Published by :
https://www.ijert.org/
An International Peer-Reviewed Journal

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14, Issue 06, ACSCON - 2026

Real Time Sign Language Translation Using AI

Ms. Deepa Srinivasa[1] Ashika P[2] Chethan G S[3] Chidhvila M[4]

[1] *Assistant Professor, Dept Of CSE, ACS College Of Engineering, Bengaluru, India*
[2] *Ashika P, Dept Of CSE Student, ACS College Of Engineering, Bengaluru, India*
[3] *Chethan G S, Dept Of CSE Student, ACS College Of Engineering, Bengaluru, India*
[4] *Chidhvila M, Dept CSE Student, ACS College Of Engineering, Bengaluru, India*

Abstract - *This project proposes an Indian Sign Language (ISL) translation system for real-time bidirectional translation of ISL to text and text to ISL which aims to address the communication gap between the hearing community and the Deaf and/or mute community. The bidirectional module is associative specifically to ISL users who are using ISL to text, or text to ISL. To recognize ISL to text, we respond by utilizing computer vision algorithms utilizing live video images to capture and recognize the user's hand gestures in ISL. Once hand gestures are made, we can apply ML models to classify the gesture so that we provide the appropriate text output. The text to ISL module maps the text that the user types in into pre-defined ISL gesture animations or still images that a hearing person can understand and begin to use [ISL]. Fast, clear two-way translation matters here - built so everyone can connect easier. Still, there's more: helping people talk every day using tech made with the Indian Deaf Community in mind. Watch how machines learn gestures instead of words, letting humans link up without speaking at all. That kind of tool helps when you do not know ISL yet need to share thoughts with someone deaf. Barriers drop once both sides find ways to reach each other. Simple design keeps it open - anyone can step in and make sense of it right away. When tech delivers answers right away, conversations flow free of hiccups - like chatting face to face. Here's how involvement grows: fresh access opens doors at school, on the job, in daily tasks.*

Key Words: Indian Sign Language, Gesture Recognition, Real-time Translation, Computer Vision, Deep Learning.

1. INTRODUCTION

Communication sits at the core of every person's life. Yet countless individuals dealing with hearing difficulties struggle to connect with those who hear. In India, sign language forms the main way deaf people share thoughts - it has its own rhythm, its own flow. Few beyond their community can follow along. Most outsiders remain unaware of how it works. Misunderstandings grow easily when knowledge fades into silence. A new effort takes shape against that gap. Technology steps forward, shaped by patterns found through machine learning. Real-time dialogue becomes possible, back and forth, without delay. Machines learn gestures, translate them swiftly, build bridges where walls once stood. Understanding flows both ways now, motion meeting voice. A new way to translate things works like this: Starting off, a camera records hand movements as they happen. Instead of relying on stored clips, it works straight from live footage. Then comes detection - each motion gets recognized instantly. Following that step, meanings behind signs are decoded without delay. After processing, symbols turn into written words one by one. At the end, sentences appear matching what was signed just moments before. Starting off, the suggested tool changes written words into sign movements, shown either through real gestures or animated versions. Instead, it leans on advanced neural networks along with Media Pipe to track key body points extraction, and OpenCV for video processing. Watching a video stream comes first. A system grabs each frame one by one. It uses Media Pipe to map out where fingers and hands sit in three dimensions. These mapped positions move forward - passed along like notes between players in a quiet game. A trained network waits, ready to interpret what those shapes mean. Each gesture gets compared against real Indian Sign Language forms. When someone signs full phrases, not just single words, the tool shifts gears. Sequences matter now. Timing matters more. So it leans on structures like LSTMs or transformers - they catch rhythm, order, flow - the way motions unfold across moments. Meaning hides in movement patterns. This setup tries to find it. Now imagine turning spoken words into sign movements step by step, so people who hear can follow along without learning ISL first. Waiting around for outside helpers fades away since this system works as things happen. When both deaf and hearing individuals meet without shared signs, conversation flows more like everyday talk. Fancy artificial intelligence runs on simple hardware, showing how clever design opens doors. Tools built

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14, Issue 06, ACSCON - 2026

this way rethink access, making space for those often left out. We have designed the system to be highly accurate while, at the same time, developing a real-time response.

1.1 PROBLEM STATEMENT

Indian Sign Language functions as an essential means of communication for millions of hearing-impaired individuals in India. However, many hearing persons do not know ISL, creating an enormous gap in communication their education, health-care systems, workplaces, and daily life. Given all the limitations, all current systems of sign language translation struggle with one or more of the following reasons: they either only translate the underlying static representation of an individual letter of an alphabet or they use a wearable sensor to capture the sign language movement gestures or they work by interpreting the gestures in real time but only continuous gestures (not isolations signs) or they usually translate in one screen direction only to or from sign to text or sign linking contextual text with interpretive translation in either direction. The reality is that there is no effective, accurate, real time, bidirectional ISL translation system that works through live video input to effectively and accurately identify isolated signs and at same time/real time gestures at the level of sentence that includes continuous gestures. Variance of individual signing differently, and available datasets, brings additional limitations to systems. Compounding these limitation is that ISL gestures are all multilayered, complicated, hand and finger components that indicate meaning in simultaneous -given the time ordered features of gesture act- sequences. Lingering all of this is that sign language translation systems create mainly negative experiences for the hearing impaired, including miscommunication, social isolation and accessiblity to education and work.A clear need exists for a system that accurately detects ISL gestures from gestures received as video input, transforms gesture data sources into meaningful text, and produces the text as ISL gestures in real time through the animation of ISL signs and animation generated through the use of AI and ML. Such a system will help reduce barriers to communication and allow people and communities hearing and hearing impaired to interact inclusively.

A. System Overview The general workflow of this system is to capture live video input from a webcam. Video frames from this webcam are processed to extract hand landmark coordinates using MediaPipe hand tracking framework. These landmark features are normalized and transformed to feature vectors. These features are processed by a trained neural network model to classify the gesture into a predefined category. The results of this classification are shown in real time using a web application developed using Flask. Translation history is stored in an SQLite database. This methodology ensures accurate gesture recognition with real-time system performance. The first step of the workflow of this system is to capture live video input from a webcam. Video frames from this webcam are processed to extract hand landmark coordinates using MediaPipe hand tracking framework. These landmark features are normalized and transformed to feature vectors. These features are processed by a trained neural network model to classify the gesture into a predefined category. In order to view the anticipated output in real time, a web application using the Flask framework is employed, and a SQLite database is used for storing translation history. This approach ensures real-time speed along with accurate identification of gestures.

B. Obtaining Video The first step is to use the camera to record the video input. This can be done by using OpenCV to access the camera and continually capture frames. Each frame is then converted to RGB format to ensure compatibility with the MediaPipe architecture. Let's use the following to represent the captured frame: Frame = {P(x, y)} where P(x, y) stands for the frame's pixel coordinates.

C. Using MediaPipe for Hand Detection and Landmark Extraction. Hands are detected and hand landmarks are extracted using the MediaPipe Hands framework. Each of the 21 landmark points that MediaPipe offers is made up of three-dimensional coordinates (x, y, z). Every landmark point is shown as follows: $L_i = (x_i, y_i, z_i)$ in where i = 1, 2,..., 21 Given that two hands are supported by the system, the total number of landmarks is: 42 landmarks total (21 × 2). Three coordinates are contained in each landmark, yielding: A total of 126 characteristics (42 × 3) These characteristics offer spatial details on hand orientation and finger locations.

D. Vector Formation and Feature Normalization The retrieved landmark coordinates are normalized with respect to the wrist position in order to increase classification accuracy and guarantee invariance to hand position and scale. Assume that the wrist coordinate is: $W = (x_0, y_0, z_0)$ Every landmark is accepted as: $y_i' = y_i - y_0$ $z_i' = z_i - z_0$ $x_i' = x_i - x_0$ The vector of normalized landmark features is shown as: F is equal to $[x_1', y_1', z_1', x_2', y_2', z_2', ..., x_{42}', y_{42}', z_{42}']$ The neural network model receives this feature vector as input. Media Pipe can be used for real-time applications because the model can withstand

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14, Issue 06, ACSCON - 2026

changes in lighting and background noise.
E. Classification of Gestures Using Neural Networks To carry out this step, a feedforward neural network is applied using Tensor Flow and Keras. The classification of the class of interest is determined using this neural network based on the normalized landmark feature vector. Layers of the Neural Network Layer of Input This layer has 126 neurons, which is equal to 126 hallmark characteristics. Layers That Are Hidden Layers are dense and completely coupled. Activation function is ReLU. Layer of Output This layer has 36 neurons corresponding to digits 0-9 and alphabets A-Z. Activation function is Soft max. Definition of classification function: C = argmax (f(F)) where f is the function of the neural network and F is the feature vector. C is the class of interest. Probability is determined by the Soft max function as: $e^{(z_i)} / \Sigma e^{(z_j)} = P(C_i)$ The predicted gesture is chosen from the class with the highest probability.
F. Output Display and Real-Time Prediction Real-time gesture recognition is made possible by integrating the learned neural network model into a FlaskBefore you begin to format your paper, first write and save the content as a separate text file. Keep your text and graphic files separate until after the text has been formatted and styled. Do not use hard tabs, and limit use of hardreturns to only one return at the end of a paragraph. Do not add any kind of pagination anywhere in the paper. Do not number text heads-the template will do that for you. web application. The system constantly records video frames, extracts characteristics, and classifies gestures. The online interface shows the anticipated gesture in real time. Additionally, the system enables users to create words and sentences from gestures that are identified. The Flask framework offers:
• Streaming videos in real time
• The display of gesture prediction
• Management of translation history As additional gestures are identified, the anticipated output is continuously updated.
G. Storage of Translation History The system stores translated text and identified motions in a SQLite database. This allows users to go over motions and texts they have already recognized. Every entry in the database contains:
• Predicting gestures
• Text that has been translated
• The time stamp This feature makes it easier to use and makes it possible to follow translation history.
H. Optimizing Performance in Real Time Prediction is done at predetermined intervals rather than processing each frame to guarantee effective performance. Without compromising accuracy, this lowers the computing strain and enhances system responsiveness. The system accomplishes:
• Minimal latency
• Accurate recognition
• Performance in real time

1.2 GOAL OF THE PROJECT

The main aim of this project is to create a real-time two-way Indian Sign Language translation system, where an impaired person is able to communicate effectively with someone not impaired. This will be done by converting hand gestures to text through computer vision and machine learning. In addition, this project will create an animation to be used in translating the text to the appropriate sign. This project aims to achieve several goals, including the creation of tools that have high repeatability, low latency, and are easy to use to ensure access, inclusivity, and communication among impaired and non impaired persons.

2. LITERATURE SURVEY

1. Real-Time Sign Language Translation using Deep CNN-RNN: In this paper, a hybrid deep learning and computer vision model is developed using a convolutional neural network to extract visual features from sign language video frames, and a recurrent neural network to model sign gesture sequences. This model can be used in realtime with low latency for sign language translation.

2. Sign 2 Speech: End-to-End Translation using Transformers : This research proposes an end-to-end solution for translating between sign languages and spoken languages using transformers.

3. Integrating Multimodal Fusion to Improve Sign Language Recognition :This article explores a method of integrating RGB video and corresponding stereo depth maps for the creation of a multimodal system

4. Understanding Hand-Body Postural Interfaces through GNNs: This paper describes a way of modelling hand and body postural interfaces (i.e., gestures) through a Graph Neural Network (GNN-based) approach. A GNN method facilitates the representation of hand and body joints as nodes of a graph, thereby providing an efficient means of describing the dynamic behavior of gestures and the structural movement of gestures.

5. Incremental Learning for Adaptive Sign Language Translation: The current research includes the idea of incremental or "online" learning so that once the model is created, it is updated incrementally as it is

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14, Issue 06, ACSCON - 2026

exposed to new types of and styles of sign language. This is done without having to "retrain" the model.

3. OBJECTIVES

The main aim of the project is to design and implement a two-way translation system in real-time that is able to translate gestures in Indian Sign Language into written text and vice versa with regards to Indian Sign Language and text translation. In particular, through the use of deep learning methodologies and powerful AI and ML methodologies. The particular aims of the project are in line with the overall aim of this project. They are:

1. Development of a Real time ISL Gesture Recognition System: The project will develop a vision-based system using OpenCV and Media Pipe that will allow "live" video based input, hand detection, and will accurately identify 3D landmarks in real time.
2. Development of a Robust Machine Model to Translate ISL to Text: The project will train a series of deep learning models, that will recognize both single and continuous sign gestures from landmark data providing accurate and timely translation, after a training phase.
3. Development of Sequential Models to Recognise Continuous Sentences: The project aims to train sequential models like LSTMs or Transformer encoder-decoder architectures that learn the temporal dependencies between sequential frames of ISL and provide accurate meaningful translations of continuous sentences from the ISL gestures.
4. Development of the Text-to-ISL Translation Module: The project will develop a pair-set that accepts text input and returns the ISL gesture options using visualisation or animations that will allow true two-way communication between Deaf and hearing persons.
5. Prioritize considerations: accuracy together with real-time Research optimized and improved architectures of neural networks, if using machine learning and the list of words is reliable and there are no specific signs, if with applicable research project reduce work with close to “real time” translate - with accuracy that is designed to be 95% or higher if the signs are alone and with great BLEU / accuracy sentences.
6. Create a user-involvement / lower cost interface: design a GUI interface so the user can interact and switch between ISL-to-Text and Text-to-ISL.

This will increase usability and be productive for education, workplace environments, health care, entertainment and day-to-day conversations.

4. METHODOLOGY

A. Experimental Configuration

A computer with a webcam and normal computer processing ability was utilized to test this system. The acquired dataset of landmarks, which contained normalized coordinates of hand landmarks retrieved using the Media Pipe framework, was utilized to train the neural network model. The real-time application based on Flask was utilized to integrate it with the model, which is in.h5 format. Live gestures were made in front of the webcam during testing in various settings, such as lighting levels, Equitable representation of the dataset. Model performance was much enhanced and computational complexity was decreased by using landmark-based features rather than raw images. The accuracy of classification is computed as follows:

(Number of Correct Predictions / Total Predictions) x 100 is the accuracy.

The model was effective in learning the spatial relationships between the hand landmarks, as shown by the high accuracy achieved.

B. Performance of Real-Time Gesture Recognition

During the testing phase, the system was found to offer effective performance with regard to real-time recognition of gestures. To ensure the recognition of gestures, the system processed video frames, which were continuously obtained at specific intervals of time from the webcam. The smooth performance of the system was ensured with the low average time required for the prediction of each frame. Moreover, the system did not show any lag during the recognition of gestures. The performance of the system in real time was ensured with the following factors:

- Effective landmark detection with the use of MediaPipe
- Lightweight neural network model
- Better feature extraction method
- Constant time interval for predictions The performance of the system during its continuous operation remained the same.

C. Identifying Numbers and Alphabets

The model was able to identify all supported gesture classes with high accuracy. The letters A through Z 0-9 are the numbers. This is due to the fact that all the gestures were well categorized based on the position of hand landmarks. This was observed in all test runs of the model. The model was able to demonstrate its high capacity to distinguish between various gestures

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14, Issue 06, ACSCON - 2026

due to the use of landmark coordinates.

D. Sturdiness and Dependability

To test its robustness, the system was subjected to various environmental settings. The system functioned dependably under:

• Typical lighting conditions within
• Various hand orientations
• Moderate changes in the background
• Varying separations from the camera The reliable landmark identification of MediaPipe greatly increased system dependability. Nevertheless, performance somewhat declined under:
• The lighting is quite dim.
• A partial blockage of the hand
• Too much motion blur The system continued to operate at respectable levels in spite of these circumstances.

E. User Interface Performance and Translation

Real-time gesture predictions were successfully exhibited via the Flask-based web interface. Users could create words and sentences and observe detected movements using the interface. Additionally, the system used a SQLite database to maintain translation history, giving usershand orientation, and distance of the camera. The following criteria were utilized to test this system:

• Accuracy of gesture classification
• Speed of prediction in real-time
• Stability of this system

F. Accuracy of Gesture Classification.

When the trained neural network model is subjected to real-time gesture input, the model exhibited high classification accuracy. The system's overall accuracy in identifying hand movements that match to letters and numbers was about 98%. The following factors contributed to the excellent accuracy: Using Media Pipe for efficient landmark feature extraction.

G. Analysis of Confusion Matrix

The majority of gesture classes were accurately categorized, according to the confusion matrix analysis. Rarely did misclassification occur, and it mostly happened between gestures with strikingly similar hand forms. The total classification result validates the landmark-based neural network approach's efficacy.

H. Evaluation Against Current Systems

The suggested system offers certain benefits over conventional image-based gesture recognition systems:

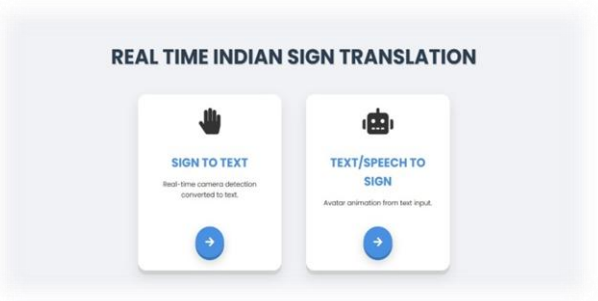
• Increased precision
• A quicker rate of processing
• Reduced processing demands
• Better performance in real time Efficiency is increased and superfluous image processing is decreased with landmark-based feature extraction.

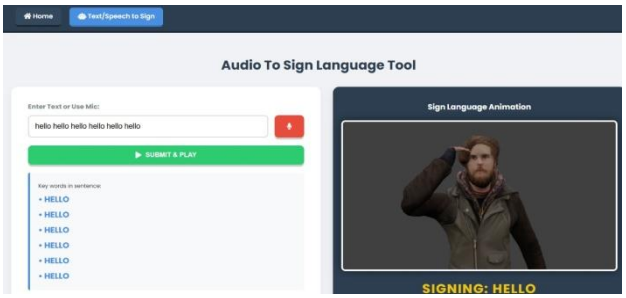
I. Results Synopsis

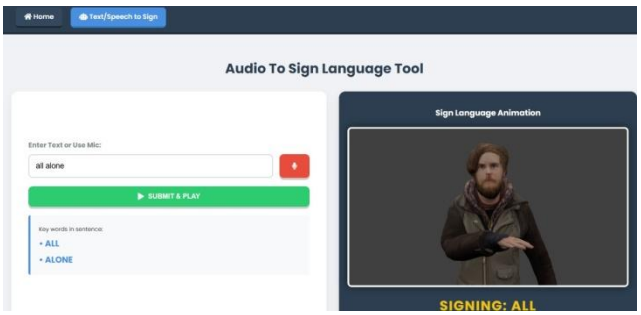
The outcomes of the trial show that the suggested real time sign language translation system offers:

• Accurate gesture recognition (~98%)
• Effective performance in real time
• Accurate classification of gestures
• Support for assistive communication that works System performance is greatly enhanced by the combination of MediaPipe and neural network-based classification.

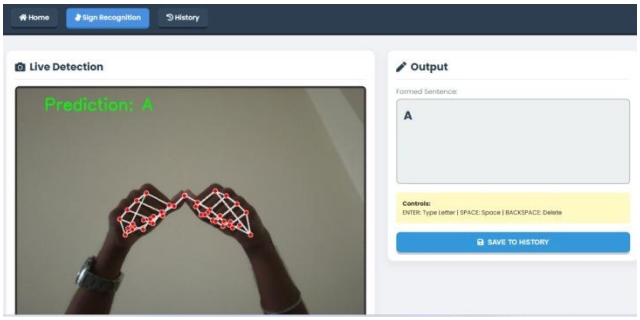
5. RESULT

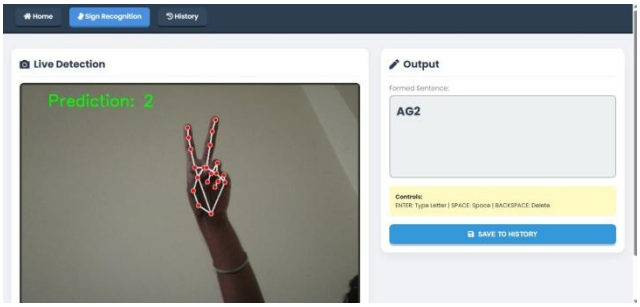






International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14, Issue 06, ACSCON - 2026





6. CONCLUSION

In order to facilitate communication between hearing impaired people and the general public, this article proposed a real-time sign language translation system that makes use of computer vision and deep learning techniques. The system used a neural network model trained with TensorFlow and Keras for gesture categorization, and MediaPipe for precise hand landmark recognition. The system effectively detected 36 gesture classes, including the alphabets (A–Z) and digits (0–9), with high accuracy and dependable real-time performance by extracting and processing 3D landmark coordinates. Continuous video capture was made possible by the integration of OpenCV, and an interactive platform for storing translation history using a SQLite database and showing translation results was offered via the Flask-based web interface. The suggested system showed excellent identification accuracy, low latency, and reliable performance under a range of environmental situations, according to experimental data. Whereas mammography is dedicated to detecting breast tumours in their earliest stages which is highly sensitive, and classify it as benign or malignant. Mammography allows for high-resolution imaging of the internal architecture of the breast and aids in the detection of worrisome lesions. It is a fundamental strategy to reduce the death rate of breast cancer patients in MGs. Furthermore, Mammography screenings may reduce the likelihood of cancer death rate in patients by 39-49%. Compared to conventional image-based techniques, the landmark-based method improved classification efficiency while lowering computational complexity. The suggested system provides an accessible, scalable, and affordable assisted communication solution. Expanding the system to allow word-level and sentence-level recognition, strengthening its resilience in difficult situations, and implementing it on mobile and embedded platforms to improve usability and accessibility in practical applications are some potential future improvements.

7. REFERENCES

[1] Real-Time Sign Language Translation Using Deep CNN RNN, IEEE Transactions on Intelligent Systems, 2024.
[2] Sign2Speech: End-to-End Sign Language to Speech Translation Using Transformers, ACM Transactions on Multimedia Computing, Communications, and Applications, 2024.
[3] Integrating Multimodal Fusion to Improve Sign Language Recognition, Springer Journal of Computer Vision and Pattern Recognition, 2024.
[4] Understanding Hand–Body Postural Interfaces Through Graph Neural Networks, IEEE Transactions on Neural Networks and Learning Systems, 2025.
[5] Incremental Learning for Adaptive Sign Language Translation, ACM International Conference on Intelligent User Interfaces, 2024.
[6] Fusion of Visual and Sensor Data for Enhanced Sign Language Translation, Springer Journal of Multimodal Interaction, 2025.
[7] A Real-Time Explainable AI Framework for Sign Language Translation, IEEE Access, 2025.
[8] Domain Adaptation Through Transfer Learning for Sign Language Translation, ACM Transactions on Accessible Computing, 2025.
[9] Time-Series Analysis for Capturing Gesture Dynamics Over Time in Sign Language Translation, IEEE Transactions on Affective Computing, 2025.
[10] Federated Learning for Privacy-Preserving Sign Language Translation Models, Springer Journal of Artificial Intelligence and Privacy, 2025.