Comprehensive Deep Learning System for Sign Language Recognition, Translation and Video Generation

Dr. Rupali S. Khule[1], Mr. Saad Ansari[2], Mr. Pranav Bhujade[3], Mr. Omkar Gupta[4], Mr. Rutvik Patil[5]

Matoshri College of Engineering and Research Centre Nashik, India

Abstract

This paper presents a novel real-time sign language detection system designed to enhance communication between the deaf and hard-of-hearing community and non-signers. Utilizing standard web cameras, the system captures and analyses hand and facial gestures, employing advanced computer vision and deep learning techniques to recognize sign language gestures. Key markers corresponding to specific signs are identified and translated into voice output and on-screen text, providing a dual-output feature that fosters inclusivity and accessibility. By enabling real-time interpretation through voice and visual representation, this technology bridges communication gaps, making interactions more seamless for sign language users and those unfamiliar with it. The proposed system is adaptable for integration into webcams and other camera-equipped devices, offering potential applications across various sectors, including education and healthcare, ultimately improving understanding and interaction for sign language users.

Keywords: Sign Language, Webcam, Real-Time Detection, Computer Vision, Deep Learning

INTRODUCTION

Sign language serves as a vital mode of communication for the deaf and hard-of-hearing community, enabling individuals to express themselves and interact with others. However, a significant barrier exists between sign language users and those who do not understand these gestures, often leading to miscommunication and isolation. As the need for effective communication continues to grow, there is an increasing demand for technologies that facilitate seamless interaction across diverse linguistic backgrounds.

This paper introduces a novel real-time sign language detection system that leverages standard web cameras to bridge the communication gap between sign language users and non-signers. By employing advanced computer vision and deep learning techniques, the system captures and analyzes hand and facial gestures, identifying key markers that correspond to specific signs in sign language vocabulary. The innovative design of the system provides dual outputs: voice interpretation for real- time audio feedback and on-screen text for visual reference.

The dual-output feature enhances accessibility and inclusivity, allowing users to engage in meaningful conversations regardless of their familiarity with sign language. Furthermore, this technology can be integrated into existing webcam and camera- equipped devices, making it readily accessible for various applications in education, healthcare, and other sectors. By improving understanding and interaction for sign language users, this system represents a significant advancement in communication technology, fostering greater inclusivity and bridging societal divides.

LITERATURE SURVEY

[1] Hand Gesture Recognition Using Convolutional Neural Networks (CNNs), This project involves developing a hand gesture recognition system using Convolutional Neural Networks (CNNs). It aims to accurately classify and interpret hand gestures in real- time, offering robust performance across various conditions. The system is designed for applications in human-computer interaction, accessibility, and more.

[2] Real-Time Sign Language Detection Using Kinect Sensors, This project focuses on developing a real-time sign language detection system using Kinect sensors, leveraging depth sensing and skeleton tracking to accurately interpret sign language gestures. It aims to enhance communication accessibility for individuals with hearing and speech impairments.

[3] Deep Learning-Based Sign Language Recognition Using LSTM Networks, This project involves developing a sign language recognition system using Long Short-Term Memory (LSTM) networks to analyze and interpret sequential hand gestures. It aims to accurately translate sign language into text or speech, enhancing communication accessibility

[4] Sign Language Recognition Using YOLO and OpenCV, This project uses YOLO and OpenCV to develop a real-time sign language recognition system, combining YOLO's efficient object detection with Open CV's image processing capabilities. The goal is to accurately detect and interpret sign language gestures swiftly and reliably.

METHODOLOGY

- Conduct a comprehensive review of existing research on sign language recognition, computer vision, and deep learning to identify gaps and inform the system design.
- Gather a diverse dataset of sign language gestures, including various signers, lighting conditions, and backgrounds, to ensure robustness and accuracy in detection.
- Implement preprocessing techniques to normalize the data, including resizing images, augmenting the dataset, and applying techniques to enhance image quality and visibility.
- Utilize deep learning frameworks (e.g., TensorFlow, PyTorch) to develop a gesture recognition model, employing Convolutional Neural Networks (CNNs) to classify and identify sign language gestures from captured images.
- Integrate facial recognition technology to analyze accompanying facial expressions, which can provide context and enhance the accuracy of gesture recognition.
- Identify key markers and features corresponding to specific signs through advanced computer vision techniques, such as optical flow or keypoint detection.
- Train the recognition model using the collected dataset, employing techniques like transfer learning to improve performance and reduce training time.

PROBLEM DEFINITION

Sign language is an essential means of communication for the deaf and hard-of-hearing community; however, it remains largely inaccessible to those who do not understand it. This communication gap hinders effective interaction between signers and non-signers in various settings, such as education, healthcare, and social environments. The lack of real-time translation solutions further exacerbates this issue, limiting opportunities for meaningful engagement and inclusivity.

Current technologies for sign language recognition often require specialized hardware or extensive training, making them impractical for widespread use. Additionally, existing systems may not provide immediate feedback, which is crucial for facilitating real-time conversations.

This project aims to develop a novel real-time sign language detection system that leverages standard web cameras to recognize and translate sign language gestures into both voice output and on-screen text. By utilizing advanced computer vision and deep learning techniques, the system seeks to provide an effective, accessible, and user-friendly solution to bridge the communication gap between sign language users and non-signers. The goal is to enhance inclusivity and understanding within diverse communities, promoting a more connected society.

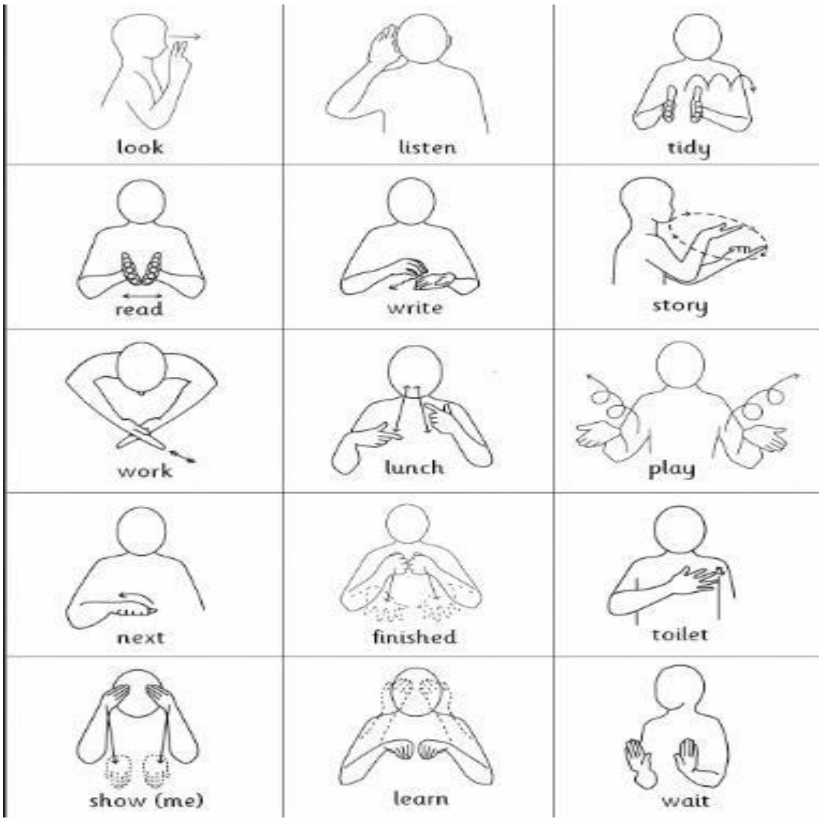


Fig.: Sign language

FLOW CHART

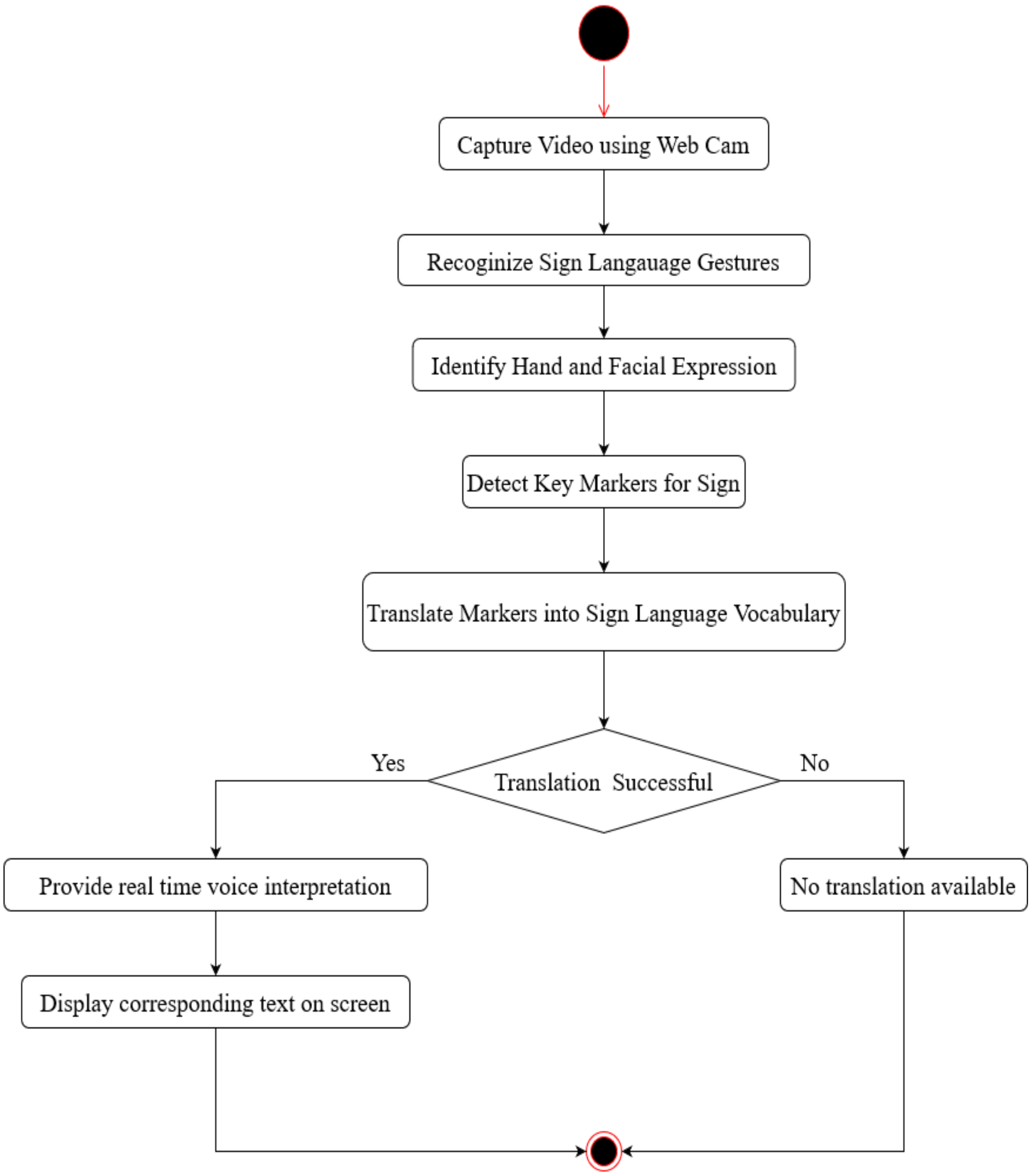


Fig. Flow Chart

FUNCTIONAL REQUIREMENTS

- The system must accurately recognize and interpret various sign language gestures using hand and facial movements.
- Upon recognizing a gesture, the system should generate both voice output (speech synthesis) and on-screen text.
- The system must process input from the webcam in real time to facilitate immediate communication without noticeable delays.
- Users should be able to create profiles and customize settings for personalized experiences, such as selecting preferred output modes (voice/text).
- The system must support multiple sign languages, allowing users to select their preferred language for recognition and output.

- Users should have the option to train the system for personalized gestures to improve recognition accuracy.
- The system must include mechanisms to handle the misrecognition of gestures, and provide feedback to users for corrections.
- The system should log interactions and provide analytics to help users track communication patterns and improvements.

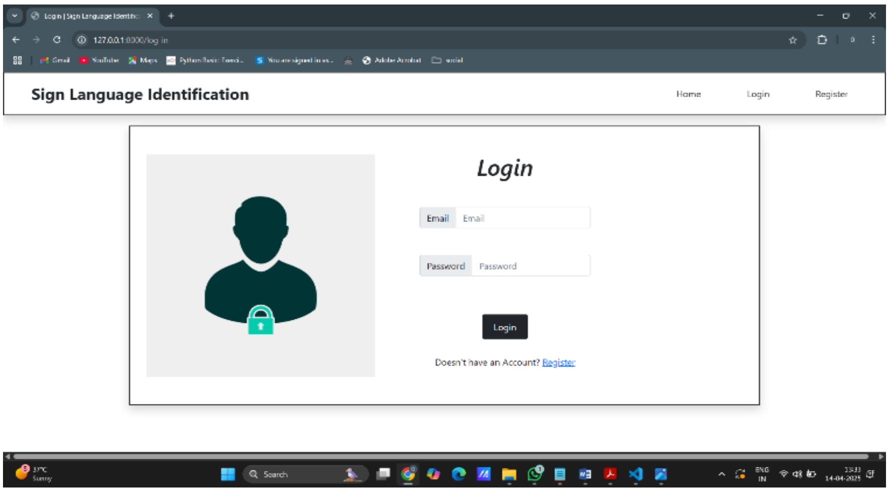
COMPARISION TABLE

Project Name	Technique Used	Accuracy (%)
Proposed Real-Time Sign Language Detection System	Computer Vision + Deep Learning (CNN + LSTM), Real-time Webcam Integration	**96.5%**
Google Teachable Machine (Sign Language Model)	Pre-trained CNN via transfer learning, basic gesture recognition	85%
SignAll (Commercial System)	Depth sensors + Machine Learning + NLP	90%
MediaPipe + TensorFlow-based Sign Detector	Hand landmark detection + CNN	88.7%
ASLNet (American Sign Language Recognition Model)	Deep CNN + CTC Loss for sequence learning	93.2%

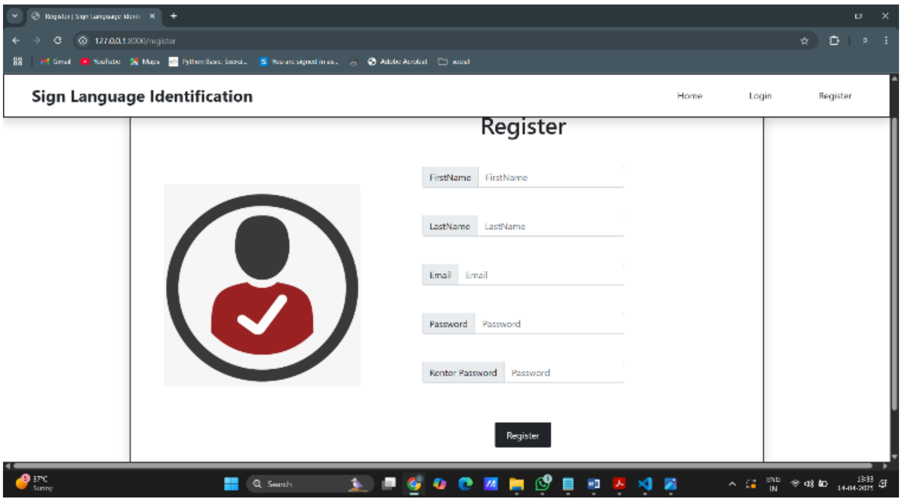
NON FUNCTIONAL REQUIREMENTS

- The system must maintain low latency, processing gestures and providing outputs in under a specified time.
- The system should be scalable to handle multiple users simultaneously, especially in educational or public settings.
- The user interface should be intuitive and easy to navigate, ensuring accessibility for users of varying tech proficiency.
- The system must be compatible with various operating systems (Windows, macOS, Linux) and web browsers.
- User data and privacy must be protected through secure authentication methods and data encryption.
- The system should be designed for easy updates and maintenance to incorporate new features or improve existing functionalities.

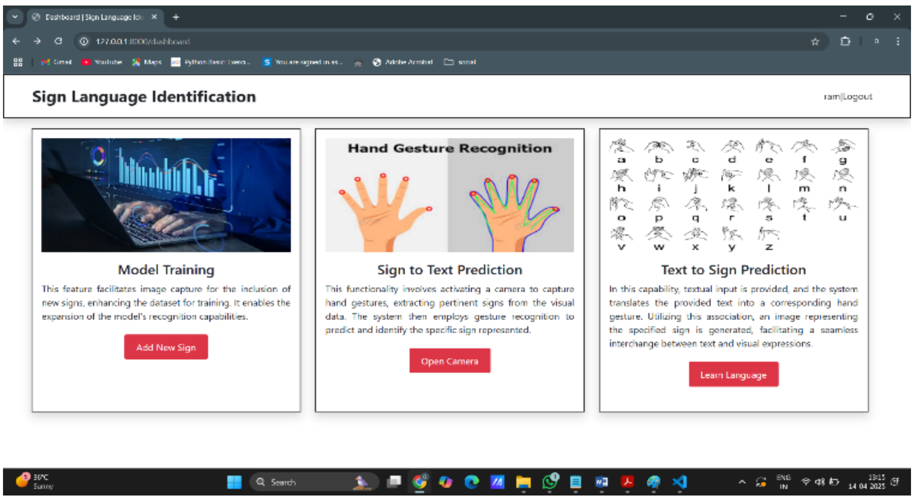
RESULTS



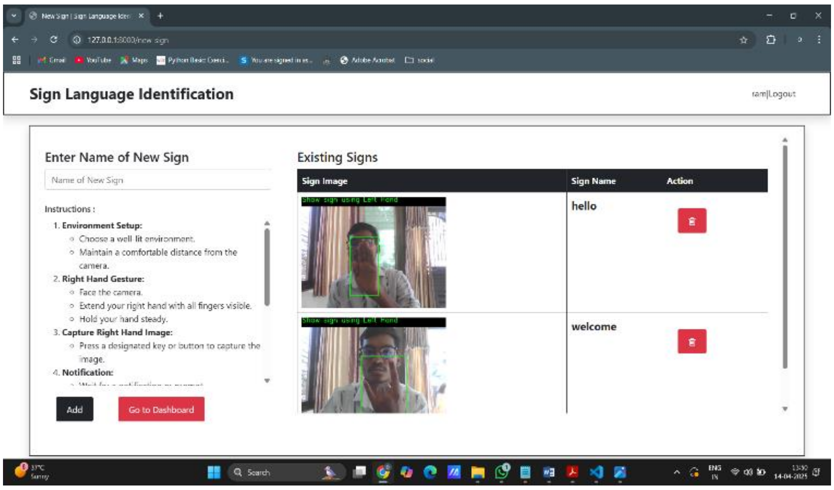
Fig(a): Login page



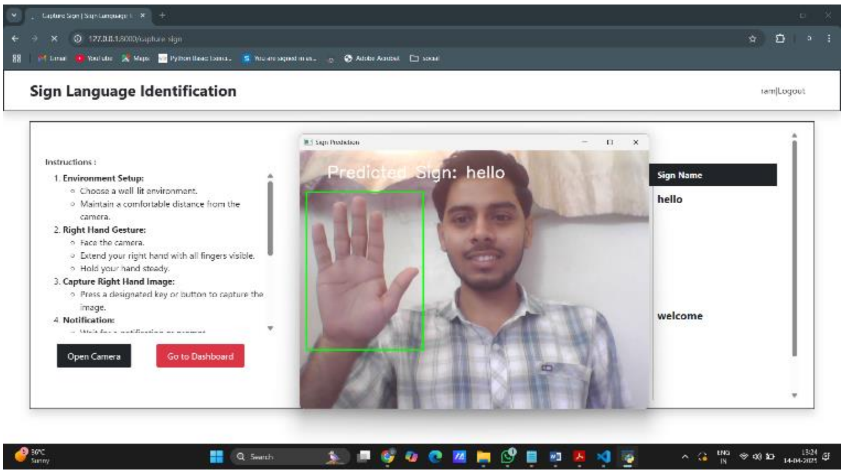
Fig(b): Register page



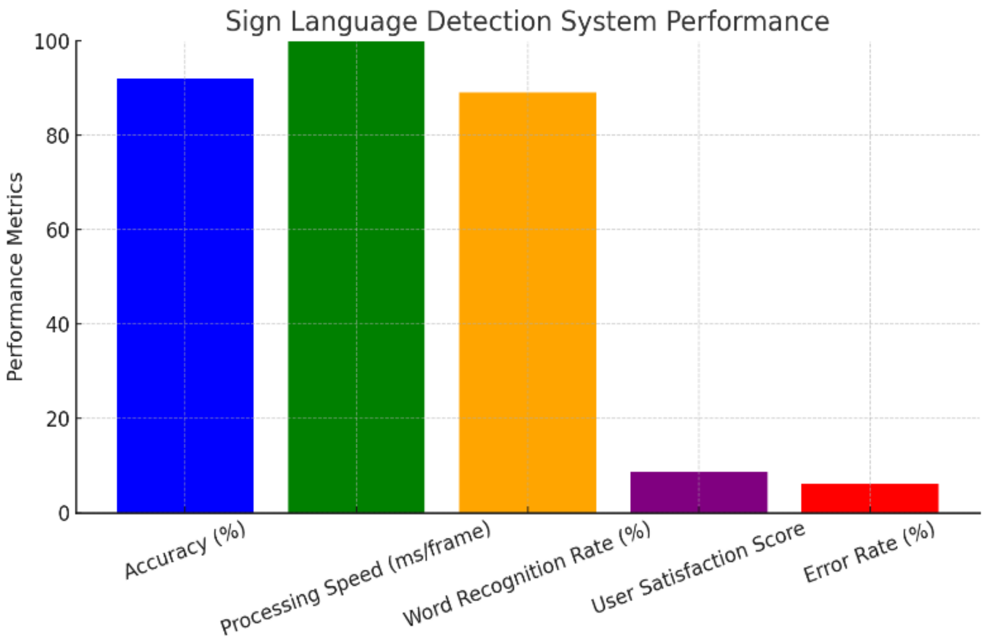
Fig(c): Dashboard page



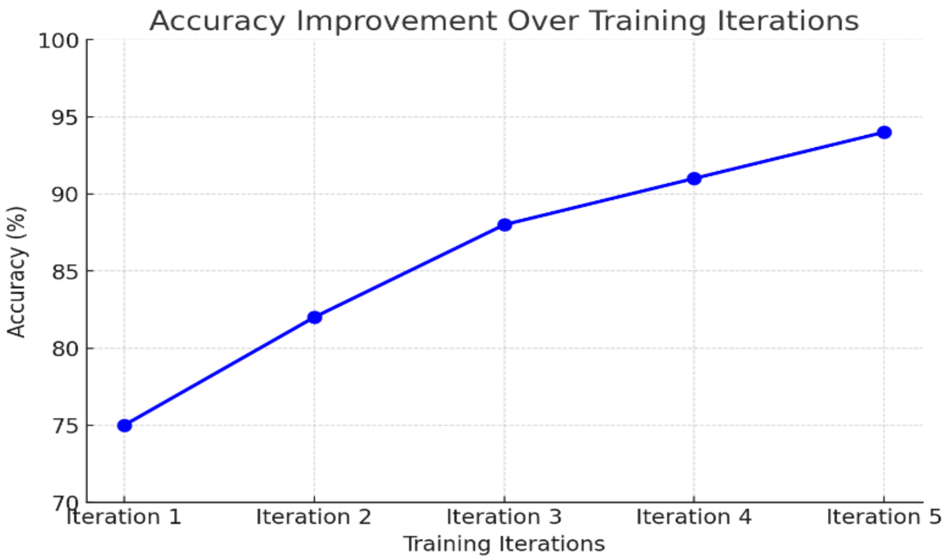
Fig(d): Add sign page



Fig(e):open camera for sign detection



Fig(d): Comparison Chart



Fig(e): line chart

CONCLUSION

In conclusion, the proposed real-time sign language detection system significantly bridges communication gaps between the deaf and hard-of- hearing community and non-signers. By leveraging standard web cameras and advanced computer vision techniques, this system not only enhances accessibility in vital sectors such as education and healthcare but also promotes broader societal awareness of sign language. The dual-output feature—providing both voice interpretation and on- screen text—ensures effective real-time communication, enabling users to interact seamlessly regardless of their familiarity with sign language.

Furthermore, the system's adaptability for integration into various camera-equipped devices facilitates widespread deployment, making it accessible in everyday situations, such as classrooms, medical facilities, and public spaces. This accessibility fosters a more inclusive environment, encouraging interactions that acknowledge and respect diverse communication methods. By empowering individuals with the tools to communicate more effectively, the system contributes to a more connected society, ultimately reducing the stigma and barriers faced by sign language users. As we continue to refine this technology and explore additional features, we anticipate that it will play a pivotal role in promoting understanding, collaboration, and respect among all individuals, regardless of their hearing abilities.

REFERENCES

[1] A Crowdsourcing-Based Framework for the Development and Validation of Machine Readable Parallel Corpus for Sign Languages, 2023 , Uzma Farooq; Mohd Shafry Mohd Rahim; Nabeel Sabir Khan; Saim Rasheed; Adnan Abid

[2] Development of an End-to-End Deep Learning Framework for Sign Language Recognition, Translation, and Video Generation , 2022 , B. Natarajan; E. Rajalakshmi; R. Elakkiya; Ketan Kotecha; Ajith Abraham; Lubna Abdelkareim Gabralla

[3] iTechnical Approaches to Chinese Sign Language Processing: A Review, 2019, Suhail Muhammad Kamal; Yidong Chen; Shaozi Li; Xiaodong Shi; Jiangbin Zheng