IJCRT.ORG ISSN : 2320-2882
INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)
An International Open Access, Peer-reviewed, Refereed Journal

# AI Based Sign Language Recognition and Language Translator

[1]Mohammed Tauqeer Ahmed, [2] Habeebur Rahman, [3] Abdul Rahman, [4] Keerti M
[1]Student, [2]Student, [3]Student,[4]Assistant Professor
[1]Department of Computer Science and Engineering,
Navodaya Institute of Technology,
Visvesvaraya Technological University (VTU),
Raichur, Karnataka, India

***Abstract***- In the modern digital era, effective communication remains a significant challenge for individuals with hearing and speech impairments. Traditional sign language communication often requires trained interpreters, which limits accessibility and real-time interaction with the general population. Although several assistive technologies have been developed, many existing solutions suffer from limited accuracy, lack real-time processing, dependency on internet connectivity, and poor user adaptability.

This paper proposes an **AI Based Sign Language Recognition and Language Translator** designed to bridge the communication gap between sign language users and non-signers. The proposed system utilizes **computer vision and deep learning techniques** to recognize hand gestures captured through a webcam and convert them into meaningful text and speech in real time. **MediaPipe Hand Tracking** is employed to extract hand landmark features, which are classified using a custom-trained neural network model. To improve reliability, a confidence-based filtering mechanism is implemented to reduce misclassification of gestures. Additionally, the system integrates an **offline speech-to-text module** and a **multilingual translation component**, enabling seamless bidirectional communication across different languages. A **Streamlit-based interactive user interface** provides real-time visualization, gesture training support, and smooth user interaction. The proposed solution offers a lightweight, scalable, and practical approach for assistive communication, accessibility technologies, and inclusive educational environments.

**Index Terms -** Artificial Intelligence, Sign Language Recognition, Hand Gesture Recognition, Deep Learning, Computer Vision, Assistive Technology, Accessibility Systems.

## I. INTRODUCTION

Communication is a fundamental human necessity; however, individuals with hearing and speech impairments often face significant barriers in daily interactions with the general population. Sign language serves as a primary mode of communication for the deaf and hard-of-hearing community, but effective communication typically requires trained interpreters, who are not always available in real-time situations. This limitation creates challenges in essential domains such as education, healthcare, workplaces, and public services. As a result, there is a growing demand for intelligent and accessible systems that can automatically translate sign language into understandable text or speech.

Traditional sign language interpretation systems rely on manual translation or hardware-based solutions such as sensor gloves, which are often expensive, intrusive, and impractical for widespread use. Vision-based approaches have gained attention as they offer a non-invasive alternative; however, earlier methods struggled with low accuracy, sensitivity to lighting conditions, and limited adaptability to different users and gestures. With recent advancements in **artificial intelligence, computer vision, and deep learning**, it has become possible to process visual data in real time and achieve reliable gesture recognition using standard camera devices.

This paper presents an **AI Based Sign Language Recognition and Language Translator** that leverages hand landmark detection and deep learning models to accurately interpret sign language gestures and convert them into meaningful text and speech. The proposed system integrates real-time gesture recognition with offline speech-to-text and multilingual translation capabilities, providing a comprehensive and accessible communication solution. By utilizing lightweight models and an interactive user interface, the system aims to enhance inclusivity, improve accessibility, and support real-world deployment in assistive communication technologies.

## II. LITERATURE REVIEW

Previous studies on sign language recognition have explored **sensor-based** and **vision-based** approaches. Sensor-based systems using data gloves and motion sensors provided accurate gesture capture but were costly, intrusive, and unsuitable for widespread use. This led to increased interest in vision-based systems that rely on camera input and conventional computer vision techniques combined with machine learning classifiers such as Support Vector Machines, Decision Trees, and Random Forests. Although these methods achieved moderate success, they were sensitive to lighting conditions, background noise, and user variations.

Recent advancements in **deep learning** have significantly improved sign language recognition performance. Convolutional Neural Networks (CNNs) and hybrid CNN–LSTM architectures have been widely used for automatic feature extraction and temporal modelling of gestures, achieving higher accuracy compared to traditional models. However, image-based deep learning approaches often require high computational resources and large datasets, which limits real-time applicability.

To address these challenges, recent research has adopted **landmark-based hand tracking methods**, such as MediaPipe, which represent gestures using skeletal key points rather than raw images. Landmark-based learning improves robustness, reduces computational complexity, and enables real-time performance on low-resource devices. Motivated by these findings, the proposed system utilizes hand landmark extraction with a neural network classifier and incorporates confidence-based filtering to enhance prediction stability and reliability in practical assistive communication environments.

## III. PROPOSED SYSTEM

The proposed **AI Based Sign Language Recognition and Language Translator** is designed to enable seamless communication between individuals with hearing or speech impairments and the general population. The system adopts a **vision-based approach** using a standard webcam to capture hand gestures in real time, eliminating the need for specialized hardware or wearable sensors.

The captured video frames are processed using **MediaPipe Hand Tracking** to extract 21 three-dimensional hand landmarks, which accurately represent finger and hand movements. These landmarks are converted into numerical feature vectors and passed to a **custom-trained deep learning neural network** for gesture classification. To enhance reliability and prevent incorrect predictions, a **confidence-threshold mechanism** is employed to label uncertain gestures as “UNKNOWN,” thereby improving system stability in real-world environments.

Once a gesture is recognized, the corresponding output is displayed as text and converted into speech for effective communication. In addition, the system incorporates an **offline speech-to-text module** that enables conversion of spoken language into text without internet dependency, along with a **multilingual translation module** to support communication across different languages. A **Streamlit-based interactive user interface** provides real-time visualization, smooth user interaction, and support for user-defined gesture training.

Overall, the proposed system is lightweight, scalable, and capable of real-time performance on low-computation devices, making it suitable for assistive communication, accessibility tools, and inclusive educational applications.

## IV. SYSTEM ARCHITECTURE AND IMPLEMENTATION

The system architecture is designed as a **modular, real-time processing pipeline** that integrates computer vision, deep learning, speech processing, and natural language translation to enable inclusive communication. The architecture consists of the following key components:

1. Input Layer

The system accepts real-time input from standard hardware devices:

- **Camera Module:** Captures live video frames containing hand gestures.
- **Microphone Module:** Captures spoken input for speech-to-text translation.
  This hardware-independent design ensures ease of deployment on common computing devices.

2. Preprocessing and Hand Tracking Module

Captured video frames are preprocessed and passed to the **hand tracking module**, which detects and tracks the user's hand in real time.

- **Hand Landmark Extraction:** 21 three-dimensional hand landmarks (x, y, z coordinates) are identified.
- **Feature Vector Formation:** The extracted landmarks are flattened into numerical feature vectors for efficient processing.

3. Gesture Recognition Module

The processed feature vectors are input to a **custom-trained deep learning neural network**:

- Classifies predefined sign language gestures.
- A **confidence-threshold mechanism** filters low-confidence predictions and labels them as UNKNOWN to prevent misclassification and improve reliability.

4. Output Generation Module

Once a gesture is correctly identified:

- The recognized gesture is converted into **text output**.
- Text is optionally converted into **speech output** using text-to-speech synthesis for natural interaction.

5. Speech-to-Text and Translation Module

To enable bidirectional communication:

- **Offline Speech-to-Text Engine:** Converts spoken language into text without internet dependency.
- **Language Translation Engine:** Translates text into multiple languages, supporting cross-lingual communication.

6. User Interface Layer

A **Streamlit-based interactive interface** manages user interaction:

- Displays real-time camera feed and recognition results.
- Allows switching between gesture translation and speech translation modes.
- Supports user-defined gesture training and system feedback.

7. System Integration and Control

All modules are seamlessly integrated to ensure:

- Low latency and real-time response.
- Smooth data flow between vision, learning, and language components.
- Scalability and adaptability for future enhancements.

The proposed system is implemented using a modular and real-time architecture to ensure efficient sign language recognition and translation. A webcam captures live video frames, which are processed using

MediaPipe to extract 21 hand landmarks representing gesture movements. These landmarks are transformed into numerical feature vectors and classified using a custom-trained deep learning model. A confidence-threshold mechanism is applied to filter uncertain predictions and improve reliability. The recognized gestures are converted into readable text and synthesized into speech for effective communication. Additionally, an offline speech-to-text module and multilingual translation component are integrated to support bidirectional and cross-language interaction through an intuitive Streamlit-based user interface.

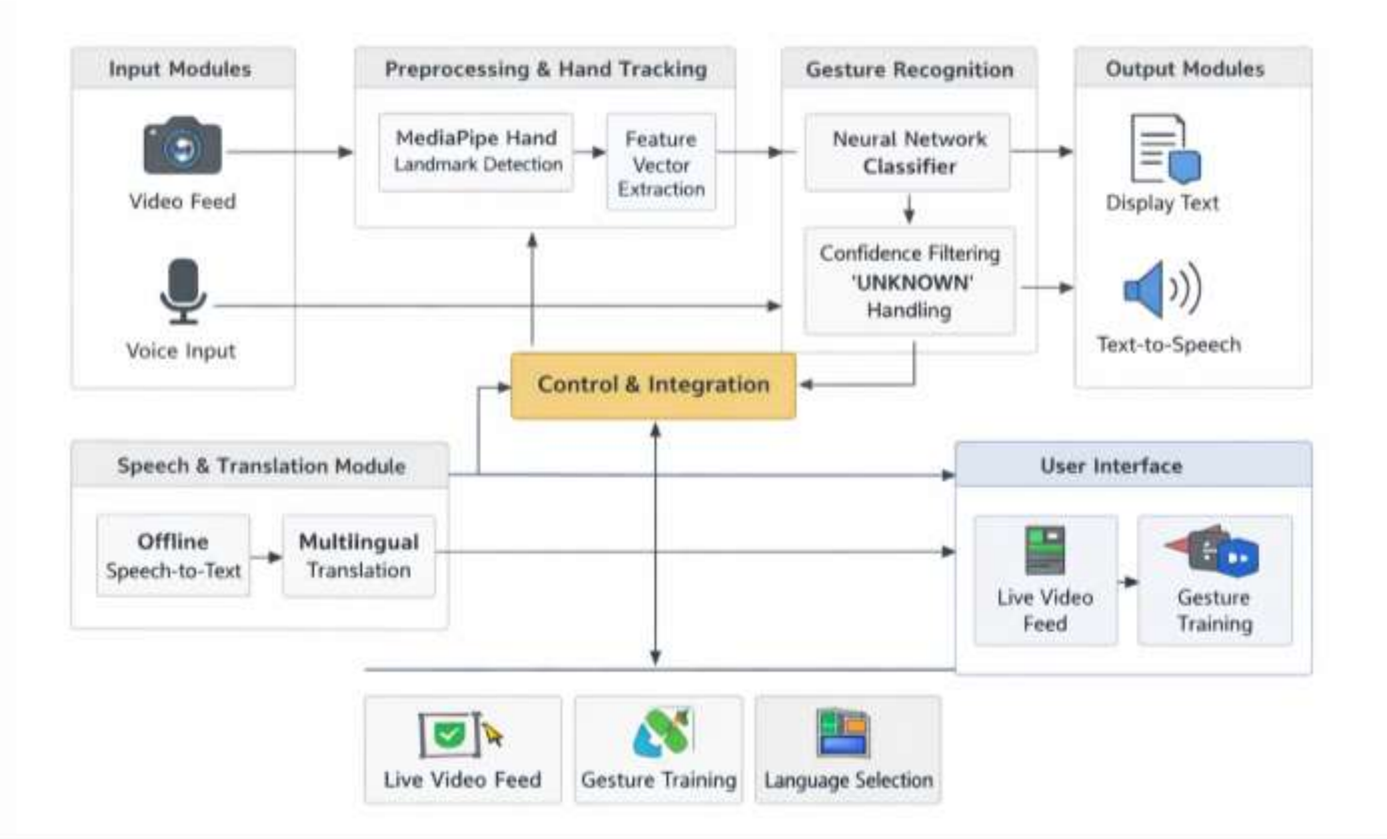


**Figure 1. System Architecture of AI Based Sign Language Recognition and Language Translator.**

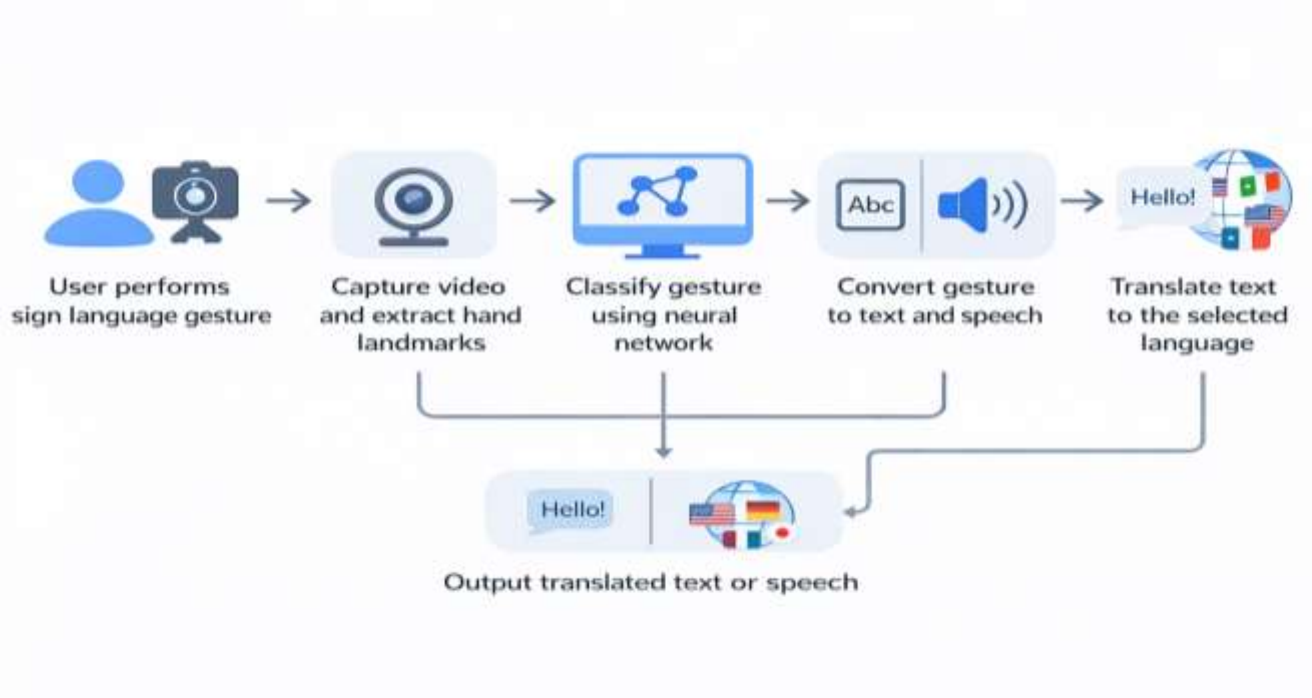


**Figure 2. Implementation Diagram of the proposed system.**

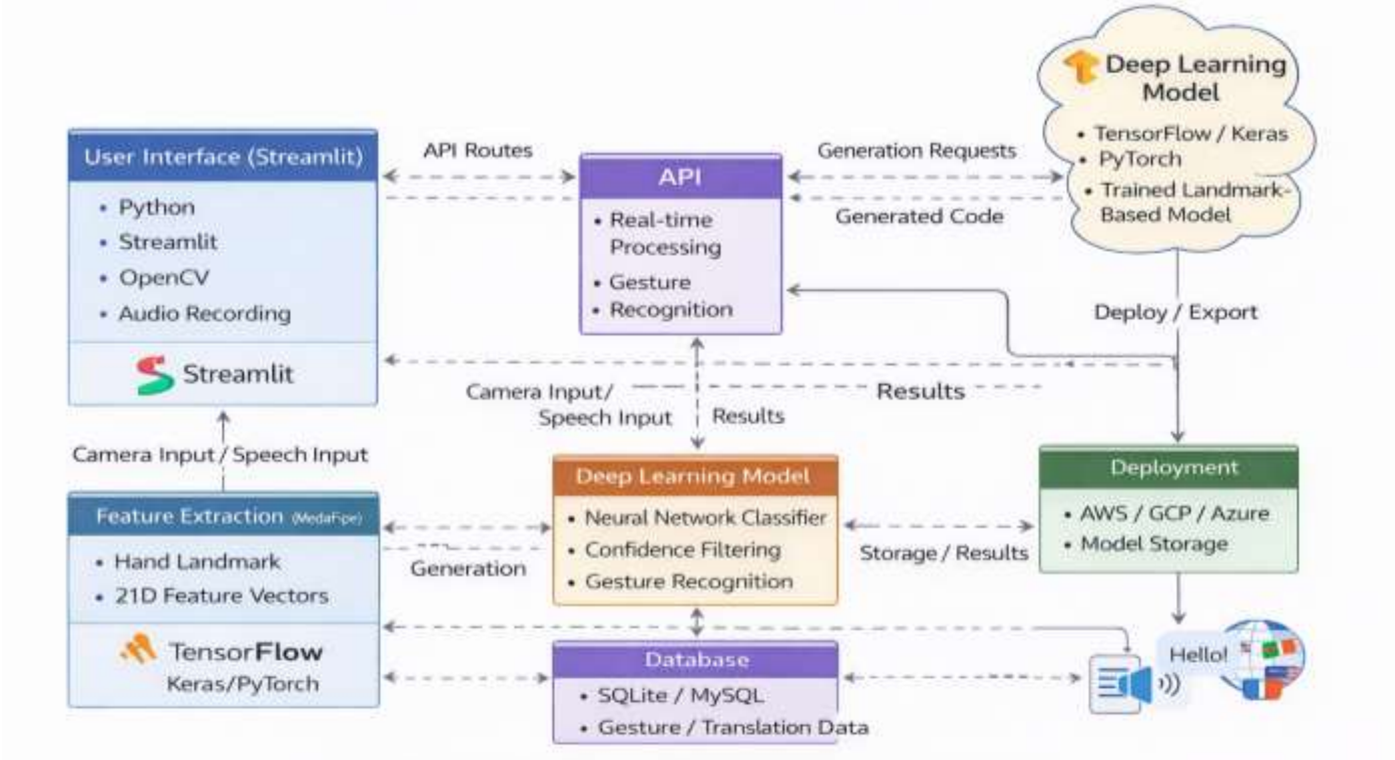


**Figure 3. Entity Relationship Diagram of the proposed system.**

## V. RESULT AND OUTPUT ANALYSIS

The performance of the proposed **AI Based Sign Language Recognition and Language Translator** was evaluated through real-time testing using predefined and user-trained gesture datasets. The deep learning–based gesture recognition model demonstrated **high recognition accuracy and stable real-time performance**, successfully identifying sign language gestures under varying lighting conditions and user hand orientations. The confidence-based filtering mechanism effectively reduced misclassification by labelling low-confidence predictions as *UNKNOWN*, thereby improving reliability and user trust.

The system generates clear outputs in the form of **recognized text and corresponding speech**, enabling smooth communication between sign language users and non-signers. Additionally, the integrated offline speech-to-text and multilingual translation modules produced accurate transcriptions and translations, supporting bidirectional and cross-language communication. Experimental observations indicate that consistent hand positioning and sufficient gesture samples improve recognition accuracy, while rapid or ambiguous gestures may result in filtered outputs. Overall, the results confirm that the proposed system is effective, lightweight, and suitable for real-time assistive communication applications.

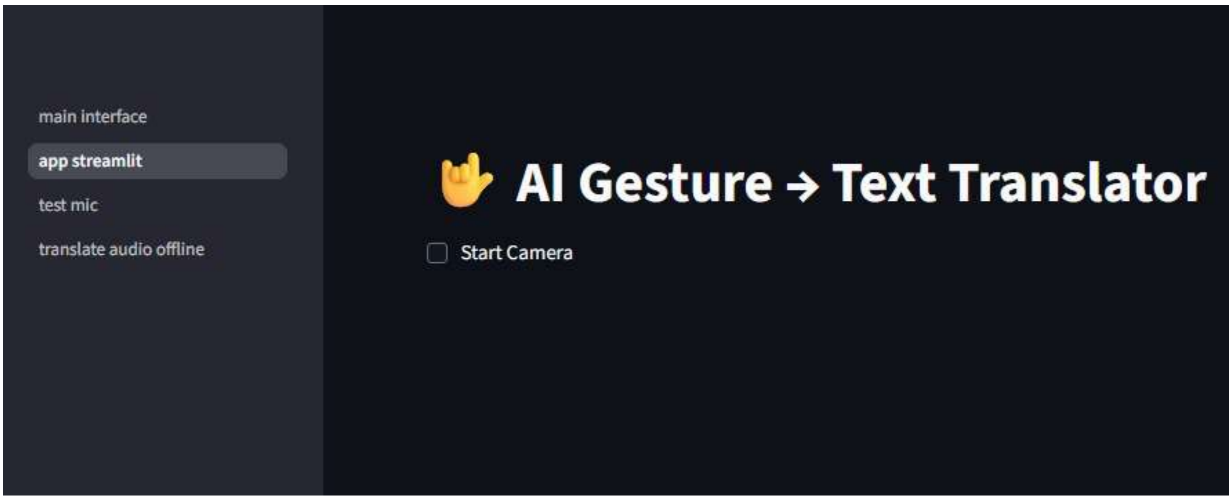


**Figure 4. Generated Output Console.**

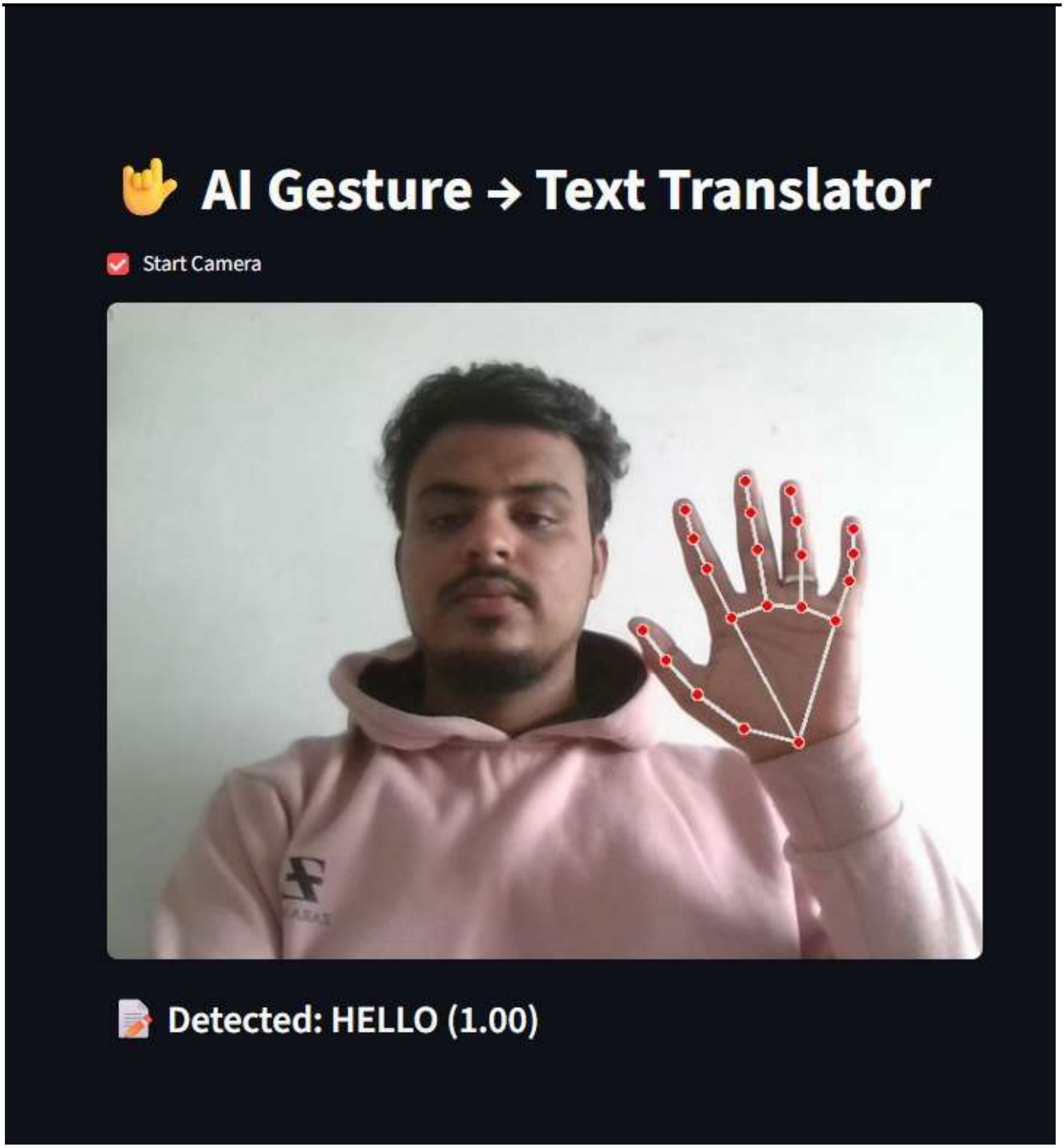


**Figure 5. Generated Output Console (Gesture Detection).**


**Figure 6. Generated Output Console (Speech to Text).**

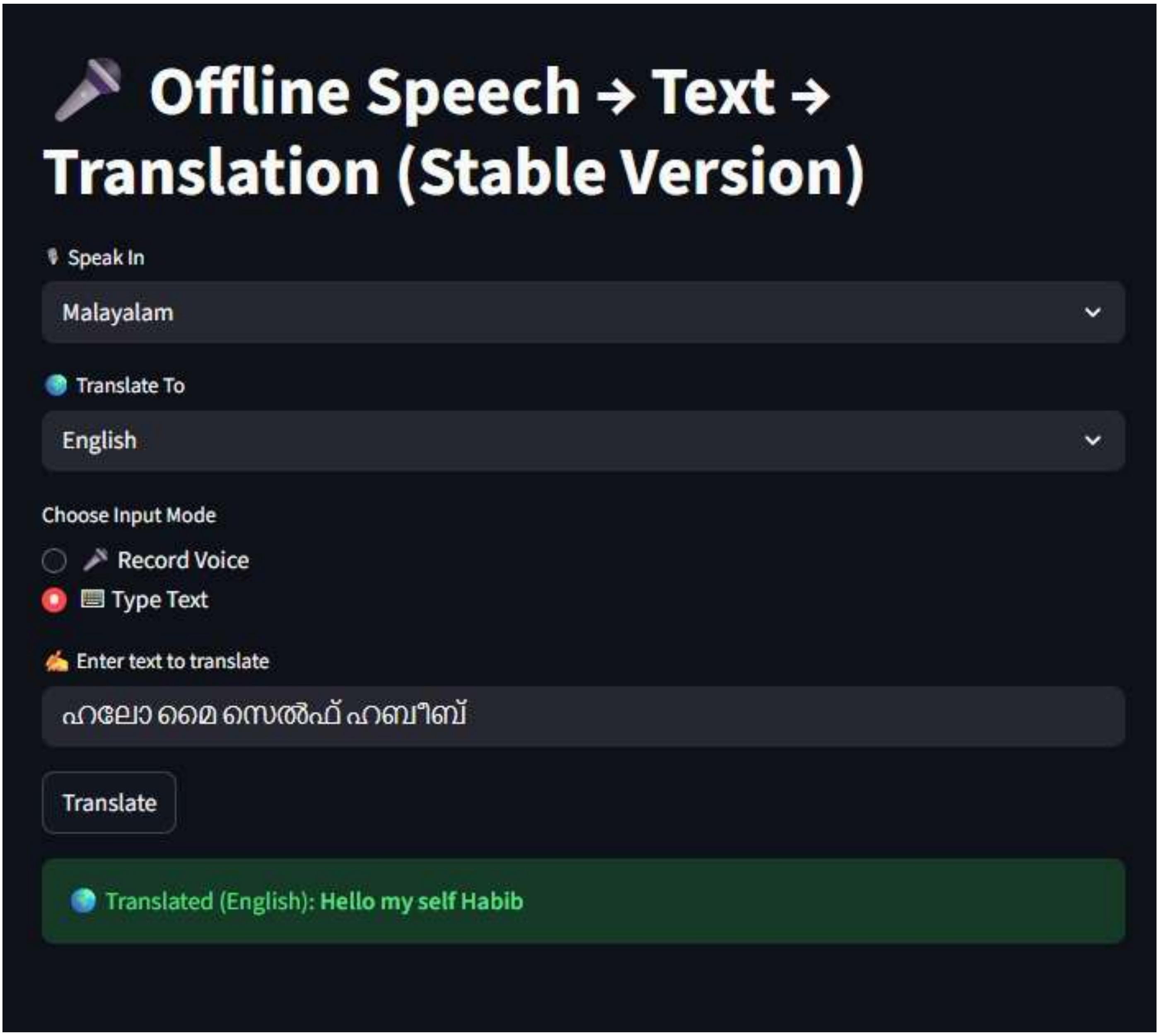


**Figure 7. Generated Output Console (Speech to Text Verified).**

## VI. CONCLUSION AND FUTURE SCOPE

This paper presented an **AI Based Sign Language Recognition and Language Translator** designed to enhance communication between individuals with hearing or speech impairments and the general population. The proposed system effectively utilizes computer vision, hand landmark extraction, and deep learning techniques to recognize sign language gestures and convert them into meaningful text and speech in real time. The integration of offline speech-to-text and multilingual translation modules further improves system accessibility and supports bidirectional communication across diverse linguistic backgrounds.

The adoption of a lightweight and interactive web-based interface enhances usability and makes the system suitable for practical deployment. Future work may focus on expanding the gesture vocabulary, supporting two-handed and continuous sign language recognition, improving model accuracy using larger and more diverse datasets, and incorporating explainable AI techniques to provide better insight into gesture classification decisions. Additionally, deploying the system on mobile or embedded platforms can further extend its real-world applicability in assistive communication and accessibility solutions.

## ACKNOWLEDGEMENT

The authors would like to express their sincere gratitude to the Department of Computer Science and Engineering, Navodaya Institute of Technology, Raichur, for providing the required infrastructure, laboratory facilities, and computational resources necessary for carrying out this research work. The authors are deeply thankful to their project guide for consistent guidance, technical support, and constructive feedback throughout all phases of the project, from problem identification to final implementation. The valuable suggestions and encouragement received from the faculty members of the department greatly contributed to improving the quality, clarity, and effectiveness of this research. The institutional support provided during the course of this work is gratefully acknowledged.

## REFERENCES

[1] T. Starner and A. Pentland, "Real-Time American Sign Language Recognition from Video Using Hidden Markov Models," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 12, pp. 1371–1375, 1998.

[2] S. Mitra and T. Acharya, "Gesture Recognition: A Survey," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 37, no. 3, pp. 311–324, 2007.

[3] C. Keskin, F. Kirac, Y. E. Kara, and L. Akarun, "Hand Pose Estimation and Hand Shape Classification Using Multi-Layered Randomized Decision Forests," *European Conference on Computer Vision (ECCV)*, pp. 852–863, 2012.

[4] Z. Cao, G. Hidalgo, T. Simon, S. E. Wei, and Y. Sheikh, "OpenPose: Realtime Multi-Person 2D Pose Estimation Using Part Affinity Fields," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 43, no. 1, pp. 172–186, 2021.

[5] F. Zhang, L. Shen, and Y. Wang, "Hand Gesture Recognition Using Deep Learning," *International Journal of Signal Processing, Image Processing and Pattern Recognition*, vol. 8, no. 11, pp. 195–204, 2015.

[6] Google, "MediaPipe: A Framework for Building Perception Pipelines," 2020. [Online]. Available: https://mediapipe.dev