A DEEP LEARNING FRAMEWORK FOR SIGN LANGUAGE RECOGNITION TO ASSIST SPEECH AND HEARING-IMPAIRED INDIVIDUALS

Project Reference No.: *49S_BE_0958*

College : *BGS College of Engineering and Technology, Bengaluru*
Branch : *Department of Computer Science and Design*
Guide(s) : *Mrs. Sanjitha S Laad*
Student(s) : *Ms. Shwetha S Nayak*
Ms. Nisarga V
Ms. Honalu Shekar M
Ms. Ayesha N

Keywords:

Indian Sign Language (ISL), Sign Language Recognition (SLR), MediaPipe Holistic, Deep Learning, Real Time Translation.

Introduction:

Effective communication is essential for human interaction, yet Deaf and hard-of-hearing individuals face challenges when communicating with non-signers. Indian Sign Language (ISL) is widely used, but its limited understanding creates barriers in daily life. Sign Language Recognition (SLR) systems help bridge this gap by converting gestures into text or speech. This project develops a real-time ISL recognition system using MediaPipe for hand landmark extraction. A CNN + LSTM model is used, where CNN captures spatial features and LSTM models temporal sequences. The model is evaluated using accuracy, precision, recall, and F1-score to ensure reliability. The system provides multilingual text and Text-to-Speech output, enhancing accessibility and usability.

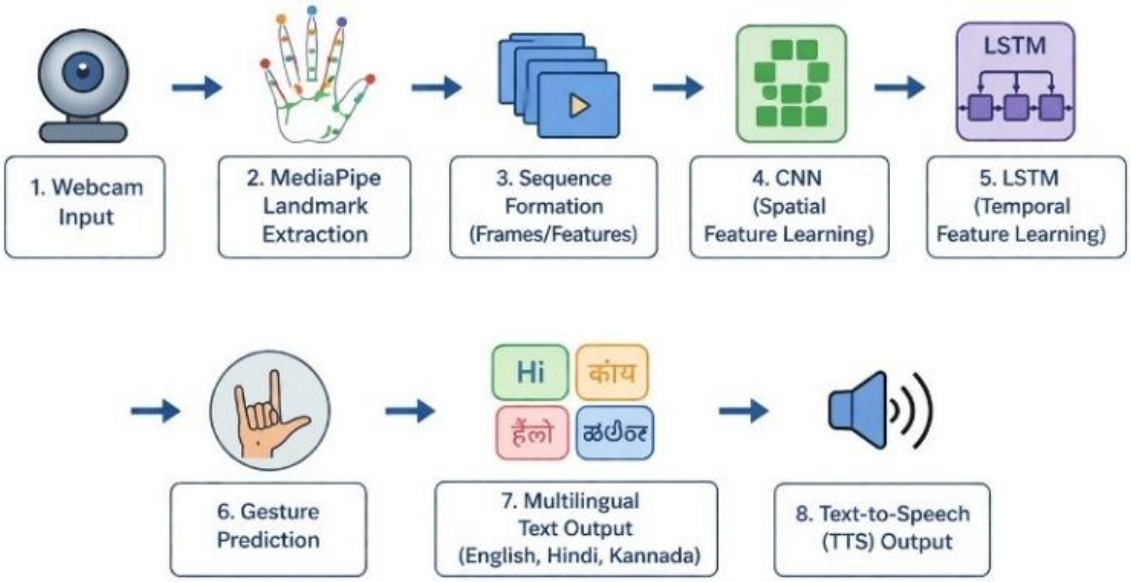

Figure 1: System Architecture

Objectives:

1. To develop a real-time Indian Sign Language (ISL) recognition system using MediaPipe Holistic for accurate hand landmark detection.
2. To create and preprocess a custom ISL dataset by collecting gesture videos, including datasets from Kaggle.
3. To design, train, and evaluate the deep learning model for efficient gesture classification.
4. To convert recognized gestures into multilingual text (English, Hindi, Kannada) for better accessibility.

Methodology:

A hybrid CNN + LSTM model is used to recognize both static and dynamic ISL gestures by capturing spatial and temporal features. A custom dataset is created using webcam recordings and Kaggle data, including variations in lighting and background. Preprocessing involves frame extraction, resizing, normalization, and cleaning to improve data quality. MediaPipe Holistic is used to extract 21 hand landmarks per frame, which are converted into sequential data for model input. The CNN learns spatial features such as hand shapes, while the LSTM captures temporal dependencies in gesture sequences. The model is trained and evaluated using accuracy, precision, recall, and F1-score. The model achieves an accuracy of 80%–98%. For real-time deployment, webcam input is processed through a Streamlit interface, and predicted gestures are converted into multilingual text and speech using TTS.

Result and Conclusion:

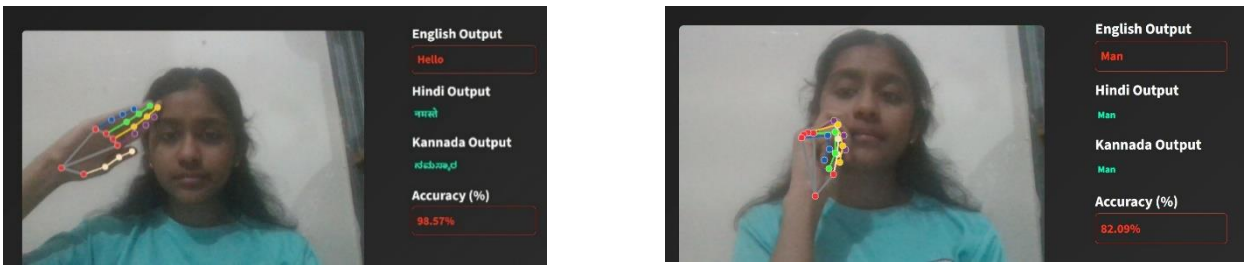

Fig 2: Sign Language Recognition

In conclusion, the project was successfully developed as a real-time Indian Sign Language (ISL) recognition system using a hybrid CNN + LSTM architecture. The model captures spatial and temporal features for accurate recognition of both static and dynamic gestures. MediaPipe enables efficient hand landmark extraction, improving system performance and robustness. The system is evaluated using accuracy, precision, recall, and F1-score to ensure reliability. It operates efficiently on standard hardware with real-time processing capability. Multilingual output in English, Hindi, and Kannada, along with Text-to-Speech (TTS), enhances accessibility and usability. Overall, the system provides a practical and scalable solution for assisting deaf and hard-of-hearing individuals.

Project Outcomes and Industry Relevance:

The proposed system is expected to achieve accurate real-time recognition of Indian Sign Language gestures using a CNN + LSTM architecture. It will provide multilingual text output along with Text-to-Speech for enhanced accessibility. The system will operate efficiently on standard hardware with low latency and robust performance under varying conditions.
The project is highly relevant to industry and society as it supports communication for deaf and hard-of-hearing individuals. It can be applied in healthcare, customer service, education, and public service sectors for real-time interaction. Organizations such as hospitals, banks, airports, and government service centers can integrate this system. It also has applications in AI-based accessibility tools and human-computer interaction systems.

Working Model vs Simulation/Study:

This project involves the development of a working model rather than a simulation or theoretical study. The system performs real-time ISL recognition using webcam input, MediaPipe for hand landmark extraction, and a CNN + LSTM model for gesture classification. It generates multilingual text and speech output, demonstrating practical functionality and real-world applicability.

References:

1. Najib, F.M. “Sign language interpretation using machine learning and artificial intelligence”. Neural Comput & Applic 37, 841–857 (2025). https://doi.org/10.1007/s00521-024-10395-9.
2. Pawar et al.,(2023). “AI Powered Sign Language Translator”. INTERANTIONAL JOURNAL OF SCIENTIFIC RESEARCH IN ENGINEERING AND MANAGEMENT. 07. 1-11. 10.55041/IJSREM26681.
3. Hu et al., (2023). “STFE-Net: A Spatial-Temporal Feature Extraction Network for Continuous Sign Language Translation”. IEEE Access. PP. 1-1. 10.1109/ACCESS.2023.3234743.
4. Lau, W.K., Chalupny, J., Grote, K. et al. “How sign language expertise can influence the effects of face masks on non-linguistic characteristics”. Cogn. Research 7, 53 (2022). https://doi.org/10.1186/s41235-022-00405-6.

Future Scope:

The future scope of this project includes:

1. **Continuous Sign Language Recognition (CSLR):** Extend the system to recognize continuous sign and full sentences using advanced models such as BiLSTM.
2. **Enhanced Text-to-Speech (TTS):** Improve the audio module to generate more natural, expressive, and human-like speech output in multiple languages, including Hindi and Kannada.
3. **Vocabulary Expansion:** Expand the dataset and translation dictionary to support a larger set of ISL gestures, improving system accuracy, scalability, and real-world usability.