University of Engineering and Technology Lahore
Department of Computer Science
Muhammad Tayyab 2022-CS-133 mu.tayyab001@gmail.com

Real-Time Sign Language to Text and Speech

This project introduces an innovative real-time solution for translating sign language into text and speech, aimed at bridging the communication gap for individuals with hearing and speech impairments. The system employs cutting-edge deep learning models, including CNN, RNN, and LSTM, alongside MediaPipe for hand landmark detection and gesture recognition. It provides a robust, low-latency, and accessible solution, suitable for consumer-grade hardware.

Research Objective

To design a real-time sign language translation system using CNN, RNN, LSTM, and MediaPipe, capable of interpreting gestures into text and speech. The objective is to enhance accessibility and inclusivity for deaf and mute individuals, offering seamless integration into daily communication.

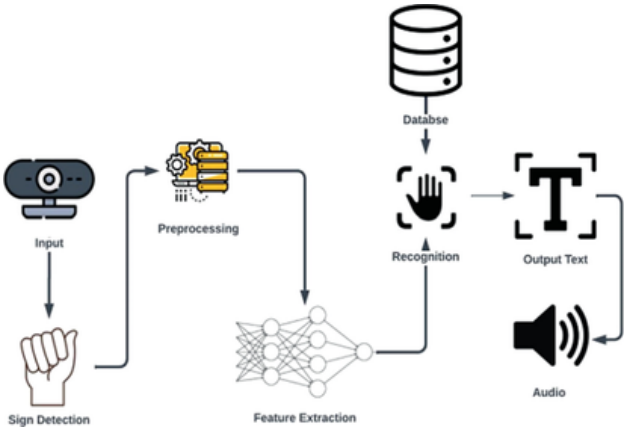
Key Concepts

- **Sign Language Recognition:** Using CNN, RNN, and LSTM for accurate gesture detection and translation.
- **MediaPipe Integration:** Real-time hand tracking for precise landmark detection.
- **Accessibility Enhancement:** Bridging communication gaps with low-latency responses.
- **Multi-Model Evaluation:** Incorporating machine learning models like Random Forest, Gradient Boosting, and SVM for gesture classification.

Methodology

System Architecture:
The system integrates:

- **MediaPipe:** For real-time hand tracking and feature extraction.
- **Deep Learning Models:**
  CNN: For spatial feature extraction.
  RNN/LSTM: For sequential gesture recognition.
- **Machine Learning:** Evaluation with Random Forest, Gradient Boosting, SVM, and others for comparative performance analysis.
- **Data Pipeline:** Preprocessing diverse datasets for adaptability to varied gestures and real-world conditions.

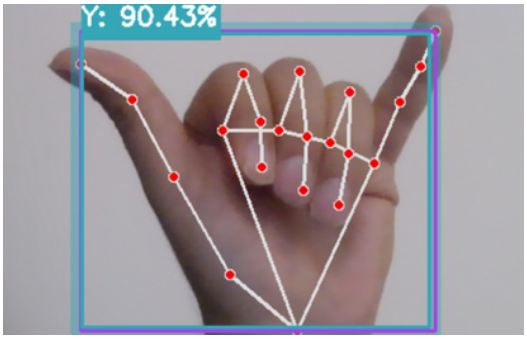


Conclusion

This project presents a novel real-time sign language translation system that bridges communication gaps for deaf and mute individuals. By leveraging MediaPipe for gesture detection alongside advanced deep learning models such as CNN, RNN, and LSTM, the system ensures accurate and low-latency translations into text and speech. The comparative evaluation of multiple machine learning models, including Random Forest, demonstrates robust performance, achieving up to 98.7% accuracy. This work advances assistive technologies, fostering inclusivity and accessibility, and sets the stage for future innovations, such as dynamic gesture recognition and support for diverse sign languages.

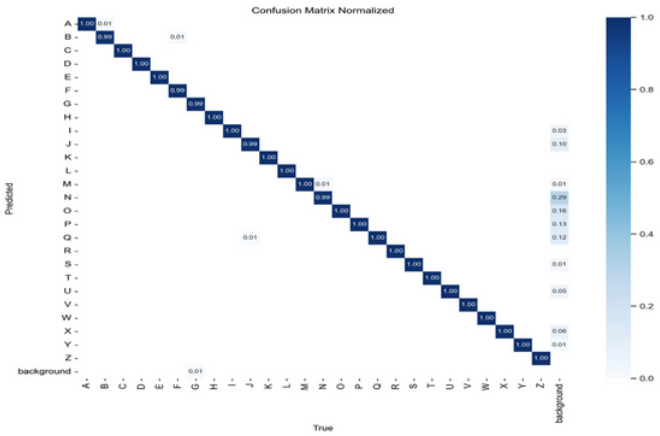
Hand Landmark Detection

The system employs MediaPipe for detecting and annotating hand landmarks, enabling real-time gesture recognition with high accuracy and reliability.



Confusion Matrix for Model Performance Evaluation

The confusion matrix provides a detailed analysis of the model's classification performance for sign language alphabets. Each row represents the true labels, while each column corresponds to the predicted labels.



Results

- **Deep Learning Performance:** CNN, RNN, and LSTM achieved seamless gesture recognition with minimal latency.
- **Random Forest Accuracy:** Demonstrated 99.7% accuracy, the highest among all tested models.
- **Scalability:** System adaptability to diverse datasets and environments enhances its usability.