Classification Of Sign Language Characters by Applying a Deep Convolutional Neural Network

Dr. Shah S.N[1], Dr. Nazirkar S.B[2], Gaikwad Tejaswini[3], Khumbhar Sakshi[4], Babar Sakshi[5]
[1]*Co-Ordinator, Department of Computer Engineering SPCOET, Someshwarnagar*
[2]*Guide, Department of Computer Engineering SPCOET, Someshwarnagar*
[3,4,5]*Students, Department of Computer Engineering SPCOET, Someshwarnagar*

***Abstract*—Communication barriers between the deaf and hard-of-hearing community and the general population remain a significant challenge in everyday interactions. Traditional methods of sign language interpretation often require human interpreters, which may not always be available, leading to delays and limited accessibility. To address this issue, the Sign Language Detection System proposes an intelligent, real-time solution that utilizes computer vision and artificial intelligence to translate sign language gestures into text and speech. The system captures live video input through a webcam and processes it using advanced hand tracking techniques powered by MediaPipe to extract precise hand landmarks. These landmarks are then analyzed using a dual mode classification approach that combines a deep learning model with a rule-based heuristic engine. This hybrid mechanism ensures high reliability and continuous operation, even in cases where the trained model is unavailable or produces low-confidence predictions. To enhance accuracy and stability, the system incorporates prediction smoothing techniques and a gesture hold-time mechanism, reducing false detections caused by rapid or transitional hand movements. The recognized gestures are converted into readable text and simultaneously transformed into audible speech using a multi-threaded text-to-speech engine, enabling seamless real-time feedback without interrupting system performance. The application is deployed using a Flask-based backend and is containerized with Docker to ensure scalability, portability, and ease of deployment across different platforms. Additionally, the system supports a flexible training pipeline that allows users to collect custom gesture datasets and improve model performance over time. The proposed system provides an efficient, scalable, and user-friendly solution for real-time sign language interpretation. It aims to bridge communication gaps, promote inclusivity, and enhance accessibility by offering a reliable and automated platform for gesture-based interaction.**

I. INTRODUCTION

Communication is a fundamental aspect of human interaction, yet individuals from the deaf and hard-of-hearing community often face significant challenges in expressing themselves and understanding others. Sign language serves as a primary mode of communication for such individuals, but the lack of widespread understanding among the general population creates a communication gap. Traditional solutions rely on human interpreters, which may not always be available, making real-time communication difficult. This highlights the need for an intelligent, automated, and accessible system that can translate sign language into text and speech effectively.

The Sign Language Detection System is developed to provide real-time gesture recognition using computer vision and artificial intelligence. The system captures hand gestures through a webcam and processes them using advanced hand tracking techniques. It then translates these gestures into meaningful text and speech outputs. By integrating machine learning models with rulebased logic, the system ensures accurate and reliable gesture recognition, even in varying conditions. The web-based implementation of the system ensures ease of access, platform independence, and a user-friendly interface. Through efficient processing and intelligent analysis, this project aims to bridge the communication gap and enhance accessibility for the deaf and hard-of-hearing community.
The Sign Language Detection System focuses on providing real-time, automated gesture-to-text and speech translation.

II. LITERATURE REVIEW

Sign language recognition has been an active area of research in the field of computer vision and human-computer interaction. Traditionally, communication for the deaf and hard-of-hearing community relied on human interpreters, which limited accessibility and real-time interaction.

In paper [1]: A Review on Smart Gloves to Convert Sign to Speech for Mute Community (Author: Khan Sohelrana, Syed Faiyaz Ahmed, Shaik Sameer) - Description: Focuses on an embedded system using Flex sensor and NodeMCU to reduce communication barrier.

[2]: A Gesture-to-Emotional Speech Conversion by Combining Gesture Recognition and Facial Expression Recognition (Author: Nan Song, Hongwu Yang) - Description: Proposes a facial expression integrated sign language to emotional speech conversion using DNN, SVM, and HMM-based synthesizer.

[3]: Hidden Markov model-based Sign Language to Speech Conversion System in TAMIL (Author: Aiswarya V, Naren Raju N, Johanan Joy Singh S, Nagarajan T, Vijayalakshmi P) - Description: Uses an accelerometer-gyroscope sensor-based hand gesture recognition module and HMM-based text-to-speech synthesizer.

[4]: A Translator for American Sign Language to Text and Speech (Author: Vi N.T. Truong, Chuan-Kai Yang, QuocViet Tran) - Description: Detects static hand signs of alphabets in ASL using AdaBoost and Haar-like classifiers. In paper

[5]: Design and Implementation of a Sign-toSpeech/Text System for Deaf and Dumb People (Author: DalalAbdulla, Shahrazad Abdulla and Rameesa Manaf, Anwar H. Jarndal) - Description: Presents an approach for designing and implementing a smart glove for deaf and dumb people.

[6]: Justin K. Chen, Debabrata Sengupta, Rukmani Ravi Sundaram," Sign Language Gesture Recognition with Unsupervised Feature Learning"

III. SYSTEM ARCHITECTURE

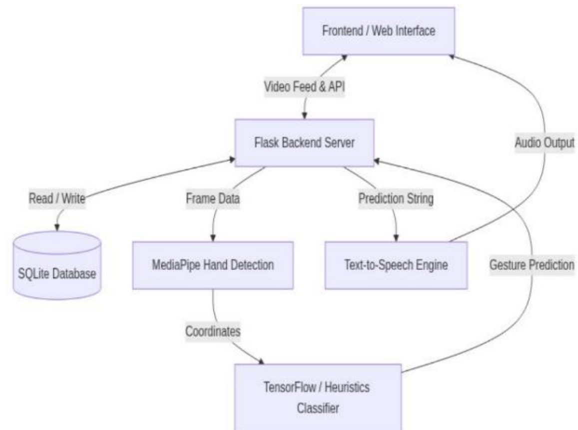


Fig. System Architecture

1. Presentation Layer (Frontend)
Displays live video feed
Shows detected gesture text output

2. Backend (Flask)
Handles video streaming and processing
Manages API requests

3. Computer Vision Module
MediaPipe for hand detection
Extracts 21 landmark points

4. Classification Engine
ML Model (TensorFlow/Keras)
Heuristic fallback logic

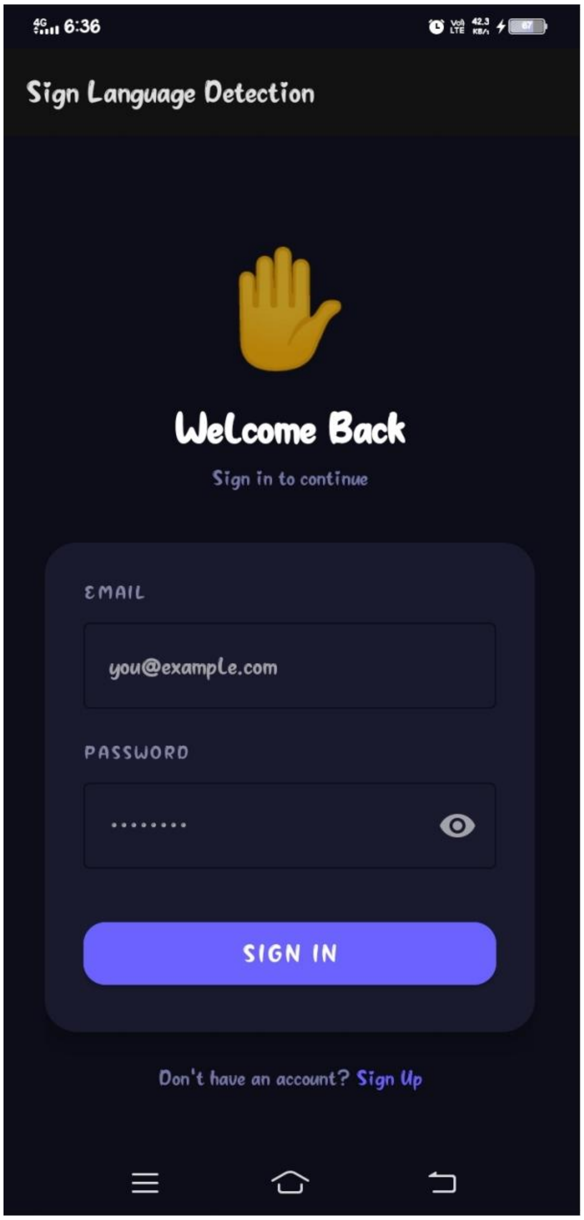
5. TTS Module
Converts text into speech

6. Deployment
Local / Docker-based deployment

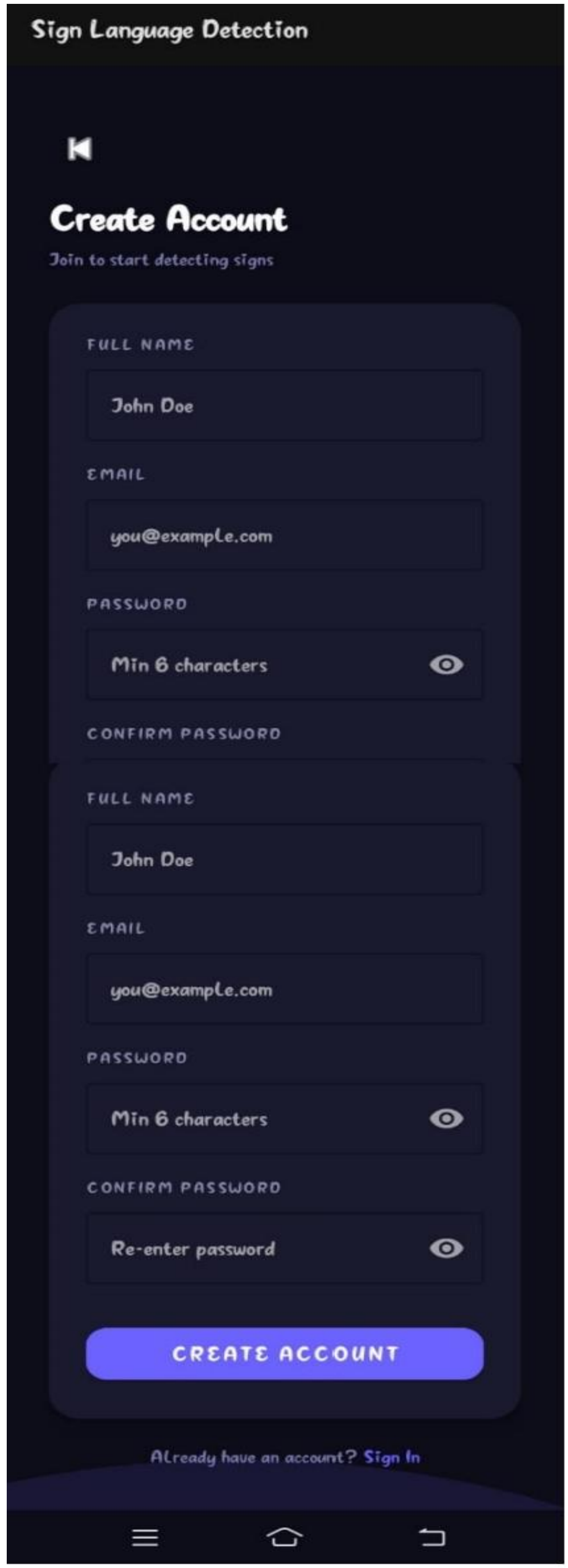
IV. METHODOLOGY

SLR can be done using one hand, two hands but the hearing-impaired people also use their facial expressions to represent their feelings. Therefore, to understand the information, hearing and speech impaired people use signs which may include detection of face and hands which acts as an input for tracking in recognition of sign language symbols. The proposed system works on the critical aspects of sign language symbols such as position of hand on particular area of a face by taking the image in YCbCr format using Image processing toolbox of MATLAB for performing skin color detection to make a template and then dividing the template to compute quad values which acts as threshold value to do matching and doing recognition of sign language symbols.

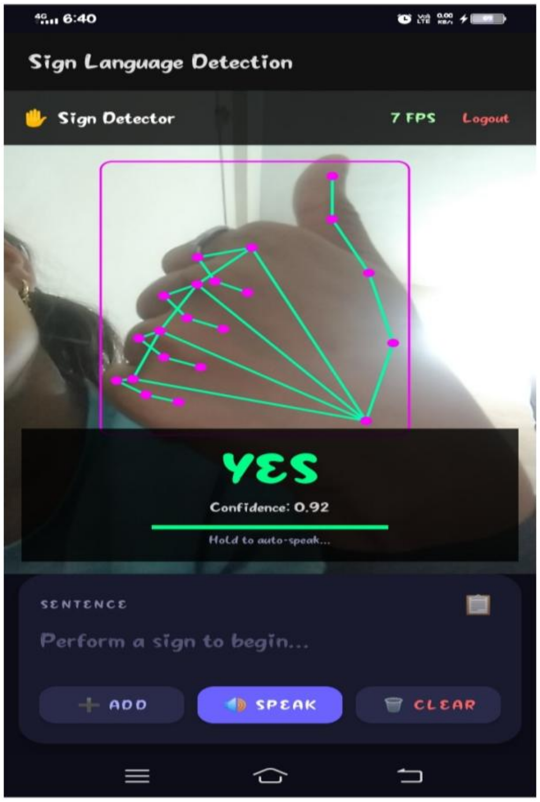
V. RESULT AND DISCUSSION



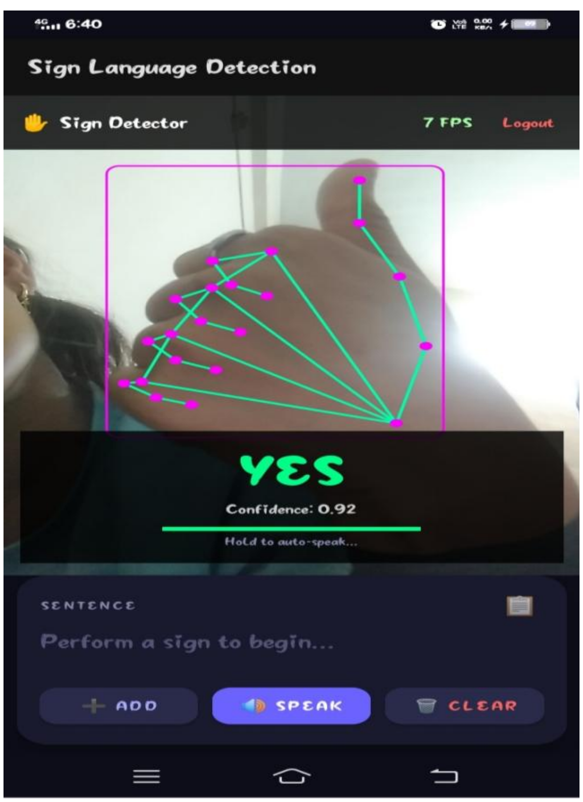
1. Home Page (System Interface)



2. Account creation



3. Hand Detection using MediaPipe



4. Text Output Display

VI. CONCLUSION

The Sign Language Detection System is developed as an intelligent real-time application that utilizes computer vision and artificial intelligence to bridge the communication gap between the deaf and hard-of-hearing community and the general population. The system captures live video input, detects hand gestures using MediaPipe, and translates them into meaningful text and speech outputs.

A key strength of the system lies in its real-time performance and stability. The implementation of prediction smoothing techniques and gesture hold-time significantly improves accuracy and reduces flickering in outputs. Additionally, the use of a fallback mechanism enhances system robustness and ensures uninterrupted functionality. The system can be further

enhanced with advanced models and expanded datasets to support more complex sign language interactions.

VII. FUTURE WORK

This paper investigated methods to model hand movement information in a language independent manner using hand movement sub units obtained through HMMs. Our investigations showed that there is a performance gap when modeling hand movement information in a language independent manner and in a language dependent manner. However, this gap is significantly reduced when combined with hand shape information and yields competitive systems. These findings are promising and they pave the path for development of sign language processing systems by sharing multiple sign language resources. Our future work will build upon these finding to address resource-constraint issues in sign language processing such as, developing systems with reduced number of signers and examples. In addition, we will also investigate whether such a multilingual approach can be applied for sign language assessment.

REFERENCES

[1] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2016.

[2] M. Tan and Q. Le, “EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks,” in *Proc. Int. Conf. Mach. Learn. (ICML)*, 2019.

[3] Google, “MediaPipe Hands: On-device Real-time Hand Tracking,” 2020. [Online]. Available: MediaPipe Official Website

[4] S. Molchanov, S. Gupta, K. Kim, and J. Kautz, “Hand Gesture Recognition with 3D Convolutional Neural Networks,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2015.

[5] R. Rastgoo, K. Kiani, and S. Escalera, “Video-based Isolated Hand Sign Language Recognition Using Deep Learning,” *IEEE Access*, vol. 8, pp. 72693–72706, 2020.

[6] O. Koller, H. Ney, and R. Bowden, “Deep Learning of Mouth Shapes for Sign Language,” in *Proc. IEEE Int. Conf. Comput. Vis. Workshops (ICCVW)*, 2015.

[7] D. Neverova, C. Wolf, G. Taylor, and F. Nebout, “Multi-scale Deep Learning for Gesture Detection and Localization,” in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2014.

[8] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “ImageNet Classification with Deep Convolutional Neural Networks,” in *Advances in Neural Information Processing Systems (NeurIPS)*, 2012.

[9] TensorFlow Team, “TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems,” 2015. [Online]. Available: TensorFlow Official Website

[10] OpenCV, “Open-Source Computer Vision Library,” 2023. [Online]. Available: OpenCV Official Website