized knowledge, creating a communication gap with the broader community. This project aims to bridge that gap by developing a system that converts American Sign Language (ASL) gestures into corresponding text and speech. Using a vision-based approach, the system captures hand gestures through a webcam, processes them with the help of Media pipe landmark detection, and classifies them using a Convolutional Neural Network (CNN) model. By mapping the gestures to alphabets and forming words, the system outputs the recognized text and simultaneously generates its audio using a text-to-speech engine. This application offers a cost-effective, accessible, and efficient solution that empowers deaf and mute individuals to communicate seamlessly with the hearing population, enhancing their inclusion and independence in society. With real-time processing, high accuracy even under varying lighting and background conditions, and a user-friendly interface, the project represents a significant step towards inclusive communication technology.

**KEY FEATURES**

1. **Real-time Hand Gesture Detection:**  
   Captures hand gestures instantly using a webcam, enabling real-time recognition and translation.
2. **Mediapipe-based Hand Landmark Extraction:**  
   Uses Google's Mediapipe framework to accurately detect hand landmarks, ensuring robustness across different backgrounds and lighting conditions.
3. **CNN-based Gesture Classification:**Employs a Convolutional Neural Network (CNN) model trained on pre-processed skeletal hand landmarks to classify American Sign Language (ASL) gestures with high accuracy.
4. **Text Conversion Module:**  
   Recognizes individual gestures, forms words by combining alphabets, and displays the output as readable text on the screen.
5. **Speech Conversion Module:**  
   Integrates a text-to-speech (TTS) engine (pyttsx3 library) to convert recognized text into clear, audible speech, facilitating two-way communication.
6. **Robust Preprocessing Techniques:**  
   Applies image preprocessing techniques such as hand segmentation, grey scaling, adaptive thresholding, and landmark drawing to improve model performance and reliability.
7. **Background and Lighting Adaptability:**  
   Ensures accurate detection and classification even in varying lighting conditions and complex backgrounds by focusing on hand landmarks rather than raw images.
8. **User-Friendly GUI Interface:**  
   Provides an intuitive desktop application built using Python Tkinter for easy interaction, supporting gesture recognition, text display, and speech output.
9. **Error Handling and Suggestions:**Includes features like special gestures for adding space between words and moving to the next alphabet, enhancing smooth user experience.
10. **High Accuracy:**  
    Achieves gesture recognition accuracy of up to 97% in normal conditions and up to 99% under clean backgrounds and good lighting.

**BENEFITS**

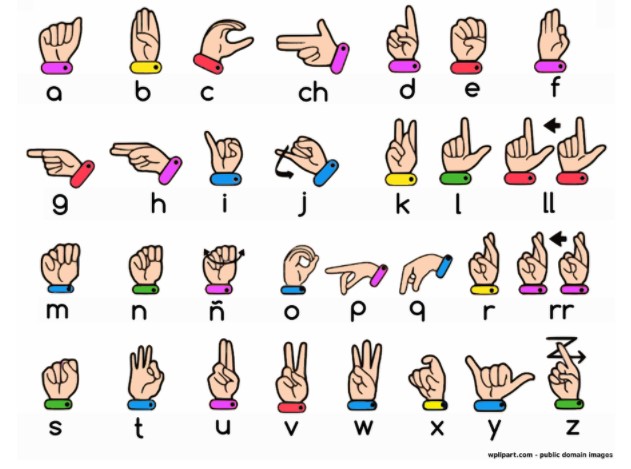
1. **Bridges Communication Gaps:**  
   Enables deaf and mute individuals to communicate easily with people who do not understand sign language, promoting social inclusion.
2. **Real-Time Translation:**  
   Provides immediate translation of gestures into text and speech, ensuring smooth and fast interaction without delays.
3. **High Accuracy and Reliability:**Delivers gesture recognition accuracy up to 97–99%, making the system dependable even in real-world conditions with varied backgrounds and lighting.
4. **Cost-Effective and Accessible:**  
   Requires only a basic webcam and standard computer setup without the need for expensive sensors or hardware, making it affordable for widespread use.
5. **User-Friendly Interface:**  
   Features a simple GUI that even non-technical users can operate easily, enhancing usability for all age groups.
6. **Adaptability to Different Environments:**Maintains performance across different lighting and background conditions by using hand landmarks instead of raw image processing.
7. **Support for Learning and Education:**  
   Can be used as a learning tool for those who wish to understand and practice American Sign Language (ASL).
8. **Flexible and Expandable:**  
   The system is designed in a modular way, allowing future upgrades such as adding more gestures, full-sentence construction, or supporting multiple sign languages.
9. **Improves Independence:**  
   Empowers deaf and mute users to communicate independently without needing human interpreters, enhancing their confidence and self-reliance.
10. **Potential for Multi-Platform Deployment:**  
    Can be further extended to mobile and web platforms, increasing its availability and ease of access.

**FUTURE IMPROVEMENTS**

1. **Sentence and Paragraph Formation**:  
   Extend the system from recognizing individual alphabets to recognizing complete words and forming meaningful sentences automatically.
2. **Support for Dynamic Gestures**:  
   Upgrade the model to detect dynamic hand movements (continuous gestures) for full sign language sentence translation, not just static fingerspelling.
3. **Mobile Application Development**:  
   Develop a lightweight Android/iOS app to make the system portable and usable anytime with smartphone cameras.
4. **Multi-Language Speech Output**:  
   Enable the text-to-speech module to support multiple languages, allowing translations into languages other than English.
5. **Voice-Based Reverse Translation**:  
   Incorporate speech-to-sign language translation, where spoken words are converted into animated sign language gestures.
6. **Integration with Augmented Reality (AR)**:  
   Use AR to overlay detected sign language translations visually in real-time, making the system more interactive and immersive.
7. **Machine Learning Model Optimization**:  
   Implement advanced deep learning techniques to reduce model size and increase processing speed for real-time mobile and web deployment.
8. **Custom Sign Addition**:  
   Allow users to add and train new custom signs specific to local or regional variations of sign language.
9. **Improved Gesture Differentiation**:  
   Enhance the system’s ability to differentiate between very similar signs more accurately through improved feature extraction techniques.
10. **Cloud-Based Recognition System**:  
    Create a cloud service where heavy processing is offloaded to servers, enabling faster and more powerful gesture recognition even on low-end devices.

**INTRODUCTION**

Communication is essential for expressing thoughts, emotions, and ideas. While verbal communication is the most common form, individuals who are deaf or mute rely heavily on sign language as their primary means of interaction. However, not everyone in society understands sign language, leading to significant communication barriers.  
The **Conversion of Sign Language to Text and Speech** project aims to bridge this gap by developing a system that can automatically recognize hand gestures of American Sign Language (ASL) and convert them into corresponding text and speech outputs. Using a vision-based approach integrated with machine learning techniques, the system captures hand gestures via webcam, processes them through hand landmark extraction, classifies them using a trained Convolutional Neural Network (CNN), and generates text and audible speech outputs.  
This project not only empowers deaf and mute individuals by giving them a voice but also enhances inclusivity by allowing non-sign language users to understand and interact with them effectively.



**OBJECTIVE**

The primary objectives of the project are:

* To design and develop a real-time system capable of detecting and recognizing static ASL hand gestures accurately.
* To translate recognized gestures into corresponding English alphabets, and then into meaningful text and speech outputs.
* To build a system that works reliably across varying backgrounds and lighting conditions without the need for specialized or costly hardware.
* To create a user-friendly desktop application that can be used by individuals with no technical expertise.
* To contribute to social inclusivity by reducing communication barriers for individuals with hearing and speech impairments.

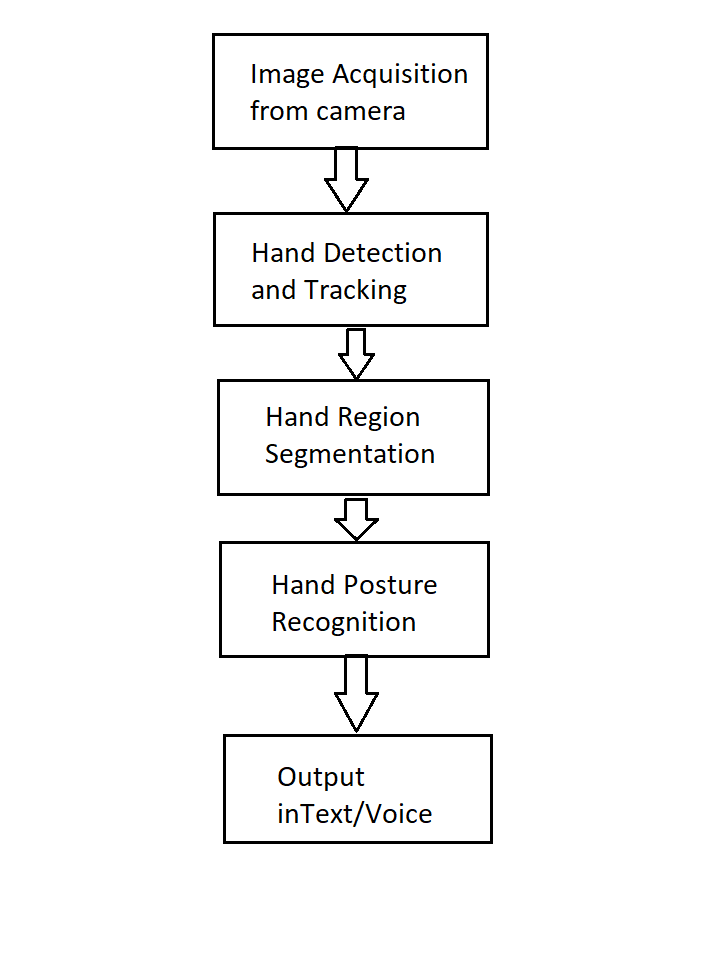
**SCOPE**

The scope of the **Conversion of Sign Language to Text and Speech** project includes:

* Developing a complete real-time gesture recognition pipeline starting from hand detection to gesture classification, text generation, and audio output.
* Training the model specifically for the American Sign Language (ASL) alphabet gestures (A-Z), allowing users to form words by combining letters.
* Providing both text display and speech output, ensuring that the system supports users in different communication scenarios.
* Designing a scalable system where additional gestures, full words, dynamic signs, and multi-language support can be added in the future.
* Focusing on affordable, accessible technologies, ensuring that the system requires only basic hardware like a webcam and standard computer setup.

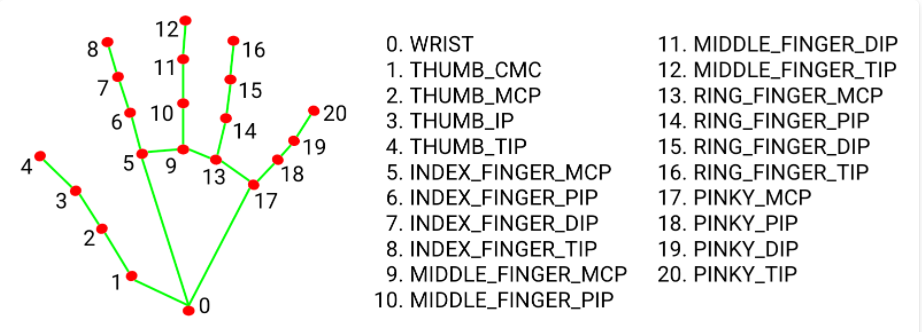
**PROCEDURE**

**Flow Chart Diagram:**



The development of the project follows these major steps:

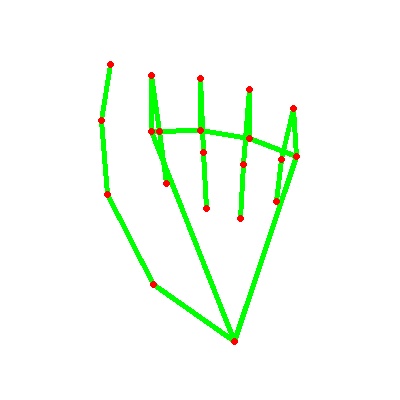
1. **Data Acquisition**:  
   The webcam captures real-time images of hand gestures. Instead of using raw images, the Mediapipe library is utilized to detect hand landmarks, which provide a structured and standardized format for each gesture.
2. **Preprocessing and Feature Extraction**:  
   After identifying the hand region, the system applies preprocessing steps such as grayscale conversion, Gaussian blurring, and thresholding to enhance important features while minimizing noise. The extracted landmarks are drawn onto a plain white background to ensure the model focuses only on hand structure, eliminating background complexities.
3. **Model Training**:  
   A Convolutional Neural Network (CNN) is trained on the skeleton images generated from hand landmarks. Initially, individual alphabet images are categorized into groups based on visual similarities for better classification accuracy. The CNN learns to identify patterns unique to each group and then further distinguishes individual letters through mathematical landmark operations.



**Mediapipe Landmark System**



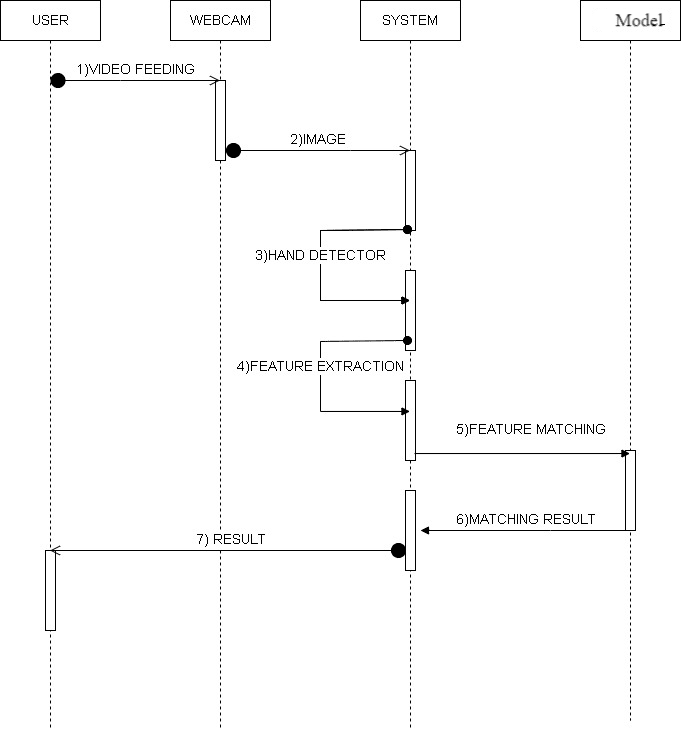
**Mapping of Mediapipe Landmarks on Hand gestures**



**Landmark points on plain white background using OpenCV library**

1. **Text and Speech Translation**:  
   Once a gesture is recognized, it is translated into a corresponding alphabet. Multiple alphabets are concatenated to form words. Recognized text is displayed on the GUI, and simultaneously, the Pyttsx3 library converts the text into speech, providing an audio output.
2. **Graphical User Interface (GUI) Development**:  
   A simple and intuitive GUI is developed using Python’s Tkinter library. It allows users to interact with the system, view recognized text, and hear the corresponding speech in real-time.

**Sequence Diagram for the Model**



**TECH STACK USED**

The technologies used for developing the project are:

* **Programming Language**:
  + Python 3.9: Chosen for its simplicity, vast library support, and ease of rapid development.
* **Libraries and Frameworks**:
  + **OpenCV**: For image acquisition and preprocessing tasks.
  + **Mediapipe**: For real-time hand landmark detection and extraction.
  + **TensorFlow and Keras**: For building, training, and testing the Convolutional Neural Network (CNN) model.
  + **Pyttsx3**: For converting recognized text into speech output.
  + **NumPy**: For efficient numerical computations during preprocessing.
  + **Tkinter**: For creating the graphical user interface (GUI).

**REQUIREMENTS**

**Hardware Requirements:**

* **Webcam**: Integrated or USB external webcam for real-time hand gesture capturing.
* **Computer/Laptop**:
  + Minimum 4GB RAM.
  + Intel i3 Processor or higher recommended.
  + At least 10GB free disk space.

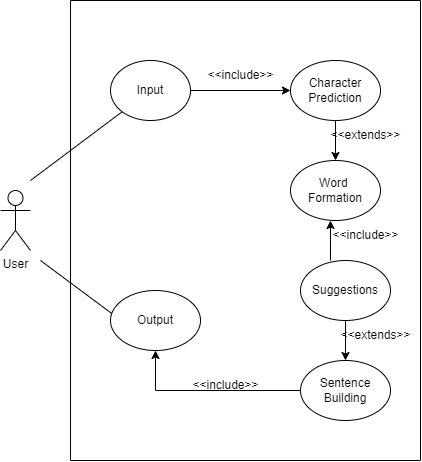
**Software Requirements:**

* **Operating System**: Windows 8 or above.
* **Programming Environment**: Python 3.9 environment installed.
* **Libraries/Frameworks**: OpenCV, Mediapipe, TensorFlow, Keras, NumPy, Pyttsx3, Tkinter.
* **IDE/Code Editor**: PyCharm (preferred), Visual Studio Code, or Jupyter Notebook.
* **Additional Tools**:
  + Pip package manager for installing dependencies.
  + Audio drivers for enabling speech output.

DESIGN THINKING MODEL

**EMPATHY**

In the empathy phase, we focus on understanding the needs, challenges, and experiences of deaf and mute individuals. Communication barriers often lead to isolation, dependency on interpreters, and difficulty in expressing basic needs.  
By putting ourselves in their situation, we recognize that traditional methods like sign language are effective among those who understand it, but ineffective in general society. Many hearing individuals are unfamiliar with sign language, which limits social, educational, and professional opportunities for the deaf and mute community.  
Thus, the need arises for a real-time system that translates sign language into text and speech, making communication smoother, faster, and accessible to everyone.



**Use Case Diagram for the Model**

**DEFINE**

Problem Statement:  
**Deaf and mute individuals face communication challenges because most of the population is not familiar with sign language, limiting their interactions in daily life.**

Key Needs Identified:

* A system that can accurately recognize sign language gestures.
* A system that translates those gestures into readable text and audible speech.
* A low-cost, easily accessible solution that does not require specialized equipment or training.

Thus, the defined problem is:  
**"How can we enable deaf and mute individuals to communicate more effectively with the hearing population in real-time without needing human interpreters?"**

**IDEATE**

Possible ideas brainstormed to solve the problem:

* Use wearable devices like smart gloves embedded with sensors.
* Develop a mobile app using camera-based gesture recognition.
* Create a real-time desktop system using computer vision and AI to detect hand gestures.
* Use landmark-based skeletal analysis to overcome lighting and background issues.
* Integrate text-to-speech modules to provide audio output for the translated gestures.

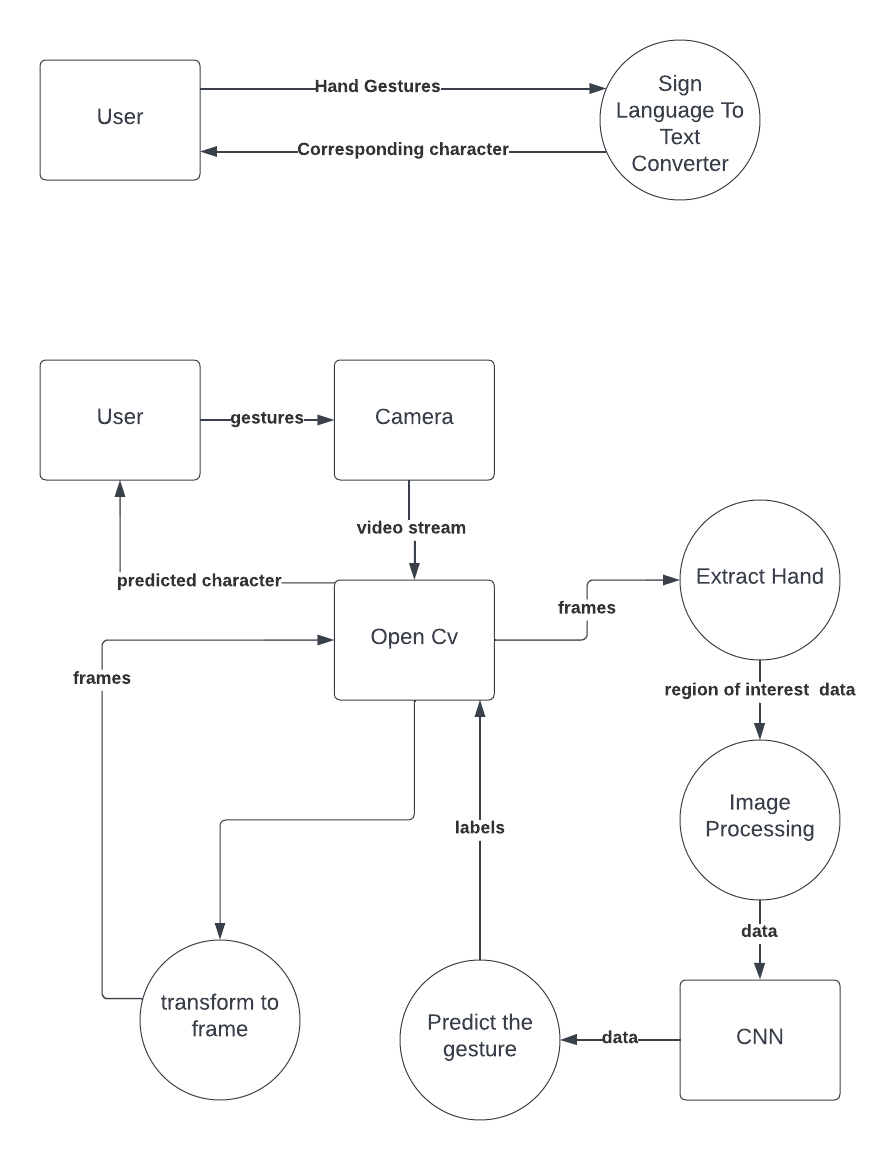
After evaluating feasibility, cost, and ease of use, the best idea chosen was:  
**Building a real-time webcam-based system using Mediapipe for hand landmark detection and a CNN model for gesture classification, combined with text-to-speech generation.**

**SOLUTION**

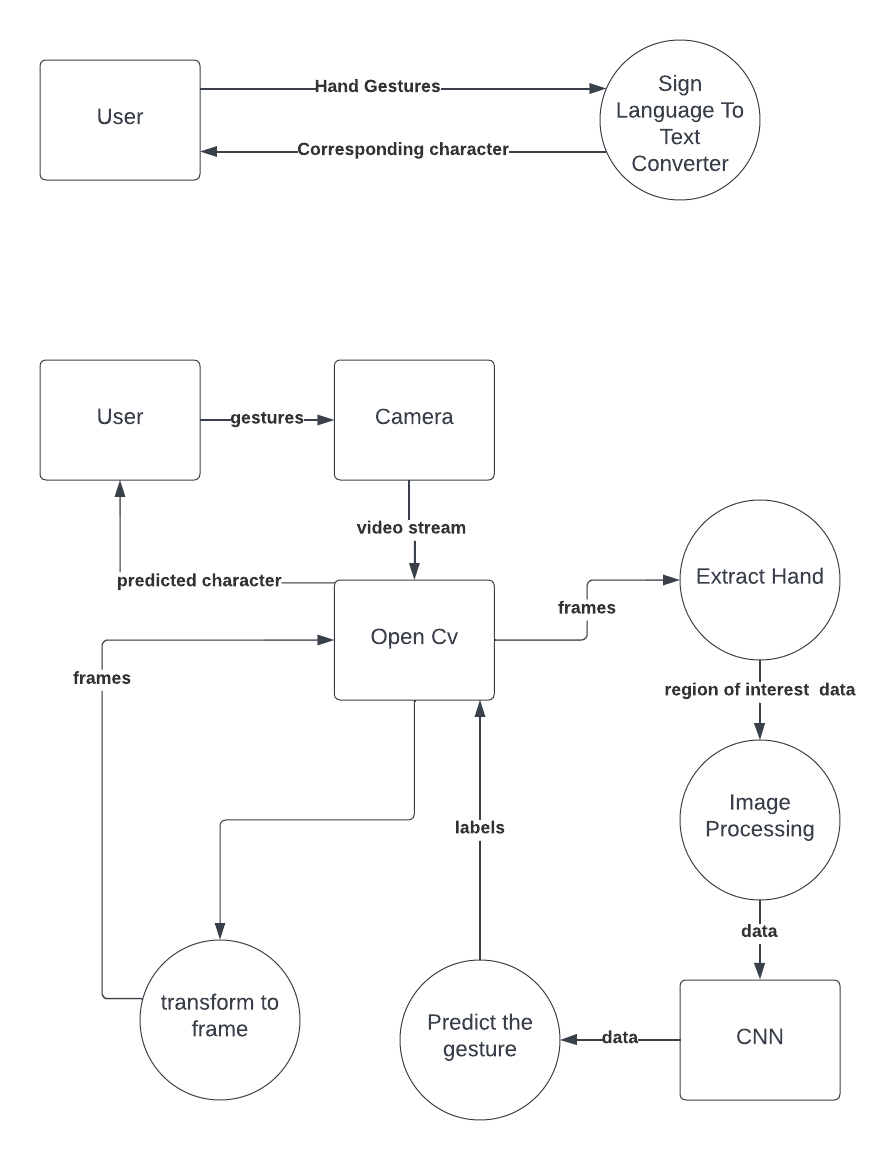
The proposed solution is a **vision-based desktop application** that captures hand gestures using a standard webcam, processes the image to extract hand landmarks, classifies the gesture using a trained CNN model, and then outputs the corresponding text and speech.  
The solution requires only a basic computer and webcam, making it affordable and accessible to everyone.

Main features include:

* Real-time hand gesture recognition.
* Robust performance across various lighting and background conditions.
* Conversion of recognized gestures into both text and speech outputs.
* User-friendly interface for smooth interaction.



**DFD Level – 0 Diagram**

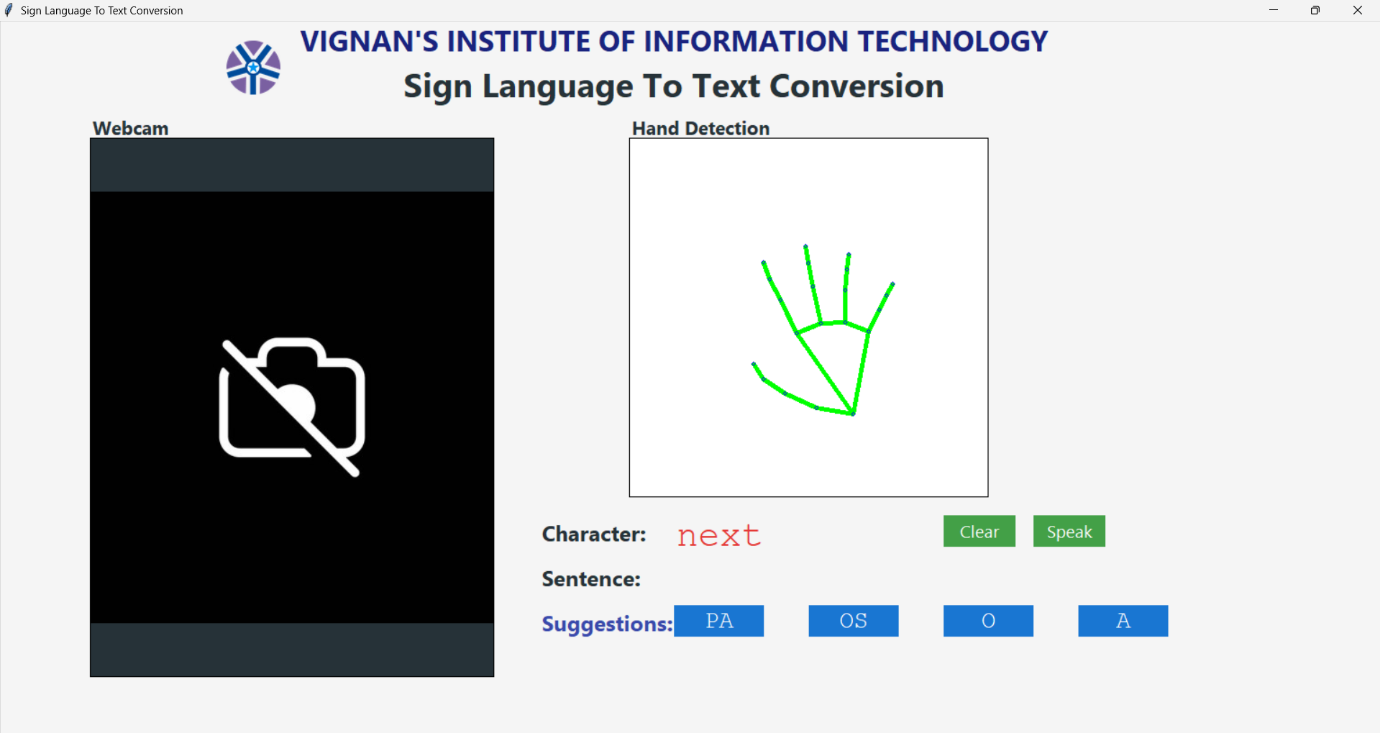


**DFD Level – 1 Diagram**

**PROTOTYPE**

**Prototype Development Steps:**

1. **Data Collection**:  
   Captured images of ASL gestures for each alphabet (A-Z) using webcam and Mediapipe.
2. **Preprocessing**:  
   Applied grayscale conversion, Gaussian blur, thresholding, and hand landmark drawing on a plain background to standardize input images.
3. **Model Training**:  
   Developed a CNN model using Keras and TensorFlow to classify gestures into respective alphabets.
4. **GUI Interface**:  
   Created a basic Tkinter GUI to display recognized alphabets and allow users to interact easily.
5. **Speech Integration**:  
   Integrated Pyttsx3 library to convert recognized text into audible speech output.



**TESTING**

Testing was carried out in multiple phases:

* **Unit Testing**:  
  Verified that each module (image capture, preprocessing, model prediction, text display, speech generation) worked independently.
* **Integration Testing**:  
  Tested the system as a whole by connecting all modules and ensuring proper communication between them.
* **User Testing**:  
  Conducted practical tests by asking users to perform random ASL alphabets in different lighting and background settings.  
  Observed system's real-time performance, recognition speed, and accuracy.
* **Result**:  
  Achieved an overall accuracy of 97% under normal conditions and up to 99% under clean backgrounds.

Challenges Found During Testing:

* Difficulty in recognition under very poor lighting.
* Confusion between very similar-looking gestures without clear separation.

Solutions:

* Encouraged use in well-lit environments when possible.
* Suggested clear finger positioning for better landmark detection.

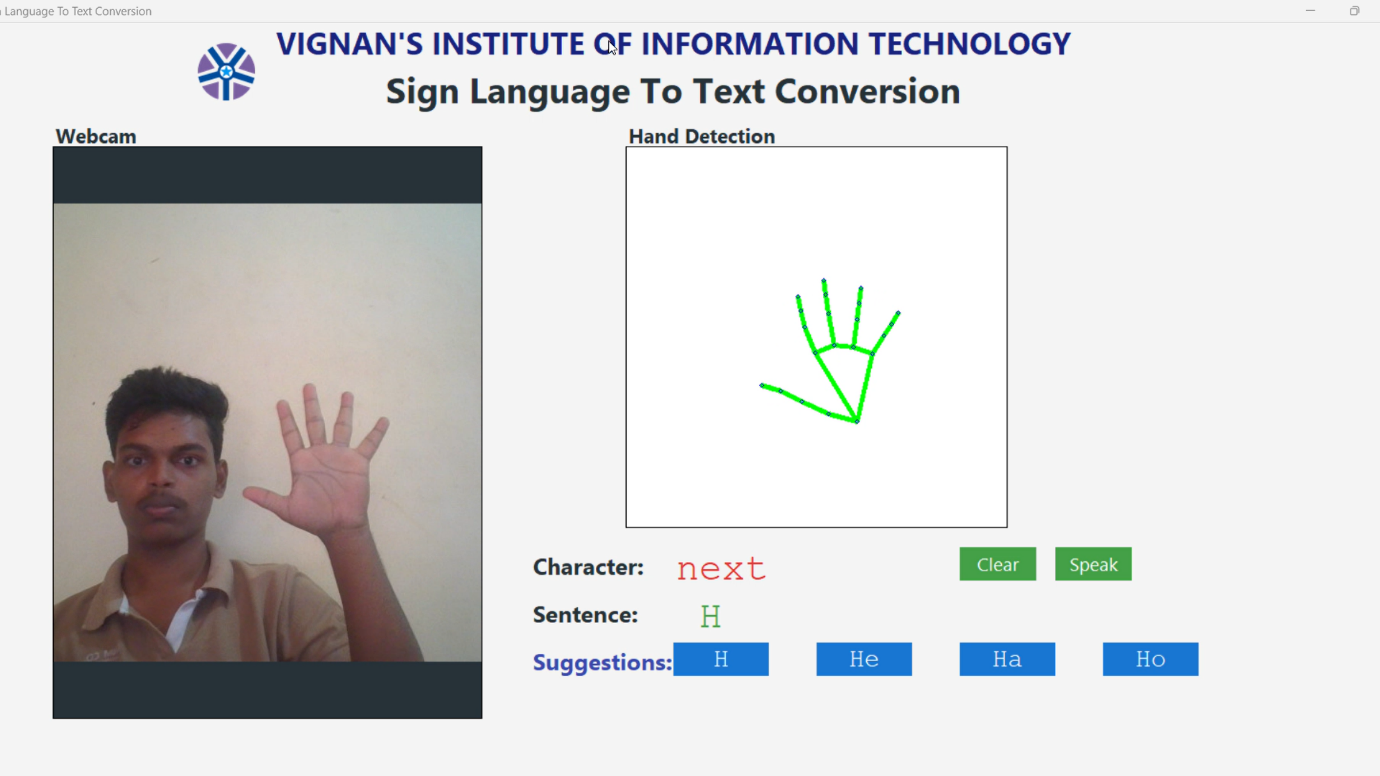
**IMPLEMENTATION**

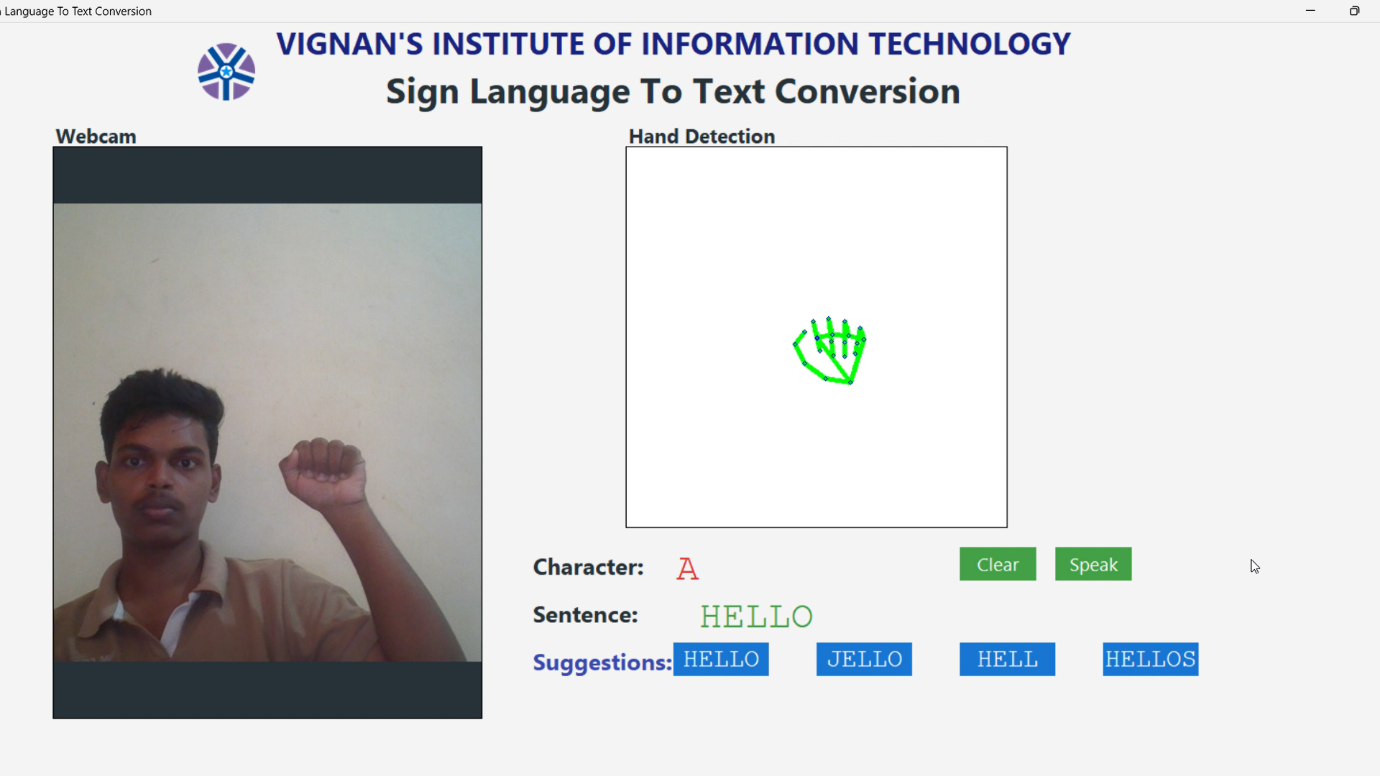
The final system was implemented as a standalone Python desktop application with the following workflow:

* Webcam captures real-time hand gesture.
* Mediapipe detects hand landmarks and extracts key points.
* Pre-processed skeleton image is passed to the trained CNN model.
* Model predicts the corresponding alphabet and outputs it as text on the GUI.
* Recognized text is converted into speech using Pyttsx3.
* Users can form words by sequentially showing gestures and use special gestures for space and next operations.

The system provides an effective, user-friendly, and low-cost solution to bridge the communication gap for deaf and mute individuals, fulfilling the core objective of the project.

IMAGES OF THE PROJECT





**GITHUB LINK OF THE PROJECT:**

<https://github.com/aashiksagar/Sign-Language-Recognition.git>