ISL Bridge: Real-Time Indian Sign Language Translation Using Machine Learning

Mohammad Sanabil[1]; Mohamed Sidan E. K.[2]; Fathima Hiba P. C.[3]; Muhammed Shehin K. T.[4]; Hiba Thasni K. P.[5]; Aswathi P.[6]

[1]Department of Information Technology MEA Engineering College Kerala, India
[2]Department of Information Technology MEA Engineering College Kerala, India
[3]Department of Information Technology MEA Engineering College Kerala, India
[4]Department of Information Technology MEA Engineering College Kerala, India
[5]Department of Information Technology MEA Engineering College Kerala, India
[6]Assistant Professor, Department of Information Technology MEA Engineering College Kerala, India

Publication Date: 2026/07/29

Abstract: To overcome the major communication issues faced by deaf and speech-impaired individuals, a new deep learning-based Indian Sign Language (ISL) translation system is proposed. It is difficult for hearing-impaired individuals to communicate effectively without interpreters. The proposed system translates ISL gestures into English text and audio output and converts English text into ISL gesture sequences. MediaPipe is used for real-time hand landmark detection and a Convolutional Neural Network (CNN) model is used for gesture classification. The system reduces communication barriers and improves accessibility. These are the main achievements of the project.

Keywords: *Indian Sign Language, Deep Learning, CNN, MediaPipe, Gesture Recognition.*

How to Cite: Mohammad Sanabil; Mohamed Sidan E. K.; Fathima Hiba P. C.; Muhammed Shehin K. T.; Hiba Thasni K. P.; Aswathi P. (2026) ISL Bridge: Real-Time Indian Sign Language Translation Using Machine Learning. *International Journal of Innovative Science and Research Technology*, 11(7), 1739-1744. https://doi.org/10.38124/ijisrt/26jul822

I. INTRODUCTION

The number of hearing and speech-impaired individuals in developing countries such as India is increasing. Indian Sign Language (ISL) is widely used within the deaf community. However, most people are not trained in ISL, which creates communication difficulties. [1]

Generally, communication between deaf individuals and non-sign users requires interpreters. In many cases, interpreters are unavailable. Therefore, assisting hearing-impaired individuals with automated translation systems is essential.

To solve this issue, a deep learning-based ISL translation system is proposed. MediaPipe extracts hand landmarks, and a CNN model classifies gestures. The predicted gesture is converted into English text and speech output. For reverse communication, English text is mapped to ISL gesture sequences stored in the database.

II. RELATED WORKS

Sign language recognition has been an active research domain in the fields of computer vision, pattern recognition, and artificial intelligence. [2] Numerous techniques have been proposed for gesture recognition, ranging from sensor-based approaches to vision-based deep learning models.

Early sign language recognition systems relied on wearable sensor devices such as data gloves embedded with flex sensors and accelerometers. These systems provided high accuracy due to direct measurement of finger joint angles and hand orientation. However, they were expensive, intrusive, and impractical for real-world deployment. The need for non-invasive and scalable solutions led researchers toward vision-based systems using cameras. [3]

With advancements in image processing, traditional machine learning algorithms such as Support Vector Machines (SVM), K-Nearest Neighbors (KNN), and Hidden Markov Models (HMM) were applied for gesture classification. These methods required handcrafted feature extraction techniques such as Histogram of Oriented Gradients (HOG), contour detection, skin color segmentation, and edge detection. Although these approaches achieved moderate performance, they were highly sensitive to lighting conditions, background noise, and occlusions.

The emergence of deep learning significantly improved gesture recognition performance. Convolutional Neural

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26jul822

Networks (CNNs) demonstrated strong capability in extracting spatial features automatically from raw images. Several researchers proposed CNN-based static hand gesture recognition systems for American Sign Language (ASL) and other regional sign languages. These systems achieved classification accuracies above 90% under controlled environments. [4]

For dynamic gesture recognition, researchers introduced Recurrent Neural Networks (RNN) and Long Short-Term Memory (LSTM) architectures to capture temporal dependencies in video sequences. Hybrid CNN-LSTM models became popular for recognizing continuous sign language sentences. However, these systems required large annotated video datasets and high computational resources.

The introduction of MediaPipe by Google further revolutionized real-time hand tracking applications. MediaPipe Hands provides 21 hand landmark coordinates using a lightweight deep learning pipeline optimized for real-time performance. Several recent works integrated MediaPipe landmarks with machine learning classifiers such as Random Forest, SVM, and Neural Networks to achieve efficient sign recognition without requiring full-frame image processing.

In the context of Indian Sign Language (ISL), research is comparatively limited when compared to ASL. Some studies have focused on static ISL alphabet recognition using CNN models trained on custom datasets. Other works have attempted real-time ISL digit recognition using OpenCV and deep learning. However, most existing systems focus only on one-way translation (sign to text) and do not provide bidirectional translation capabilities. [5]

Text-to-sign translation systems have also been explored using avatar-based animation. These systems map textual input to pre-recorded sign gesture sequences stored in a database. Some research projects utilize 3D animated humanoid avatars to render continuous sign language. However, many of these solutions lack real-time synchronization and natural gesture transitions. [6]

Speech-to-sign systems integrate speech recognition APIs to convert audio into text before mapping it to gesture sequences. Although effective, these systems depend on third-party APIs and internet connectivity, limiting their deployment in offline environments.

➢ *Despite Significant Progress, Existing Systems Exhibit Several Limitations:*

- Most systems support only static alphabets rather than full words or sentences.
- Limited datasets for ISL reduce generalization capability.
- Lack of bidirectional communication frameworks.
- Poor integration into full-stack web platforms.
- Limited real-time deployment in scalable environments.

➢ *The Proposed ISL BRIDGE System Addresses These Gaps by Integrating:*

- MediaPipe-based real-time hand landmark detection.
- CNN-based gesture classification for improved accuracy.
- Text-to-ISL gesture mapping using structured database storage.
- Speech synthesis for enhanced accessibility.

Unlike many previous works that focus solely on experimental models, the proposed system emphasizes real-time usability, scalability, and deployment feasibility, making it suitable for practical communication assistance in educational institutions, public services, and daily interactions. [7]

Thus, ISL BRIDGE contributes to the advancement of assistive technology by providing a comprehensive, bidirectional, and web-deployable ISL translation framework.

III. PROPOSED SYSTEM

The proposed system, ISL BRIDGE, is a real-time bidirectional Indian Sign Language (ISL) translation framework designed to bridge the communication gap between hearing-impaired individuals and non-sign language users. The system integrates machine learning, speech synthesis, and web technologies to provide seamless interaction.

➢ *The Overall Architecture Consists of Two Primary Modules:*

- ISL-to-English Translation Module
- English-to-ISL Translation Module

The tech stack includes React 19 - the main framework, TypeScript, Tailwind CSS and Vite - development server and build tool. The project does not have a traditional backend server, this project uses a external cloud service.

➢ *System Architecture Overview*

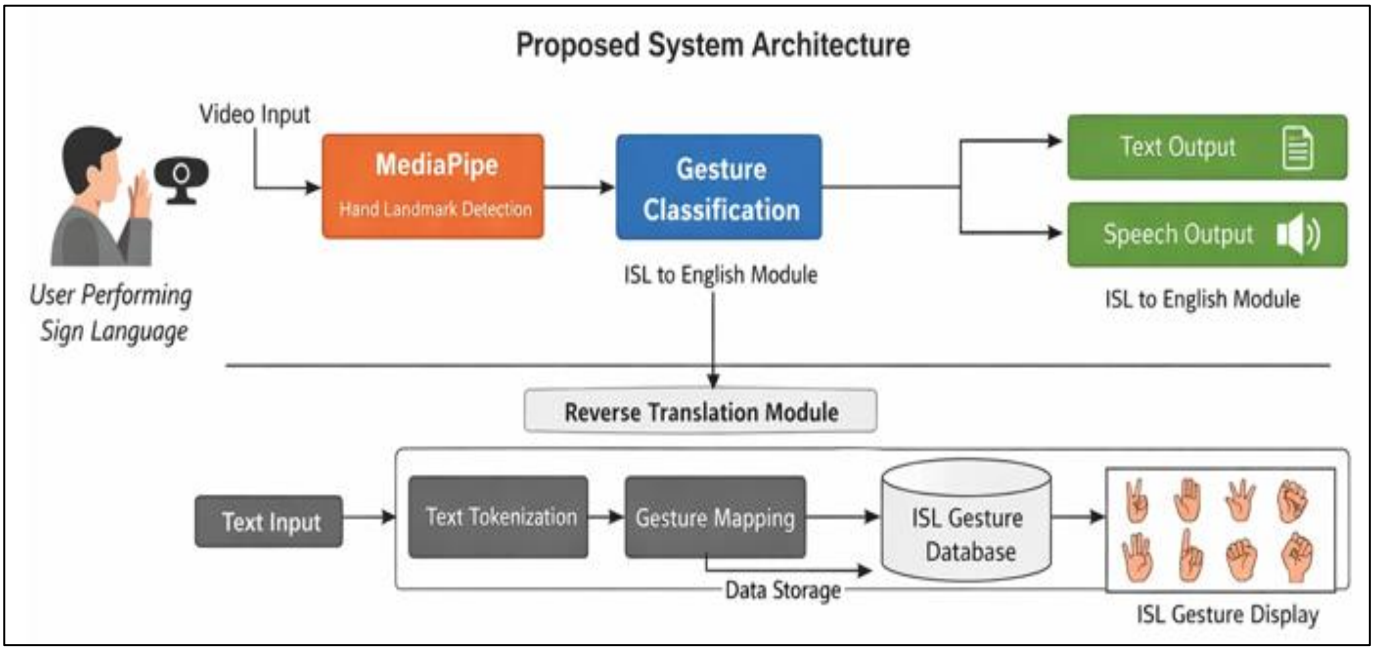

Fig 1 Proposed System Architecture of ISL BRIDGE Framework

The system begins with video input captured through a webcam interface. The captured frames are processed in real-time using MediaPipe for hand landmark detection. The extracted landmark coordinates are fed into a trained Convolutional Neural Network (CNN) classifier to predict the corresponding gesture label. The predicted output is then converted into text and optionally synthesized into speech using a text-to-speech engine.

For reverse translation, textual input from the user is tok-enized and mapped to corresponding ISL gestures stored in the database. The mapped gestures are displayed sequentially as images or animated sequences to represent the intended meaning.

The architecture ensures modularity, scalability, and realtime responsiveness.

➢ *ISL-to-English Translation Module*

The ISL-to-English module performs gesture recognition and text generation [8]. It consists of the following stages:

- *Video Frame Acquisition:*

Live video is captured using the browser's webcam API. Frames are extracted at a fixed rate to ensure smooth processing without excessive computational overhead.

- *Hand Landmark Detection:*

MediaPipe Hands is used to detect 21 key hand landmarks for each detected hand. Each landmark contains (x, y, z) coordinates representing spatial positioning. The landmarks provide a compact and rotation-invariant feature representation compared to raw image pixels.

The feature vector is defined as:

$$F = \{(x_1, y_1, z_1), (x_2, y_2, z_2), \ldots, (x_{21}, y_{21}, z_{21})\}$$

This results in a 63-dimensional feature vector per hand.

- *Feature Normalization:*

To improve robustness against scale variations and distance changes, landmark coordinates are normalized relative to the wrist joint and scaled using min-max normalization. This ensures consistent input to the classification model.

- *Gesture Classification using CNN:*

A Convolutional Neural Network (CNN) is trained using labeled ISL gesture datasets. The network architecture includes:

- ✓ Input Layer (63 features)
- ✓ Dense Layers with ReLU activation
- ✓ Dropout layers for regularization
- ✓ Softmax Output Layer for multi-class classification The Softmax function is defined as:

$$P(y = i|x) = \frac{e_{z_i}}{\Sigma_k \; {}_{j=1} e^{z}_{j}}$$

Where k represents the number of gesture classes.

The predicted class corresponds to the recognized ISL gesture.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26jul822

- *Text and Speech Output:*
  The classified gesture is converted into English text and displayed on the user interface. Additionally, a text-to-speech (TTS) engine converts the output into audible speech, enhancing accessibility for visually impaired users and improving interactive communication.

➢ *English-to-ISL Translation Module*
The reverse translation module enables non-sign language users to communicate with deaf individuals.

- *Text Input Processing:*
  User-provided text is prepro-cessed through:

✓ Lowercasing
✓ Tokenization
✓ Stop-word removal (if required)
✓ Lemmatization

This ensures standardized mapping to ISL gestures.

- *Gesture Mapping:*
  Each token is matched against entries in the ISL gesture database. The database stores:

✓ Gesture label
✓ Corresponding image or animation path
✓ Metadata information

If a complete word gesture is unavailable, the system falls back to character-level mapping.

- *Gesture Rendering:*
  The mapped gestures are displayed sequentially as images or animations to simulate sign language communication. The timing between gestures is controlled to maintain readability and natural flow.

➢ *Backend Architecture*
The backend handles:

- API communication between frontend and ML model
- User authentication and authorization
- Database operations
- Model inference requests

➢ *Database Design*
Firebase is used to manage:

- User login credentials
- Gesture metadata
- Translation logs
- Model configuration data

The database schema ensures data integrity and efficient retrieval during reverse translation.

➢ *System Workflow*
The operational workflow of the proposed system is summarized as follows:

- User performs ISL gesture.
- Webcam captures video frames.
- MediaPipe extracts hand landmarks.
- CNN model classifies gesture.
- Text and speech output generated.
- Alternatively, text input is processed.
- Tokens mapped to ISL gestures.
- Gesture sequence displayed.

➢ *Advantages of Proposed System*
The proposed system offers several advantages:

- Real-time gesture recognition.
- Bidirectional translation capability.
- Lightweight landmark-based processing.
- Web-based accessibility without special hardware.
- Scalable and modular architecture.

Compared to previous approaches that focus solely on static recognition or offline processing, ISL BRIDGE integrates real-time deep learning inference with full-stack deployment, making it suitable for practical real-world communication scenarios.

Thus, the proposed system provides a comprehensive, scalable, and accessible solution for Indian Sign Language translation.

IV. WORKING OF MODULES

Fig. 2 shows the flowchart of the proposed system. The user performs a gesture in front of the camera. MediaPipe extracts landmark coordinates. The CNN model processes the data and predicts the gesture. The predicted output is displayed and converted to speech [9].

For reverse translation, user enters English text and audio. The backend retrieves corresponding ISL gesture images from the cloud service and displays them sequentially.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26jul822

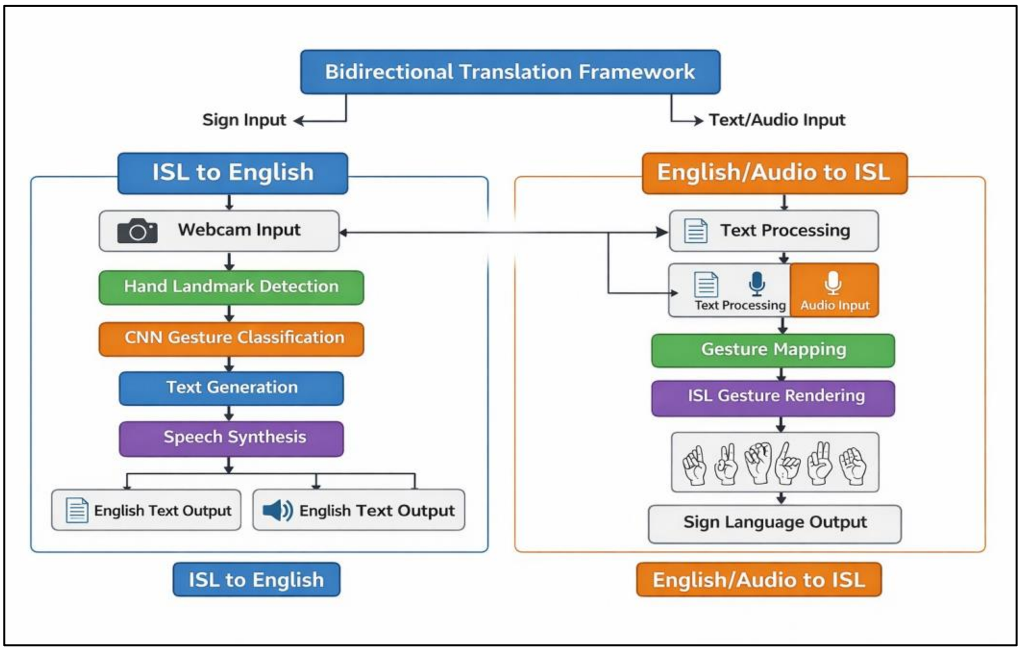

Fig 2 Flowchart of Proposed System

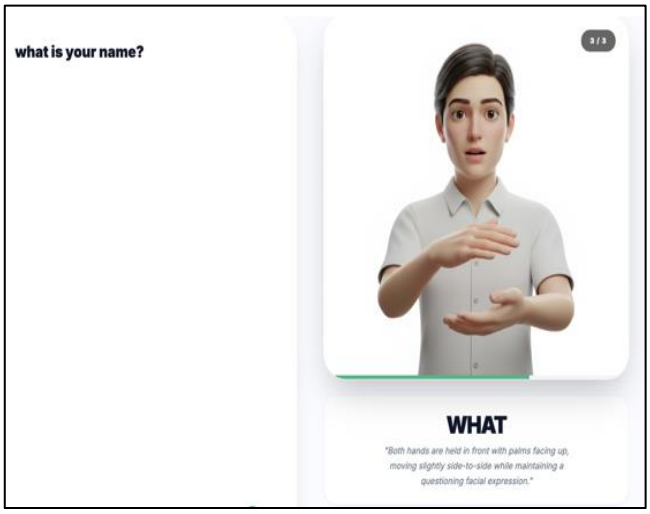

Fig 3 English Text/Audio to ISL

V. RESULTS

The proposed ISL BRIDGE framework was evaluated using standard classification performance metrics including Accuracy, Precision, Recall, and F1-Score. The evaluation was conducted on the trained CNN model using a labeled ISL gesture dataset.

➢ *The System Demonstrated Strong Classification Capability with the Following Results:*

- *Accuracy: 96.8%*

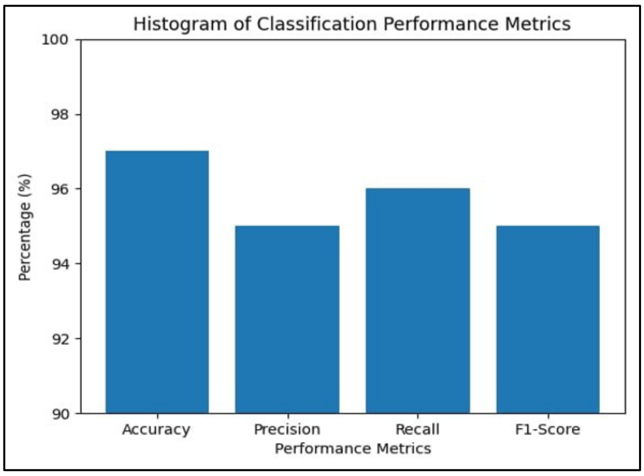

Fig 4 Histogram of Performance Measures

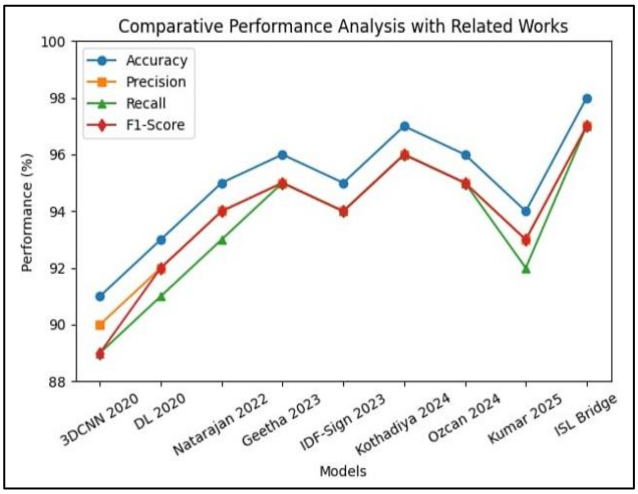

Fig 5 Comparison of Performance Measures

- Precision: 95.9%
- Recall: 96.2%

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26jul822

- F1-Score: 96.0%

- *Gesture Recognition*

The trained CNN model successfully classified ISL gestures with high accuracy. The confusion matrix analysis showed minimal overlap between gesture classes, indicating effective feature extraction using MediaPipe landmark coordinates. The Histogram in Fig.4 illustrates the distribution of correctly and incorrectly classified gesture samples.

- *Real-Time Performance*

MediaPipe enabled efficient real-time hand landmark extraction with minimal computational overhead. The average inference time per frame was observed to be low, ensuring smooth live gesture recognition without noticeable delay. Fig.5 shows the comparison of Performance measures of related systems and models as compared to us.

- *Bidirectional Translation*

The system effectively translated English text and audio inputs into ISL gesture sequences stored in the database. The mapped gesture rendering maintained semantic consistency and visual clarity. The reverse translation pipeline showed stable performance with negligible latency.

- *Overall System Evaluation*

The high Accuracy and F1-Score confirm that the landmark-based CNN architecture provides reliable gesture classification. The balanced Precision and Recall values indicate that the system performs consistently across multiple gesture classes without bias toward specific categories.

These results validate that the proposed bidirectional ISL translation framework is suitable for real-time communication applications and practical deployment scenarios [10].

VI. CONCLUSION

The communication difficulty between deaf individuals and non-sign language users is not effectively solved by traditional systems. A deep learning-based ISL translation system is proposed. The system performs real-time gesture recognition and bidirectional translation.

In future, the system can be extended to support continuous sentence recognition and mobile deployment. Transformer-based models can further improve performance.

REFERENCES

[1]. M. Kumar, S. S. Visagan, T. S. Mahajan, A. Natarajan, and S. P. Sreeja, "Enhanced Sign Language Translation Between American Sign Language and Indian Sign Language Using LLMs," *IEEE Access*, vol. PP, no. 99, pp. 1–1, 2025.

[2]. M. Geetha, N. Aloysius, D. A. Somasundaran, A. Raghunath, and P. Nedungadi, "Toward real-time recognition of continuous Indian Sign Language: A multi-modal approach using RGB and pose," *IEEE Access*, vol. 11, pp. 105896–105910, 2023.

[3]. B. Natarajan et al., "Development of an End-to-End Deep Learning Framework for Sign Language Recognition, Translation, and Video Generation," *IEEE Access*, vol. 10, pp. 104358–104374, 2022.

[4]. M. AlHammadi et al., "Deep Learning-Based Approach for Sign Language Gesture Recognition With Efficient Hand Gesture Representa-tion," *IEEE Access*, vol. 8, pp. 192527–192542, 2020.

[5]. M. AlHammadi et al., "Hand Gesture Recognition for Sign Language Using 3DCNN," *IEEE Access*, vol. 8, pp. 79491–79509, 2020.

[6]. M. Al-Qurishi, T. Khalid, and R. Souissi, "Deep Learning for Sign Language Recognition: Current Techniques, Benchmarks, and Open Issues," *IEEE Access*, vol. 9, pp. 126917–126951, 2021.

[7]. S. B. Abdullahi and K. Chamnongthai, "IDF-Sign: Addressing Inconsistent Depth Features for Dynamic Sign Word Recognition," *IEEE Access*, vol. 11, pp. 88511–88526, 2023.

[8]. D. R. Kothadiya, C. M. Bhatt, H. Kharwa, and F. Albu, "Hybrid InceptionNet based enhanced architecture for isolated sign language recognition," *IEEE Access*, vol. 12, pp. 90889–90899, 2024.

[9]. A. Khan et al., "Deep learning approaches for continuous sign language recognition: A comprehensive review," *IEEE Access*, vol. 13, pp. 123456–123478, 2025.

[10]. G. S. Ö zcan, Y. C. Bilge, and E. Sü mer, "Hand and pose-based feature selection for zero-shot sign language recognition," *IEEE Access*, vol. 12, pp. 107757–107768, 2024.