Journal of Ubiquitous Computing and Communication Technologies (ISSN: 2582-337X)
www.irojournals.com/jucct/

Indian Sign Language Recognition and Translation System

Senthil Arun M.[1], Suresh P.[2], Prakash K.[3], Syed Ibrahim A.[4], Merlin Gethsy D.[5]

[1,2,3,4]UG Scholar, [5]Assistant Professor, Department of Computer Science and Engineering, V V College of Engineering, Tisaiyanvilai, India.

E-mail: [1]senthilarun72@gmail.com, [5]merlingethsy@gmail.com

Abstract

The research focuses on overcoming the substantial communication barriers faced by deaf people in India because of the lack of professional sign language interpreters. The paper stresses the significance of ISL (Indian Sign Language) in enabling efficient communication in different sectors such as education and health care. The paper describes a system for recognizing and translating ISL that uses computer vision and machine learning methods. Specifically, a web camera is used to record hand gestures, which undergo the process of being tracked using MediaPipe framework. In addition, the system assesses the spatial and temporal aspects of hand movements by applying Random Forest, CNN, and BiLSTM models. Recognized signs are translated into text form, arranged in sentences and delivered through text-to-speech synthesizer. Thus, the web-based application does not need special hardware, which makes it more user-friendly. Experiments show that the system is able to recognize hand gestures and generate text and speech outputs in diverse environments thus helping to overcome the problem of communication gaps between deaf and hearing individuals. The efficacy of the comprehensive system test is excellent in practical terms; the Bi-LSTM model yielded an accuracy rate of 92.5% in dynamic word recognition, whereas the CNN models yielded an accuracy rate of 95.2% for alphabets and 96.3% for numerical gestures, and an excellent accuracy rate of 99.8% for static words, thus beating the accuracy benchmark of 81.0% set by conventional base reference systems.

DOI: https://doi.org/10.36548/jucct.2026.2.006
Extended version of article received from **International Conference on Next Generation Engineering Innovations, AI and Sustainable Automation – 2026**
Received: 29.03.2026, received in revised form: 02.05.2026, accepted: 20.05.2026, published: 01.06.2026

Keywords: ISL - Indian Sign Language, BiLSTM, CNN, Sign Language Recognition, Sign Language Translation, Random Forest, Hand Tracking, Gesture Recognition.

1. Introduction

It is essential that sign language serves as a means of communication for the deaf and hearing impaired. In India, there are millions of deaf Indians who communicate through ISL. However, majority of the hearing populace remains unaware of ISL, leading to a major difference between the means of communication adopted by deaf individuals and others. The problem affects various facets of life, including the provision of education, health care, employment, and other public utilities. It is customary to use human interpreters for communication between the deaf and the hearing populations. Such interpreters may be costly and unavailable in several situations.

Advances in computer vision and machine learning have made the development of automatic gesture translation systems feasible. A sign language recognition system detects and decodes human gestures to translate them into meaningful text or speech through cameras and intelligent software systems. Hand gesture recognition is used in order to translate sign languages into meaningful text or speech.

The proposed Indian Sign Language Recognition and Translation System implements advanced deep learning techniques for the detection of gestures in real-time. The system employs MediaPipe Hands for the detection of hand landmarks in video feed through a webcam and their spatial feature extraction. Dynamic gestures are identified using a bidirectional long short-term memory (BiLSTM) model trained on the dynamic movement of gestures over several frames. Static gestures, such as letters and numbers, are identified using machine learning classifiers, which include Random Forest and CNN.

The gestures are formed into sentences, corrected by local LLM-based grammar correction, translated to several other languages, and synthesized as speech using text-to-speech synthesis technology. The entire process, starting from the recognition of the gestures to the output of the synthesized speech, forms a complete communication cycle in the proposed method. In addition, the system avoids the use of specialized hardware devices and is implemented entirely using the browser, along with a basic webcam.

2. Literature Review

Recognizing sign language has emerged as a prominent topic in computer vision and deep learning because of its significance in providing better means of communication to those who are deaf or hard-of-hearing. A variety of papers has explored the topics of gesture detection, sign classification, and sign language translation using machine learning or deep learning algorithms.

One such example is the research paper written by Kumar et al. [1]. They developed a framework for translating signs from American Sign Language to Indian Sign Language through combining gesture recognition and Large Language Models for correcting grammar and improving the context. In another example, Gong et al. [2] proposed a framework named SignLLM, which translates visual sign sequences into language-like representation and then interprets them with the help of Large Language Models.

A word-level sign language recognition approach based on MediaPipe and LSTM-GRU network architecture was proposed by Navendu and Sahula [3], showing that sequential deep learning models are efficient for detecting hand movement pattern. The model was able to make predictions in real time by recognizing dynamic gestures. In their study, Deshpande et al. [4] used CNN to develop a model for sign language recognition. They used the CNN model to detect images of hand gestures obtained via a camera and then convert them to readable output.

In another study, Kurik et al. [5] used MediaPipe and LSTM networks for sign recognition and were successful in achieving real-time prediction of gestures using two-handed signs. It further showed that landmarks for recognizing gestures can be used to extract relevant features. Using an LSTM model for gesture recognition, Kumar et al. [6] also developed a framework in computer vision which detects hand and body landmarks to enable sign-to-text conversion. Banu et al. [7] used MediaPipe hand landmarks along with random forest classifier to develop a real-time sign language recognition system. This technique has provided successful results regarding accurate sign language recognition and proved that the machine learning classifiers are able to provide accurate classification for landmark-based features with very low computational power requirement.

Another study was carried out by Kumar et al. [8], where they presented an Indian Sign Language Recognition Model based on mediaPipe landmarks. This study highlighted the effectiveness of landmark-based visual representation of sign language under varying

backgrounds and illuminations. On the other hand, Sharma et al. [9] presented a real-time sign language recognition and translation system that used MediaPipe landmarks and random forest algorithms to convert sign language into text. Furthermore, another relevant research study is provided by Al Abdullah et al. [10]. They performed a systematic literature review on Sign language recognition techniques and concluded that deep learning models including CNNs and recurrent neural network had significantly enhanced the sign language recognition performance and capability to use real-time techniques.

While previous studies have presented promising findings related to sign language recognition, many existing systems concentrate on static gesture classification or translating gesture text. There has been very little work on a holistic real-time Indian Sign Language communication system, which consists of a static and dynamic gesture recognition module, sentence generation module, language translation module, and speech synthesis module, all under one web-based umbrella application. In an attempt to bridge this gap, the current work presents a real-time Indian Sign Language recognition and translation system, which involves MediaPipe hand landmark extraction, static gesture recognition with random forest and CNN, and dynamic gesture recognition with BiLSTM. The detected gestures can be turned into sentences and then translated to speech output.

3. Proposed Methodology

The proposed system aims to develop an automated Indian Sign Language (ISL) recognition and translation system that can interpret hand gestures and convert them into readable text and speech. The system utilizes computer vision techniques and deep learning models to analyze hand movements captured through a webcam. The overall workflow includes several stages such as data acquisition, image preprocessing, hand landmark extraction, feature extraction, and gesture classification. Initially, the system captures hand gesture images using a standard webcam, and the captured images are processed using MediaPipe Hands, which detects the hand region and extracts 21 landmark points representing key joints of the fingers and wrist. These landmarks provide spatial information about the hand pose, and the extracted landmark coordinates are normalized and used as input features for machine learning models. For gesture recognition, the system uses different models depending on the type of gesture. Static gestures such as alphabets and numbers are classified using Random Forest and Convolutional Neural Network (CNN) models, while dynamic gestures that involve movement

over time are recognized using a Bidirectional Long Short-Term Memory (BiLSTM) network capable of analyzing temporal sequences of hand movements. Finally, the recognized gestures are converted into words and assembled into sentences, enabling real-time communication assistance between deaf and hearing individuals.

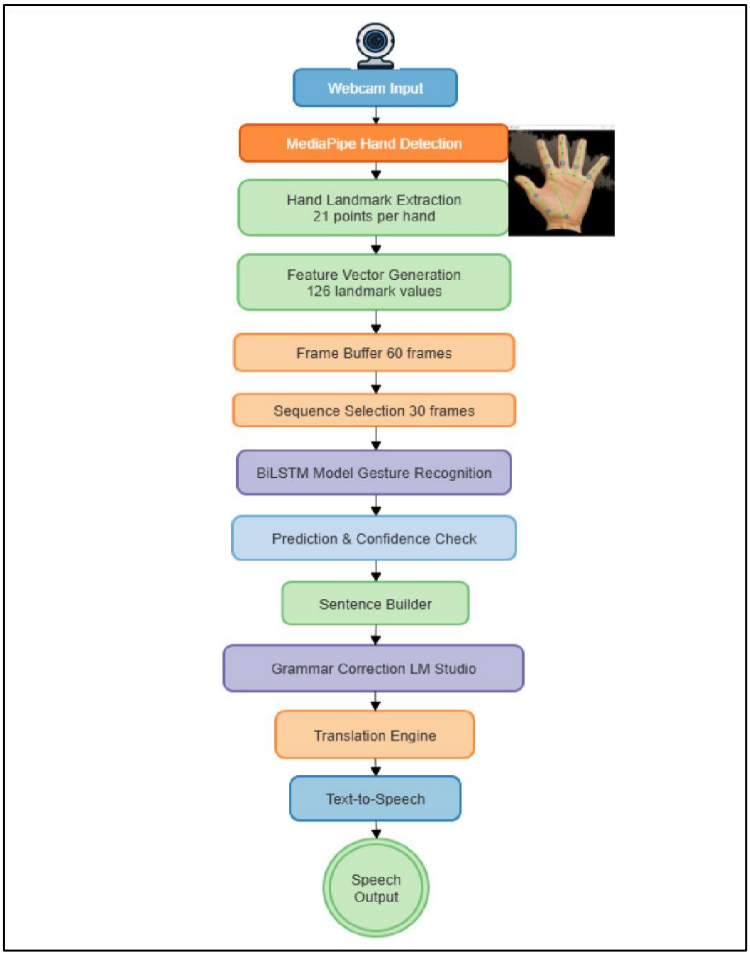

Figure 1. Overall Workflow of the Proposed System

Figure 1 illustrates the overall workflow of the proposed Indian Sign Language (ISL) recognition and translation system. The system begins with capturing real-time video input from a webcam. The captured frames are processed using MediaPipe Hands, which detects the hand region and extracts landmark points representing the joints of the fingers and wrist. These landmarks provide spatial information about the hand pose and are used to generate feature vectors for gesture analysis. The extracted features are stored temporarily in a frame buffer and a sequence of frames is selected for gesture recognition. These feature sequences are then passed to the Bidirectional Long Short-Term Memory (BiLSTM) model, which analyzes the temporal patterns of hand movements and predicts the performed gesture. After prediction, a

confidence check is applied to ensure reliable recognition results. The recognized gesture is then added to a sentence builder module where words are combined to form meaningful sentences. Finally, the generated sentence undergoes grammar correction and translation, and the output is converted into speech using a text-to-speech module. This process enables real-time communication by translating sign language gestures into understandable text and speech.

The system architecture of the ISL recognition and Translation system consists of several modules, each playing an integral part in identifying the hand gestures as shown in Figure 2.

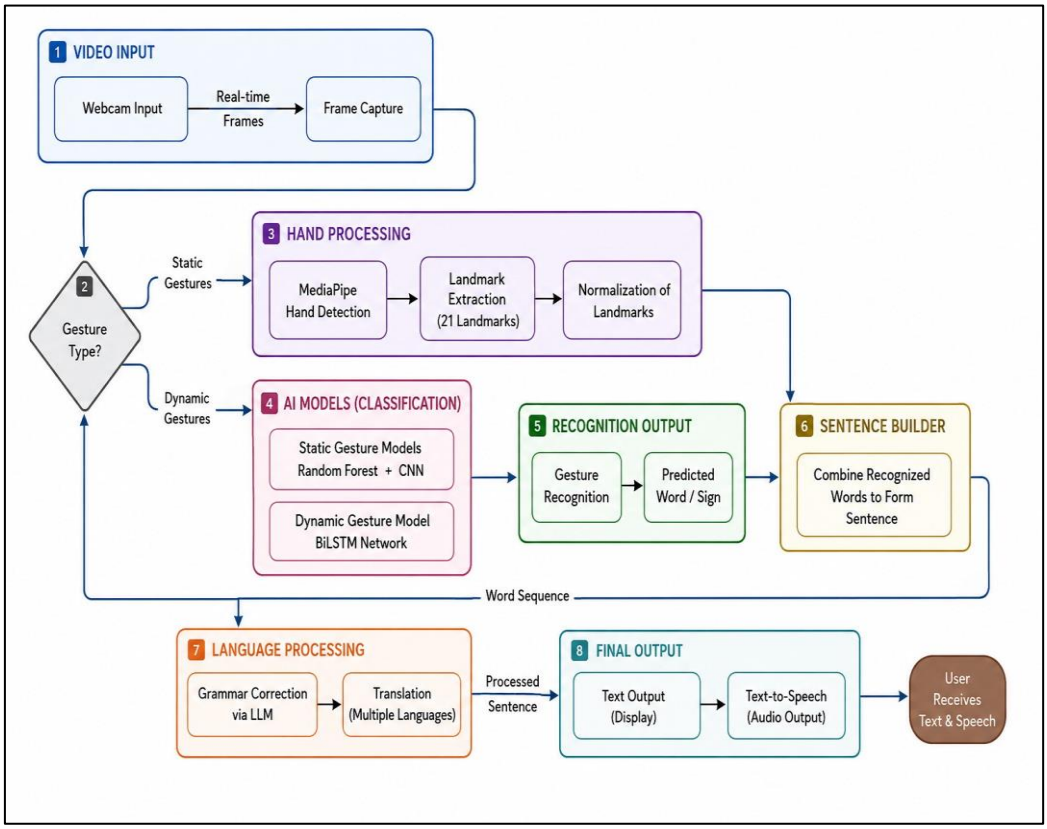

Figure 2. Architecture of the Proposed ISL Recognition and Translation Framework.

The system starts with video input modules where webcam continuously feeds frames of the user doing sign language gesture. The hand gesture from the frame gets detected in the hand detection module and landmarks are extracted using the media pipe library. The landmarks obtained are passed to feature extraction modules where the landmarks get converted into a feature vector format used by the machine learning model for further processing. Feature vectors are analyzed by gesture classification module, where the machine learning models are used to identify the gestures performed by analyzing features vectors. Some examples include

random forest, CNN model, and Bi-LSTM models. Once the performed gesture gets identified, the information passes to sentence builder modules where the recognized words are concatenated to build meaningful sentences. Further, the sentence could be translated in any required language, as well as speech generation could be implemented with the help of text-to-speech module.

3.1 Dataset

The data set utilized in the design of the Indian Sign Language Recognition and Translation system was obtained from the freely available Kaggle repository for Indian Sign Language gesture recognition [11]. The data set consists of multiple hand gesture images representing the Indian sign language alphabets, numerals, and gestures. This data set comprises images taken under different lighting and background conditions, thereby increasing the robustness of the recognition model when tested in real-time environments. The images contain both hand pose and gesture variations, thus providing a good dataset for training machine learning and deep learning models.

To perform preprocessing of the images, the hand gesture images were resized and normalized before features could be extracted. MediaPipe Hands was then used to detect and extract 21 hand landmark points in each gesture image. Hand landmark points are coordinate values representing detailed information on the finger positions and hand orientations. These landmarks were used to train the random forest, CNN, and BiLSTM models.

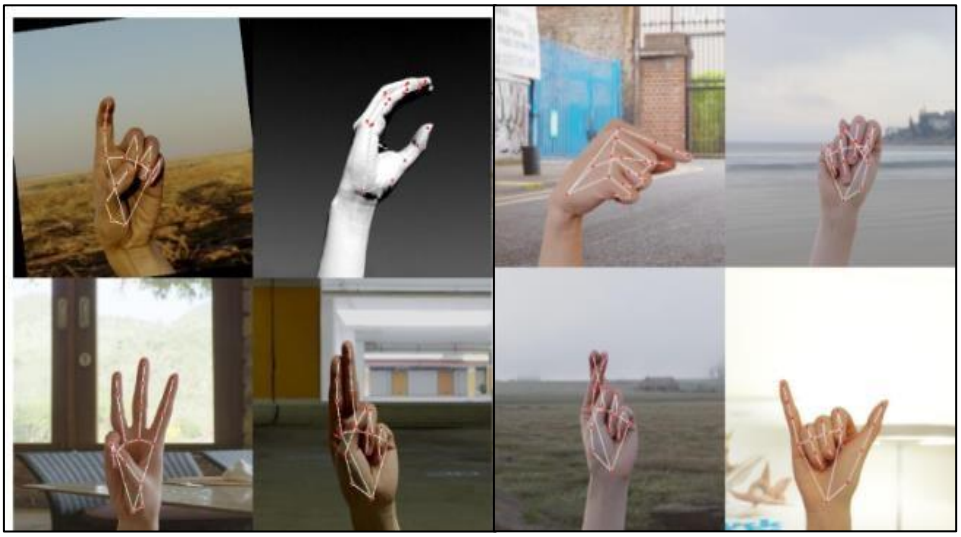

Figure 3. Input Hand Gesture for Sign Language Recognition

The dataset used for this research study has been sourced from the Kaggle Indian Sign Language Recognition repository (see Figure 3) and is being utilized for the training and evaluation process of the proposed gesture recognition model. This is an image-based dataset designed for ISL gesture classification and can be utilized for machine and deep learning tasks such as automatic sign language translation, assistive communication systems, and gesture recognition. The dataset comprises labeled images for hand gestures categorized into different classes, where each class folder corresponds to individual alphabet, numeral, or frequently used gestures within Indian Sign Language. A typical set of ISL gesture dataset comprises around 12,000 to 31,000 images in 26 classes, each containing almost 1,200 images per class that are quite enough for training the models. The images have been collected under varying lighting conditions, different hand poses, and backgrounds to enhance the robustness and generalizability capabilities of the gesture recognition model in real-time application scenarios. To facilitate the development of models effectively, the data set was divided into three categories, training set, validation set, and test set, where a ratio of 80:10:10 was considered while performing this operation. For this purpose, the training set was used to train the gesture model, whereas the parameters of the recognition process were tuned through the validation set. Furthermore, the performance of the developed gesture recognizer was analyzed by using the testing set. Before carrying out the classification process, all images were resized and normalized in order to extract the hand region with 21 landmarks of fingers and wrists using the MediaPipe Hands model. These extracted landmarks were then fed into the Random Forest, CNN, and BiLSTM classifier to recognize both static and dynamic gestures. In this way, a dataset was created that greatly helped improve the accuracy of the recognition process.

3.2 Hand Detection

Hand detection plays an important role in identifying and tracking hand gestures accurately. In the proposed system, MediaPipe Hands is used to detect hands in real-time video streams. MediaPipe uses a machine learning pipeline that can detect up to two hands simultaneously and extract 21 key landmark points for each hand as shown in Figure 4. These landmarks represent different parts of the hand including the wrist, finger joints, and fingertips. The coordinates of these points are normalized and converted into feature vectors that can be used as input for classification models. By tracking the movement of these landmarks across multiple frames, the system can capture both static hand poses and dynamic hand gestures.

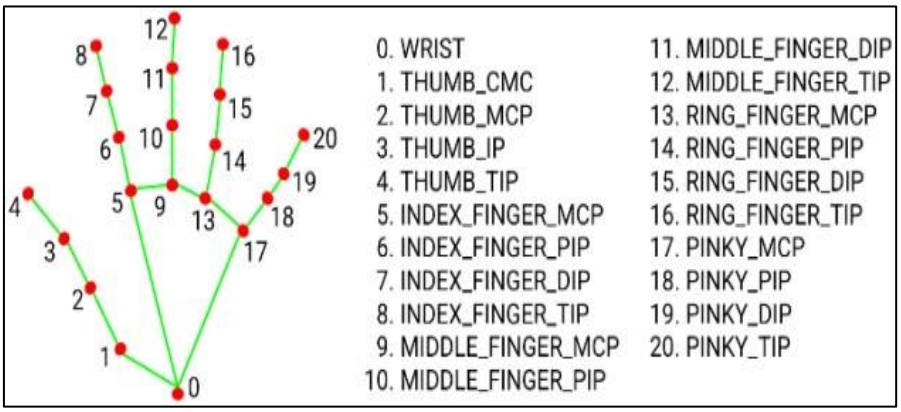

Figure 4. MediaPipe-based Hand Landmark Detection with 21 Key Points

3.3 Feature Extraction

Feature extraction is a crucial step in the proposed Indian Sign Language recognition system since it involves converting raw hand gesture data into meaningful numerical features that can be processed by machine learning algorithms. In this study, MediaPipe Hands has been used to obtain 21 hand landmark points from each detected hand gesture. These landmarks include important finger joints, fingertips, and wrist positions giving details about hand structure and gesture posture.

Landmark points were normalized to eliminate variations due to the hand distance, orientation, and camera position. From the landmarks, important features like relative joint positions, finger alignment, hand shape, and trajectory were obtained. For static hand gestures, the spatial features of landmarks were used for classification through Random Forest and CNN models. However, for dynamic gestures, sequences of landmark frames were gathered over time and processed through the BiLSTM model.

Feature extraction improved the performance of the proposed model by enabling it to discriminate between similar hand gestures. Additionally, feature extraction simplified the process of processing raw image frames. Thus, the model was efficient enough to be used in real-time sign language recognition systems.

3.4 Gesture Classification

The classification stage is responsible for identifying the gesture performed by the user based on the extracted features. The proposed system uses different machine learning models to recognize static and dynamic gestures. For static gestures such as alphabets and numbers, a

Random Forest classifier and a Convolutional Neural Network (CNN) are used. The Random Forest model consists of multiple decision trees that collectively determine the most likely class label for a given gesture. CNN models are capable of learning complex patterns from the extracted feature vectors, improving classification accuracy.

For dynamic gestures that involve hand movement, a Bidirectional Long Short-Term Memory (BiLSTM) network is used. LSTM networks are specifically designed to analyze sequential data, making them suitable for recognizing gestures that evolve over time. The bidirectional structure allows the model to analyze gesture sequences in both forward and backward directions, enabling more accurate recognition. The final output of the classification stage is the predicted gesture label along with a confidence score indicating the reliability of the prediction.

4. Results and Discussion

The proposed Indian Sign Language (ISL) Recognition and Translation System was evaluated by testing it on real-time webcam input to assess its ability to recognize hand gestures and translate them into text and speech. The system uses MediaPipe for hand gesture detection and hand landmarks. Hand landmarks are analyzed using the Random Forest, CNN, and BiLSTM algorithms to recognize static and dynamic gestures. The experimental results confirm that the proposed system can accurately detect hand gestures in real-time and provide meaningful output for aiding communication. Also, the system works reliably under normal lighting conditions and different hand positions.

4.1 Hand Landmark Detection

Figure 5. Real-Time Hand Landmark Detection and Tracking

Detection of the hand of the user along with the extraction of significant hand landmarks constitutes the first stage of the process. This stage was achieved using the MediaPipe framework, which successfully detected the hand region as well as identified 21 landmarks on the hand. Hand landmarks were necessary for detecting the movements of the hand. Real-time detection of hand gestures was possible without any delay in recognition. Figure 5 demonstrates the achievement of extracting hand landmarks through the use of the MediaPipe framework.

4.2 Word Gesture Recognition

Figure 6. Recognition of the ISL Gesture Corresponding to the Word “Good”

The proposed model has proved quite efficient in recognition of the signs at the word level with the help of the landmark-based features extracted from the signs. The dynamic gestures have been analyzed by applying the BiLSTM model to them. Figure 6 shows the example recognition of the word-level gesture of "Good". The output has been recognized quite efficiently, showing the efficiency of the system in recognizing gesture sequences and producing corresponding words.

4.3 Number Gesture Recognition

Apart from the word gestures, the model has also been evaluated on recognizing numerical gestures. The static gestures used for indicating numerals were also able to be identified by the random forest and convolutional neural network models. As shown in Figure 7, it was successfully able to recognize the hand gesture that denotes the digit “3”.

Figure 7. Recognition of the Numerical Hand Gesture Representing Digit “3”

4.4 Sentence Generation Output

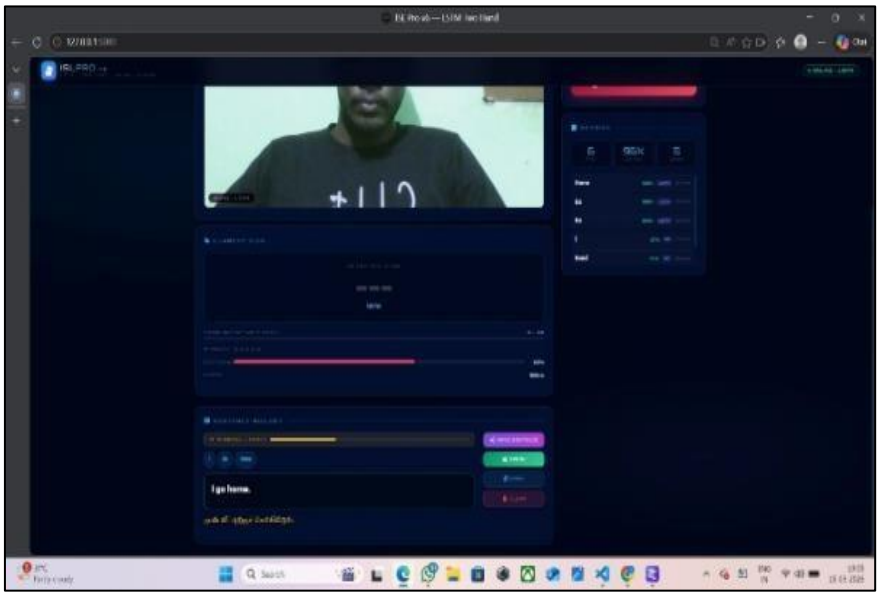

Figure 8. Sentence Generation from Recognized Sign Language Gesture

Having recognized individual gestures, the system uses the identified words to construct a sentence. In the sentence builder stage, the recognized outputs are placed in order to create a sensible sentence structure. Figure 8 demonstrates how the recognized gestures are transformed into a sentence “I go home.” This process makes the use of the system more practical by allowing constant communication.

4.5 Final Output with Translation

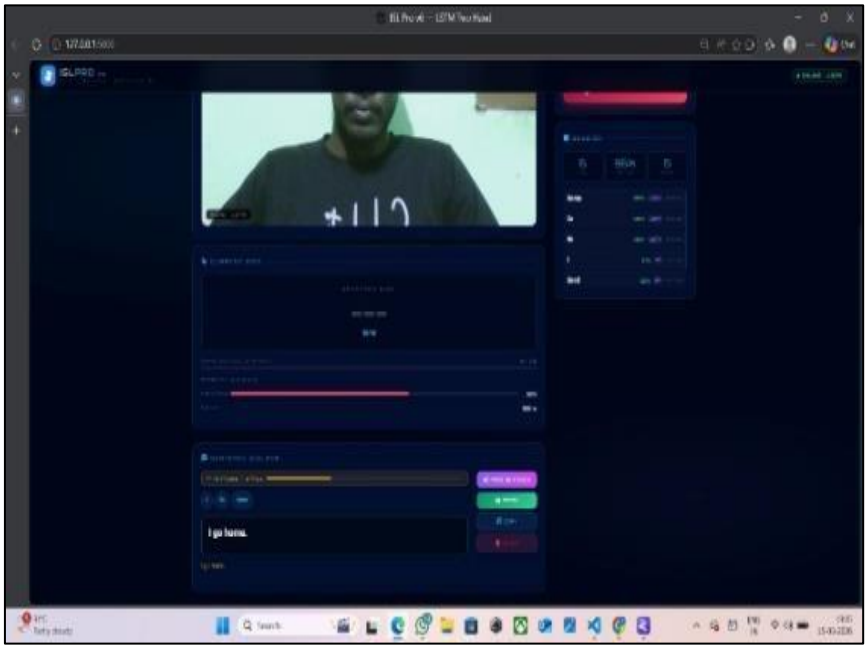

Figure 9. Final System Output with Multilingual Translation and Text-to-Speech Conversion

Figure 9 illustrates the final output of the system. The last stage of the process involves the transformation of the generated sentence into translation and speech output. The sentence can be voiced out with the help of the text-to-speech module making it easier for the hearing person to comprehend the message conveyed. This makes the system more accessible and efficient for use in interactions between deaf and hearing people.

4.6 Comparative Analysis of Validation Accuracy

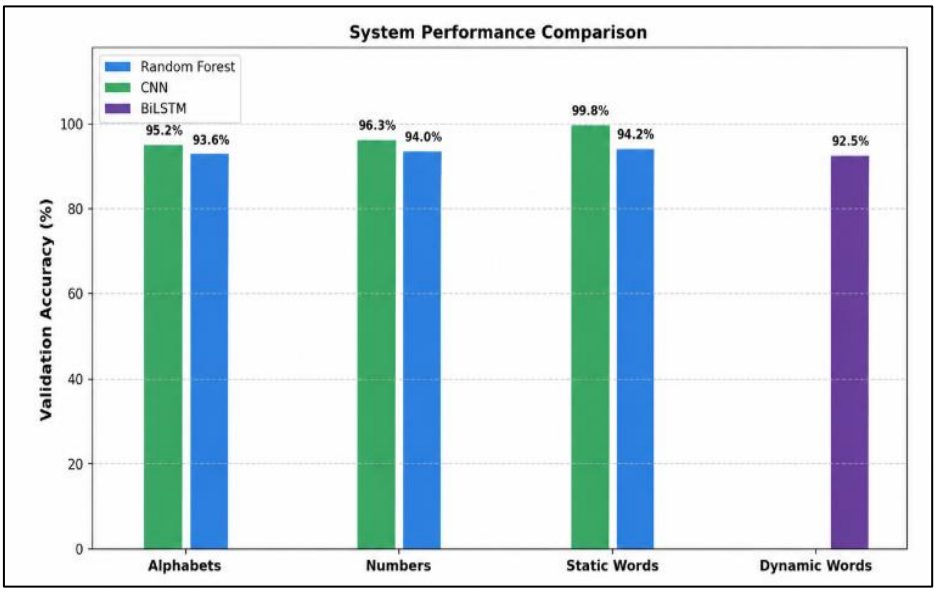

Figure 10. Comparative Analysis of Validation Accuracy

The chart above (Figure 10) shows the validation accuracy obtained from the proposed machine learning and deep learning algorithms for various classes of Indian Sign Language gestures. The accuracy level obtained for the CNN model was 95.2% for alphabet recognition and 96.3% for the numerical gestures recognition. This clearly proves that the CNN model is capable of recognizing and categorizing different static gestures of the hands. In the recognition of static words, the model was able to achieve a high accuracy level of 99.8%. The BiLSTM model was used to classify dynamic words, recording an accuracy rate of 92.5%. This shows that the model is capable of capturing the pattern of hand gestures.

The proposed Indian Sign Language recognition and translation system can be applied in several real-world scenarios to assist communication between deaf and hearing individuals. One of the primary applications is in educational institutions, where the system can help students learn sign language and improve accessibility for deaf learners. In healthcare environments, the system can assist patients who rely on sign language to communicate with doctors and medical staff. The system can also be integrated into public service centers, banks, and government offices to provide inclusive communication support. Furthermore, the technology can be incorporated into mobile applications and smart devices to enable real-time gesture recognition using smartphone cameras. By providing an automated translation of sign language gestures into text and speech, the system helps reduce communication barriers and promotes social inclusion.

5. Conclusion

The research work proposed a Real-Time Indian Sign Language (ISL) Recognition and Translation System aimed at enhancing the communication process between deaf and hearing people. The proposed system utilized computer vision and deep learning techniques in recognizing the hand gestures acquired via webcam and translating them into meaningful texts and speech. MediaPipe was used to accurately detect hand landmarks, whereas the Random Forest, CNN, and BiLSTM models were applied to recognize both the static and dynamic gestures. The proposed system was capable of generating sentence-level outputs and provided functionalities such as translation and text-to-speech conversion to make the communication process easy and comprehensible. The experiments conducted in this regard indicated that the proposed framework could carry out gesture recognition efficiently in real-time environments with effective performance and response. Furthermore, the proposed system does not require

any specialized hardware. The following improvements could be made in the future for this system: the expansion of the gesture database to incorporate more sophisticated words, sign sentences, and regional dialects of the Indian Sign Language. Finally, incorporating capabilities for mobile applications, cloud processing, and multilingual communication would enhance its usability, enabling the system to be implemented in academic and healthcare settings, as well as in public communication domains.

References

[1] M. Kumar, S. Sarvajit Visagan, T. Mahajan, A. Natarajan and P. S. Sreeja, "Enhanced Sign Language Translation Between American Sign Language and Indian Sign Language Using LLMs," in IEEE Access, vol. 13, 2025, 156270-156284.

[2] Gong, Jia, Lin Geng Foo, Yixuan He, Hossein Rahmani, and Jun Liu. "Llms are Good Sign Language Translators." In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2024, 18362-18372.

[3] Navendu, Kumar, and Vineet Sahula. "Word Level Sign Language Recognition using MediaPipe and LSTM-GRU Network." In 2024 IEEE International Symposium on Smart Electronic Systems (iSES), IEEE, 2024, 13-18.

[4] Deshpande, Aditi, Ansh Shriwas, Vaishnavi Deshmukh, and Shubhangi Kale. "Sign Language Recognition System Using CNN." In 2023 International Conference on Intelligent and Innovative Technologies in Computing, Electrical and Electronics (IITCEE), IEEE, 2023, 906-911.

[5] Kurik, Glory Cornelia Patining, Setiawan Hadi, and Deni Setiana. "Sign Language Recognition System for the Deaf Using Mediapipe Based on the Long-Short Term Memory Method." In 2024 International Conference on Information Technology Systems and Innovation (ICITSI), IEEE, 2024, 21-27.

[6] Kumar, N. Suresh, Chintapudi Harika, Chennamsetty Harika, Chintala Sankar Reddy, and Dasari Pavan Kalyan. "Sign to Text: Automated Sign Language Interpretation using LSTM and Computer Vision." In 2024 3rd International Conference on Automation, Computing and Renewable Systems (ICACRS), IEEE, 2024, 1414-1419.

[7] Banu, Priya V., Raja T. Narasimma, A. Bharath, K. S. Sriram, and S. Vinoth. "Real-Time Sign Language Recognition with MediaPipe Hand Landmarks and Random Forest Classification." In 2025 Tenth International Conference on Science Technology Engineering and Mathematics (ICONSTEM), IEEE, 2025, 1-8.

[8] Kumar, Vishal, R. Sreemathy, Mousami Turuk, Jayashree Jagdale, and Agrima Agarwal. "Real-Time Indian Sign Language Recognition Using Skeletal Feature Maps." In 2023 14th International Conference on Computing Communication and Networking Technologies (ICCCNT), IEEE, 2023, 1-6.

[9] Sharma, Garvit, Prakshep Gusain, Aman Verma, Harsha Saini, Rishi Kumar, and Guru Prasad. "Real-Time Sign Language Recognition and Translation Using Mediapipe and Random Forests for Inclusive Communication." In 2025 2nd International Conference on Computational Intelligence, Communication Technology and Networking (CICTN), IEEE, 2025, 886-890.

[10] Al Abdullah, Bashaer A., Ghada A. Amoudi, and Hanan S. Alghamdi. "Advancements in Sign Language Recognition: A Comprehensive Review and Future Prospects." IEEE Access 12 (2024): 128871-128895.

[11] Dataset: https://www.kaggle.com/datasets/satwikpasumarthi/indian-sign-language-recognition