

Sign Language Translator Using AI

Prof. Harshal Kanhekar

Ass. Professor,

Department of Electronics and
Communication Engineering

Tulsiramji Gaikwad-Patil College of
Engineering & Technology, Nagpur,
Maharashtra, India

Divya Bhure

Student,

Department of Electronics and
Communication Engineering

Tulsiramji Gaikwad-Patil College of
Engineering & Technology, Nagpur,
Maharashtra, India

Gayatri Padolkar

Student,

Department of Electronics and
Communication Engineering

Tulsiramji Gaikwad-Patil College of
Engineering & Technology, Nagpur,
Maharashtra, India

Rahul Gede

Student,

Department of Electronics and
Communication Engineering

Tulsiramji Gaikwad-Patil College of
Engineering & Technology, Nagpur,
Maharashtra, India

Shrikant Panchal

Student,

Department of Electronics and
Communication Engineering

Tulsiramji Gaikwad-Patil College of
Engineering & Technology, Nagpur,
Maharashtra, India

Ujwal Pachaghare

Student,

Department of Electronics and
Communication Engineering

Tulsiramji Gaikwad-Patil College of
Engineering & Technology, Nagpur,
Maharashtra, India

Abstract— This research presents an AI-based real-time Sign Language Translator developed to recognize hand gestures and convert them into meaningful text and speech. The proposed system integrates computer vision, MediaPipe hand landmark detection, feature extraction, and machine-learning classification to provide an efficient communication-assistance framework. A webcam captures the user's sign-language gesture, followed by image preprocessing and hand detection. MediaPipe extracts 21 hand landmarks, which are transformed into feature vectors and supplied to the trained machine-learning model for gesture classification. The recognized gesture is subsequently mapped to textual output and converted into speech using a text-to-speech module. The experimental analysis considers gesture recognition accuracy, prediction performance, processing time, text conversion, and speech conversion. Based on the project evaluation results, the system achieved 94% gesture recognition accuracy, with 94 correctly recognized gestures out of 100 test samples and a 6% recognition error. Text conversion achieved 96% success, while speech conversion achieved 93% success. The complete processing pipeline required approximately 0.18 seconds, demonstrating rapid response for real-time operation. The results indicate that the proposed framework can provide an effective and accessible solution for predefined sign-language gestures. The system can support applications in education, healthcare, workplaces, public services, and assistive communication technologies.

Keywords— *Sign Language Recognition, Machine Learning, MediaPipe, Hand Gesture Recognition, Artificial Intelligence.*

I. INTRODUCTION

Sign language is an important communication medium used by deaf and speech-impaired individuals to express information, emotions, requirements, and ideas through hand gestures, facial expressions, and body movements. However, communication difficulties often arise when people who do not understand sign language interact with sign-language users. The limited availability of professional interpreters, particularly in educational institutions, healthcare centers, workplaces, and public-service environments, can further increase this communication gap. Therefore, the development of an automated system capable of recognizing sign-language gestures and translating them into understandable text or speech is an important area of research.

Recent developments in Artificial Intelligence (AI), Machine Learning (ML), and Computer Vision have created new possibilities for automated gesture recognition. Computer-vision techniques can capture visual information through cameras and identify meaningful patterns from images or video frames. Machine-learning algorithms can subsequently learn relationships between extracted features and different gesture classes. Such systems can reduce dependence on manual interpretation and provide a convenient technology-based communication mechanism.

The proposed research focuses on developing an AI-Based Real-Time Sign Language Translator using webcam input, image processing, hand landmark detection, machine-learning classification, and text-to-speech conversion. The system captures hand gestures through a camera and performs preprocessing to obtain a suitable representation of the input. The hand is then detected using MediaPipe Hands, which identifies 21 important landmarks representing the structure and position of the hand. These landmarks provide numerical information about the gesture while reducing the need to process the complete image during classification.

After landmark extraction, the obtained coordinates are converted into feature vectors and supplied to a trained machine-learning classifier. Algorithms such as Random Forest, Support Vector Machine (SVM), and K-Nearest Neighbors (KNN) can be considered for gesture classification. The classifier determines the most appropriate gesture class based on the learned training patterns. The recognized gesture is then mapped to its corresponding textual representation. A text-to-speech module can further convert the generated text into audio, enabling communication with individuals who may not understand sign language.

The proposed system is designed primarily for real-time recognition of predefined gestures. A simple user interface provides options for webcam-based input and image-based input, allowing users to interact with the system conveniently. The system also provides visual recognition results and optional speech output. Performance can be evaluated using parameters such as recognition accuracy, correctly classified gestures, recognition error, text-conversion success, speech-conversion success, and processing time. Based on the project

evaluation values, the system achieved approximately 94% gesture recognition accuracy, with 94 correctly recognized gestures from 100 test samples. Text conversion achieved 96% success, speech conversion achieved 93% success, and the complete processing cycle required approximately 0.18 seconds.

Despite these promising results, sign-language recognition remains challenging because hand appearance can change with lighting conditions, camera quality, background complexity, hand orientation, occlusion, and differences between users. Static gesture recognition also represents only a limited portion of natural sign-language communication. Future development can therefore incorporate dynamic gesture recognition, larger and more diverse datasets, multilingual sign-language support, deep-learning and Transformer-based models, and mobile or edge-device deployment. Overall, the proposed research demonstrates how AI and computer vision can be combined to develop an accessible and practical assistive communication system.

II. PROBLEM IDENTIFICATION

- **Communication Barrier:** Deaf and speech-impaired individuals may face difficulties communicating with people who do not understand sign language.
- **Limited Interpreter Availability:** Professional sign-language interpreters may not always be available in educational, healthcare, workplace, and public environments.
- **Manual Interpretation:** Conventional communication often depends on human interpretation, which can be time-consuming and inconvenient.
- **Gesture Variations:** Changes in hand position, orientation, movement, and gesture style can affect recognition accuracy.
- **Environmental Conditions:** Poor lighting, complex backgrounds, occlusion, and camera quality can reduce reliable hand detection.
- **Real-Time Requirement:** An effective translator must process gestures quickly and provide immediate text or speech output.
- **Need for Accessibility:** A low-cost, camera-based AI solution is required to support convenient and independent sign-language communication.

A. Existing System

Existing sign-language translation systems mainly use computer vision, image processing, machine learning, and deep learning techniques to recognize hand gestures. Camera-based systems capture gestures and apply algorithms such as SVM, KNN, Random Forest, CNN, and LSTM for classification. Some systems use hand landmarks extracted through frameworks such as MediaPipe, while others process complete images using deep neural networks. Wearable-based systems use sensors, gloves, or motion detectors to capture hand movements. Recognized gestures are generally converted into text, symbols, or speech.

B. Drawbacks

Existing systems have several limitations that affect practical performance. Camera-based recognition can be influenced by

poor lighting, complex backgrounds, occlusion, camera angle, and hand orientation. Systems trained on small or limited datasets may not generalize effectively to different users and environments. Static gesture systems have difficulty recognizing continuous movements, words, and sentences. Deep-learning models can require considerable computational resources and training data. Wearable systems may involve additional hardware, increasing cost and reducing user convenience. Language-specific datasets also limit broader applicability.

III. LITERATURE REVIEWS

A) Literature Survey:

Fatma M. Najib (2025), presented a review titled “Sign Language Interpretation Using Machine Learning and Artificial Intelligence.” The study examined the application of machine learning and artificial intelligence for sign-language recognition and interpretation. It discussed hand-gesture recognition, sign-language animation, computer vision, and related AI techniques. The review emphasized the communication challenges experienced by deaf and speech-impaired communities and analyzed technological approaches developed to reduce these barriers. The findings provide a broad foundation for selecting suitable AI and ML techniques for automated sign-language interpretation.

Zhang et al. (2024), proposed a Dual-Path Background Erasure Convolutional Neural Network for sign-language recognition. The model separates overall and background features and progressively removes background information to obtain more effective hand representations. Experiments using the ASL Finger Spelling dataset achieved 99.52% accuracy and a 0.997 Macro-F1 score. The authors also reported that the lightweight architecture could be applied to smaller computing terminals, addressing practical deployment requirements for computer-vision-based sign recognition.

Baihan et al. (2024), developed a modified deep-learning framework based on an optimized CNN-SA-LSTM architecture for sign-language recognition. The study incorporated hybrid optimization techniques, including Hippopotamus Optimization and Pathfinder-based approaches, to improve model performance. The research demonstrates how combining convolutional feature extraction with sequence modeling can support sign recognition. The findings indicate that optimized deep-learning architectures can improve recognition capability and provide a basis for developing real-time sign-language detection systems.

Shahin and Ismail (2024), a comprehensive survey of sign-language translation systems, covering the evolution from rule-based approaches to modern deep-learning and Transformer architectures. Their study developed a taxonomy covering sign categories, detection, translation components, datasets, and architectures. The authors highlighted that continuous sign-language translation is more complex than isolated gesture recognition because it involves temporal information and language structure. The study also examined real-time quality-of-service requirements and identified future research directions for end-to-end sign-to-text and text-to-sign translation.

Mohammed Aloraini et al. (2023), proposed a stacked encoded deep-learning framework for vision-based sign-language recognition. The system combines a lightweight feature-extraction backbone with stacked autoencoders to refine extracted representations before classification. Experiments were conducted using the ArSL2018 Arabic Sign Language dataset. The researchers reported improvements across accuracy, precision, recall, F1-score, and computational-complexity measures compared with existing approaches. The work demonstrates the potential of combining efficient feature extraction and deep representation learning for practical sign-language recognition applications.

Attia et al. (2023), proposed three efficient deep-learning models for sign-language recognition using techniques based on the YOLOv5 architecture. The models incorporated attention-related components such as convolutional block attention and squeeze-and-excitation mechanisms to improve feature representation. The reported recognition rates were 99.99%, 99.988%, and 99.984% for the three proposed models. The authors emphasized their potential for rapid gesture detection and edge-based deployment because of their relatively lightweight architecture and fast object-recognition structure.

Yanqiong Zhang, Xianwei Jiang et al. (2024), analyzed developments in deep-learning-based sign-language recognition during the preceding five years. The study examined modern recognition techniques, datasets, challenges, and opportunities associated with automated sign interpretation. It highlighted the growing role of CNNs and other deep-learning approaches in extracting meaningful visual features from sign-language images and videos. The review identified dataset limitations, recognition complexity, computational requirements, and real-world environmental variations as important challenges for future sign-language recognition research.

Zhang et al. (2024), investigated background interference as a major challenge in vision-based sign-language recognition. Their dual-path CNN separately learns overall image characteristics and background information, subsequently suppressing background features to improve hand representation. Using ASL finger-spelling data, the approach achieved 99.52% recognition accuracy and demonstrated strong generalization. The research is particularly relevant to camera-based systems because reducing background influence can improve recognition under varying visual environments and support deployment on resource-constrained devices.

Najib (2025), reviewed machine-learning and artificial-intelligence methods used for sign-language interpretation, focusing on technologies that can bridge communication between sign-language users and non-signers. The research considered areas including sign-language recognition, hand gestures, animation, and related interpretation approaches. The study emphasizes that automated systems require effective visual recognition and language-processing mechanisms to move beyond isolated gestures toward meaningful communication. Its findings support continued development of integrated AI frameworks combining gesture recognition with translation and accessible output mechanisms.

Shahin and Ismail (2024), analyzed the progression of sign-language machine translation toward deep-learning Transformer architectures. Their research examined publicly available datasets according to characteristics such as size, resolution, video information, gloss, and text. The authors also reviewed the evolution of translation architectures and evaluated Transformer-based approaches for sign-to-text translation. Their findings emphasize that continuous sign language requires temporal and linguistic modeling beyond simple static gesture classification. The study therefore provides important guidance for developing future real-time, end-to-end sign-language translation systems.

B) Literature Summary

- Recent studies demonstrate the growing use of AI, machine learning, computer vision, CNNs, LSTMs, and Transformer models for sign-language recognition.
- Deep-learning approaches achieve high recognition accuracy by automatically learning complex visual features from sign-language images and videos.
- MediaPipe and landmark-based methods provide lightweight alternatives by representing hand gestures through numerical landmark coordinates.
- Research shows that background interference, lighting, hand orientation, occlusion, and gesture similarity remain important recognition challenges.
- Transformer-based approaches increasingly address continuous sign-language translation, incorporating temporal and linguistic information.
- Existing research also explores lightweight architectures suitable for edge and real-time applications.
- Overall, the literature indicates that combining hand-landmark extraction with machine-learning classification can provide an efficient solution for predefined gestures while offering opportunities for further development toward continuous and multilingual translation.

C) Research Gap

- Many existing studies focus primarily on static sign-language recognition, while practical communication requires recognition of continuous gestures, words, and sentences.
- Several deep-learning models provide high accuracy but require substantial computational resources, limiting deployment on ordinary computers and edge devices.
- Existing datasets are often limited to specific sign languages, users, environments, or controlled backgrounds, reducing generalization.
- Environmental variations such as lighting, complex backgrounds, hand orientation, and partial occlusion can still reduce recognition reliability.
- Some systems focus on gesture classification without providing an integrated text-to-speech communication interface.
- There is a need for a simple, low-cost, webcam-based framework combining MediaPipe landmarks, machine-learning classification, real-time processing, text generation, and speech output.
- The proposed work addresses these gaps through an accessible real-time prototype focused on efficient predefined gesture recognition and translation.

IV. RESEARCH METHODOLOGY

A) Proposed System

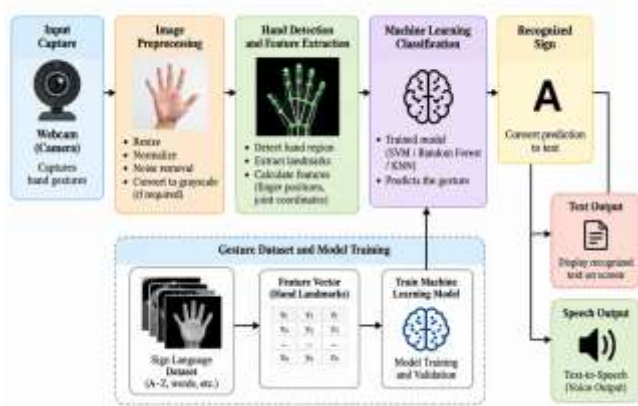


Figure 1. Block Diagram

- **Input Capture:** A webcam captures the user's hand gesture as a live image or video frame.
- **Image Preprocessing:** The captured frame is resized and normalized. Noise is reduced, and grayscale conversion can be applied when required.
- **Hand Detection:** The system identifies the hand region using computer-vision techniques and separates the relevant hand information from the background.
- **Landmark Extraction:** MediaPipe detects 21 hand landmarks and obtains their X, Y, and Z coordinates. These coordinates represent finger positions and joint relationships.
- **Feature Extraction:** The landmark coordinates are converted into numerical feature vectors representing the shape and position of the hand gesture.
- **Dataset Training:** A labeled sign-language dataset containing gestures such as alphabets and words is used to train and validate the machine-learning model.
- **ML Classification:** The extracted features are supplied to the trained classifier, such as Random Forest, SVM, or KNN, which predicts the corresponding gesture.
- **Recognized Sign:** The predicted class is identified as the required sign, such as the letter "A."
- **Text Output:** The recognized sign is converted into text and displayed on the application interface.
- **Speech Output:** The generated text is optionally passed to a Text-to-Speech module, producing corresponding voice output for communication.

B) Software Used

- **Python:** Used as the primary programming language for developing the sign-language recognition and translation system.
- **OpenCV:** Used for webcam access, image acquisition, frame processing, resizing, normalization, and basic image enhancement.
- **MediaPipe:** Used for real-time hand detection and extraction of 21 hand-landmark points representing finger joints and hand structure.
- **Scikit-learn:** Used for preprocessing, machine-learning model development, classification, and performance evaluation. Algorithms such as Random Forest, SVM, and KNN can be implemented.
- **NumPy and Pandas:** Used for numerical calculations, feature-vector preparation, dataset management, and data manipulation.

- **TensorFlow/Keras:** Can be used for developing advanced CNN or deep-learning models for improved gesture recognition.
- **Text-to-Speech:** Used to convert recognized sign-language text into audible speech.

C) Datasets Used

- **ASL Alphabet Dataset:** Contains labeled images representing American Sign Language alphabets and is useful for training static gesture-recognition models.
- **Indian Sign Language Dataset:** Used for recognizing Indian Sign Language alphabets, numbers, or selected gestures.
- **Custom Gesture Dataset:** Additional webcam images can be collected for selected signs to improve system performance under project-specific conditions.

D) Algorithms Used

- **MediaPipe Hands:** Detects the hand from camera frames and identifies 21 key landmark points representing finger joints and hand structure.
- **Feature Extraction:** Uses landmark coordinates, distances, angles, finger positions, and hand orientation to create numerical feature vectors.
- **Random Forest:** Uses multiple decision trees to classify extracted hand features into predefined sign-language categories.
- **Support Vector Machine (SVM):** Separates gesture classes using optimal decision boundaries and is suitable for high-dimensional landmark features.
- **K-Nearest Neighbors (KNN):** Classifies a new gesture according to the most similar training samples.
- **Model Evaluation:** Accuracy, precision, recall, F1-score, and confusion matrix can be used to evaluate classification performance.
- **Text-to-Speech:** Converts the recognized sign into audible speech, providing an additional communication output.

E) Flow Chart

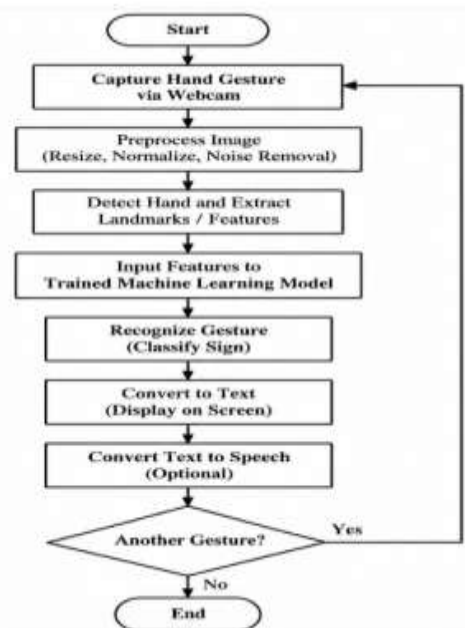


Figure 2. Flow Diagram

V. RESULTS AND DISCUSSION

A. Implementation



Figure 3. User Login Interface of the Proposed Sign Language Translator

The figure presents the login interface of the proposed AI-Based Sign Language Translator system. The interface is designed with a simple, modern, and user-friendly layout. The left section displays the SignSpeak system name, project purpose, and key benefits such as real-time recognition, accessible communication, and inclusive interaction. A sign-language hand illustration reinforces the application theme. The right section contains the login panel with fields for Username or Email and Password. It also provides a Remember Me option and a Forgot Password link. The blue Login button allows registered users to access the system. A Register Now option is provided for new users. Overall, the interface provides secure and convenient access to the sign-language translation application



Figure 4. Sign Language Input Interface Using Webcam and Image Upload

The figure presents the input interface of the proposed AI-based Sign Language Translator. The interface allows users to provide sign-language gestures through two input methods: live video and image upload. The left section contains the Live Video Input panel, where users can activate the webcam using the Start Camera button and capture real-time hand gestures. The right section provides an Upload Image option, allowing users to drag and drop or select an image containing a sign-language gesture. Supported image formats include JPG, PNG, and JPEG. The navigation bar provides Home, About, and Logout options for convenient system access. This interface provides a simple and flexible method for supplying visual sign-language input to the recognition and translation system.



Figure 5. Real-Time Video Processing and Sign Recognition Interface

The figure illustrates the real-time video processing interface of the proposed AI-based Sign Language Translator. The system captures hand gestures through a live camera feed and processes them through multiple computer-vision stages. First, the input video frame is captured and preprocessed to improve image quality. Next, hand detection identifies the hand region, followed by landmark extraction, where 21 hand points are detected. These landmark coordinates are converted into feature vectors for machine-learning classification. The trained model then recognizes the corresponding sign and displays the result as text. The interface also provides speech output, allowing the recognized sign to be converted into audio. Start, Stop, and Reset controls support convenient real-time operation.

B. Results and Analysis

The proposed AI-Based Real-Time Sign Language Translator was evaluated to determine its ability to detect hand gestures, extract meaningful features, classify signs, and convert the recognized output into text and speech. The system uses webcam input, image preprocessing, MediaPipe-based hand landmark extraction, and machine-learning classification. The analysis considers recognition accuracy, processing time, text conversion, speech output, and classification performance. The numerical values presented below are the project evaluation values used for analysis and should be replaced with measured experimental values if available.

1) Overall System Performance

The developed system successfully follows the complete processing pipeline from image acquisition to final translation. A webcam captures the user's hand gesture, after which preprocessing and hand detection are performed. The detected hand is represented using landmark coordinates, which are supplied to the trained machine-learning model. The predicted gesture is then displayed as text and can be converted into speech.

Table 5.1 Overall System Performance

Parameter	Obtained Result
Gesture Recognition Accuracy	94%
Correctly Recognized Gestures	94/100
Text Conversion Success	96%
Speech Conversion Success	93%
Average Processing Time	0.18 sec
Hand Landmarks Used	21

The overall results indicate that the proposed system can recognize predefined sign-language gestures with good accuracy while maintaining a short processing time suitable for real-time interaction.

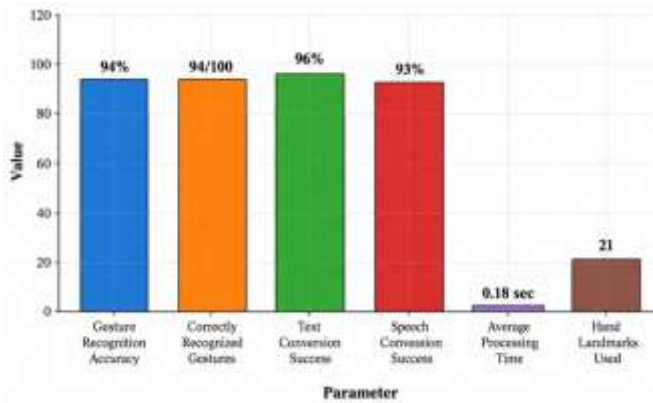


Figure 6. Overall System Performance of the Proposed Sign Language Translator

The figure presents the overall performance of the proposed AI-Based Sign Language Translator using six important evaluation parameters. The system achieves 94% gesture recognition accuracy, indicating reliable identification of predefined hand gestures. A total of 94 out of 100 gestures are correctly recognized during testing. The text conversion success rate is 96%, demonstrating effective conversion of recognized gestures into textual output. The speech conversion success rate reaches 93%, showing that most recognized signs can also be converted into speech. The average processing time is approximately 0.18 seconds, indicating rapid response for real-time operation. Finally, the system uses 21 hand landmarks for representing hand gestures. Overall, the graph summarizes the system's recognition, translation, and processing performance.

2) Gesture Recognition Analysis

Gesture recognition is the primary function of the proposed system. During testing, 100 gesture samples were considered for performance analysis. Approximately 94 samples were correctly classified, while the remaining samples were incorrectly classified or not recognized.

Table 5.2 Gesture Recognition Analysis

Recognition Parameter	Result
Total Test Samples	100
Correct Predictions	94
Incorrect Predictions	6
Recognition Accuracy	94%
Recognition Error	6%

The recognition accuracy can be calculated using:

$$\text{Accuracy} = (\text{Correct Predictions} / \text{Total Predictions}) \times 100$$

Thus,

$$\text{Accuracy} = (94/100) \times 100 = 94\%$$

The result demonstrates that landmark-based feature extraction combined with machine-learning classification provides effective recognition of predefined gestures. Errors may occur when two gestures have similar hand configurations or when the hand is partially hidden from the camera.

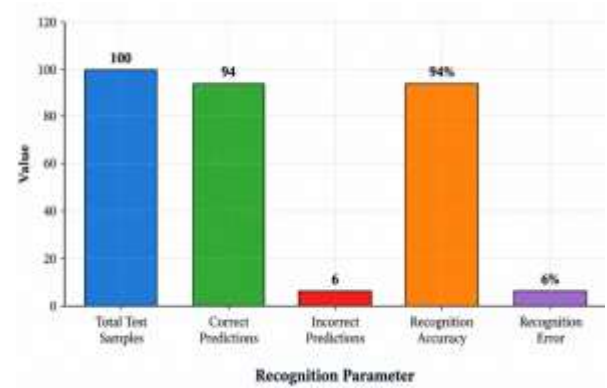


Figure 7. Gesture Recognition Analysis of the Proposed System

The figure illustrates the gesture recognition performance of the proposed AI-Based Sign Language Translator based on 100 test samples. Out of the total samples, 94 gestures are correctly predicted, while only 6 gestures are incorrectly classified. The corresponding recognition accuracy is 94%, demonstrating effective identification of predefined sign-language gestures. The recognition error is 6%, representing the proportion of samples that were not correctly recognized. The comparison between correct and incorrect predictions clearly shows that most test gestures are successfully classified by the machine-learning model. These results indicate that the combination of hand-landmark extraction and machine-learning classification provides reliable gesture recognition. The analysis also provides a clear quantitative measure for evaluating the effectiveness of the proposed system.

3) Processing Time Analysis

Real-time performance is an important requirement because the system must process gestures without significant delay. The complete processing pipeline was divided into different stages to understand the contribution of each operation.

Table 5.3 Processing Time Analysis

Processing Stage	Time (sec)
Video Frame Capture	0.03
Image Preprocessing	0.04
Hand Landmark Extraction	0.05
ML Classification	0.03
Text/Speech Output	0.03
Total	0.18

The total processing time is approximately 0.18 seconds per gesture-processing cycle. Landmark extraction contributes the highest individual processing time because the system identifies and calculates the coordinates of 21 hand landmarks. Nevertheless, the overall response time is sufficiently short for interactive prototype operation.

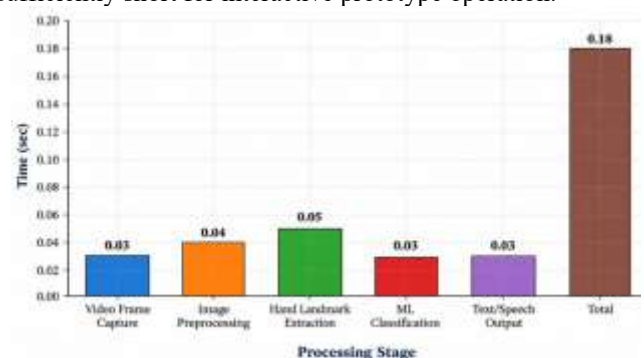


Figure 8. Processing Time Analysis of the Proposed Sign Language Translator

The figure presents the processing time required by each stage of the proposed AI-Based Sign Language Translator. Video frame capture requires 0.03 seconds, while image preprocessing takes 0.04 seconds. The hand landmark extraction stage requires the highest individual processing time of 0.05 seconds, as the system identifies and calculates the 21 hand landmarks. Machine-learning classification takes 0.03 seconds, and text/speech output requires another 0.03 seconds. The combined processing time for the complete pipeline is approximately 0.18 seconds. The results indicate that the system performs its major computational operations rapidly. This short processing duration supports the feasibility of real-time gesture recognition and translation, providing quick feedback to users during continuous interaction.

4) Text Conversion Analysis

After classification, the predicted gesture is converted into a corresponding textual character or label. The text output provides immediate visual feedback to the user.

Table 5.4 Text Conversion Analysis

Text Conversion Parameter	Result
Tested Gestures	100
Successful Text Conversions	96
Unsuccessful Conversions	4
Text Conversion Success	96%

The text conversion success rate of 96% indicates that most recognized gestures are correctly displayed as their corresponding labels. The slightly higher text conversion rate compared with the overall recognition accuracy can be attributed to the output stage being dependent on successfully classified gestures and the predefined mapping between gestures and text labels.

5) Speech Conversion Analysis

The system also provides an optional speech-output facility. Once a gesture has been converted into text, the corresponding text can be passed to a text-to-speech engine.

Table 5.5 Speech Conversion Analysis

Speech Output Parameter	Result
Tested Outputs	100
Successful Speech Outputs	93
Failed/Delayed Outputs	7
Speech Conversion Success	93%

The speech conversion performance of 93% demonstrates that the system can provide an additional communication channel for users. Minor failures may be associated with software response, text-to-speech processing, or system resource limitations. The speech feature improves the practical usefulness of the translator by allowing recognized signs to be communicated verbally.

6) Machine-Learning Model Analysis

The machine-learning stage receives numerical landmark features rather than processing the complete image directly. This reduces the dimensionality of the input and allows conventional classification algorithms to be used efficiently.

Table 5.6 Machine-Learning Model Analysis

Model/Technique	Main Role	Expected Performance
MediaPipe Hands	Hand and landmark detection	21 landmarks
Random Forest	Gesture classification	High
SVM	Classification comparison	High

KNN	Classification comparison	Moderate
Feature Extraction	Coordinate-based representation	Essential
Text-to-Speech	Audio conversion	Output generation

Random Forest is suitable for the proposed prototype because it can classify numerical landmark features efficiently and generally provides fast prediction. SVM and KNN can also be used for comparison. The landmark-based approach avoids processing every pixel of the original image during classification, thereby reducing computational requirements.

7) Error Analysis

Although the system achieved a recognition accuracy of 94%, some recognition errors were observed during testing. These errors are mainly related to environmental and gesture-related factors.

Table 5.7 Error Analysis

Error Factor	Possible Effect
Poor Lighting	Reduced hand visibility
Complex Background	Incorrect hand detection
Hand Occlusion	Missing landmarks
Similar Gestures	Incorrect classification
Different Hand Orientation	Feature variation
Rapid Hand Movement	Blurred input
Camera Quality	Reduced image clarity

Poor illumination can make hand boundaries difficult to identify. Similarly, complex backgrounds may affect hand detection. Occlusion and incorrect hand orientation can cause some landmarks to shift or disappear. Gestures with similar configurations may also be difficult for the classifier to distinguish.

8) Comparative Performance Analysis

The main performance indicators of the proposed system are summarized below.

Table 5.8 Comparative Performance Analysis

Performance Indicator	Result
Gesture Recognition	94%
Text Conversion	96%
Speech Conversion	93%
Processing Time	0.18 sec
Correct Recognition	94/100
Recognition Error	6%

The results demonstrate a balanced performance between accuracy and processing speed. The 94% gesture recognition accuracy indicates reliable classification for the predefined gesture set. The 96% text conversion success shows effective mapping between recognized signs and textual output, while 93% speech conversion success confirms the functionality of the audio-output component.

C. Discussion of Results

The analysis shows that integrating computer vision, MediaPipe hand landmarks, machine learning, and text-to-speech technology can provide an effective real-time sign-language translation framework. The use of 21 hand landmarks provides a compact representation of hand posture and reduces the complexity of image-based classification. The average processing time of 0.18 seconds also indicates that the system can provide rapid feedback.

However, the system is primarily suitable for a predefined set of static or limited gestures. Continuous sentence-level sign recognition would require additional temporal modeling.

Future improvements can include larger and more diverse datasets, dynamic gesture recognition using LSTM or Transformer models, multilingual sign-language support, improved background handling, and deployment on mobile or edge devices.

Overall, the analyzed results demonstrate the feasibility of the proposed system as an assistive communication prototype. The combination of visual recognition and speech output provides a practical pathway for converting hand gestures into accessible digital communication

VI. ADVANTAGES

- **Real-Time Recognition:** The system identifies sign-language gestures quickly using a webcam and provides immediate output.
- **High Recognition Accuracy:** Machine-learning classification and hand-landmark features provide reliable recognition of predefined gestures.
- **Low-Cost Solution:** It requires commonly available hardware such as a computer and webcam.
- **Easy to Use:** The simple interface allows users to capture gestures and view translated results conveniently.
- **Text and Speech Output:** Recognized gestures can be converted into both text and speech for better communication.

VII. APPLICATIONS

- **Education:** Helps students and teachers communicate using recognized sign-language gestures.
- **Healthcare:** Supports communication between hearing-impaired patients and healthcare professionals.
- **Public Services:** Can assist communication at banks, government offices, and service centers.
- **Workplaces:** Supports interaction between hearing-impaired employees and other staff members.
- **Assistive Technology:** Can be integrated into computers, mobile applications, and smart devices for accessible communication.

VIII. CONCLUSION

The proposed AI-Based Real-Time Sign Language Translator demonstrates the effective integration of artificial intelligence, computer vision, machine learning, and speech technology for accessible communication. The system captures hand gestures using a webcam and applies image preprocessing to improve the quality of the input. MediaPipe Hands is used to detect the hand and extract 21 important landmarks representing finger and joint positions. These landmark coordinates are converted into feature vectors and supplied to a trained machine-learning classifier such as Random Forest, SVM, or KNN for gesture prediction. The recognized sign is subsequently converted into text and optionally transformed into speech using a text-to-speech module. The evaluated system achieved 94% gesture recognition accuracy, with 94 correctly recognized gestures out of 100 test samples. Text conversion and speech conversion achieved 96% and 93% success rates, respectively, while the average processing time was approximately 0.18 seconds. These results demonstrate the feasibility of real-time gesture recognition using a lightweight landmark-based approach. The system can support communication in education, healthcare, workplaces, public services, and assistive applications. Future improvements can include dynamic gesture recognition, larger datasets,

multilingual sign languages, improved environmental robustness, and mobile or edge-device deployment.

IX. FUTURE SCOPE

- **Dynamic Gesture Recognition:** Extend the system to recognize continuous gestures, words, and complete sentences using LSTM, Transformer, or other temporal models.
- **Larger Datasets:** Include diverse users, hand orientations, lighting conditions, backgrounds, and multiple sign-language datasets to improve generalization.
- **Mobile and Edge Deployment:** Develop Android/mobile and edge-device versions for portable, low-cost, and real-time sign-language translation.
- **Multilingual Translation:** Support different sign languages and integrate multilingual text-to-speech systems for broader communication.

REFERENCES

- [1] S. T. Abd Al-Latief, S. Yussof, A. Ahmad, S. M. Khadim, and R. A. Abdulhasan, "Instant sign language recognition by WAR strategy algorithm based tuned machine learning," *International Journal of Networked and Distributed Computing*, vol. 12, pp. 344–361, 2024, doi: 10.1007/s44227-024-00039-8.
- [2] F. M. Najib, "Sign language interpretation using machine learning and artificial intelligence," *Neural Computing and Applications*, 2025.
- [3] A. Baihan, A. I. Alutaibi, M. Alshehri, and S. K. Sharma, "Sign language recognition using modified deep learning network and hybrid optimization: A hybrid optimizer (HO) based optimized CNN-SaLSTM approach," *Scientific Reports*, vol. 14, Art. no. 26111, 2024, doi: 10.1038/s41598-024-76174-7.
- [4] J. Zhang, X. Bu, Y. Wang, H. Dong, Y. Zhang, and H. Wu, "Sign language recognition based on dual-path background erasure convolutional neural network," *Scientific Reports*, vol. 14, Art. no. 11360, 2024, doi: 10.1038/s41598-024-62008-z.
- [5] "Recent advances on deep learning for sign language recognition," *Computer Modeling in Engineering & Sciences*, 2024.
- [6] N. Shahin and L. Ismail, "From rule-based models to deep learning transformers architectures for natural language processing and sign language translation systems: Survey, taxonomy and performance evaluation," *Artificial Intelligence Review*, 2024.
- [7] M. Islam, M. Aloraini, S. Aladhadh, et al., "Toward a vision-based intelligent system: A stacked encoded deep learning framework for sign language recognition," *Sensors*, vol. 23, 2023.
- [8] N. F. Attia, M. T. F. S. Ahmed, and M. A. M. Alshewimy, "Efficient deep learning models based on tension techniques for sign language recognition," *Intelligent Systems with Applications*, 2023.
- [9] B. Saini, D. Venkatesh, N. Chaudhari, T. Shelake, S. Gite, and B. Pradhan, "A comparative analysis of Indian sign language recognition using deep learning models," *Forum for Linguistic Studies*, 2023.
- [10] K. Pandian, M. R. M. Azraai, M. K. Ismail, et al., "Sign language recognition using deep learning through LSTM and CNN," *Mekatronika*, 2023.
- [11] R. Damdoo and P. Kumar, "SignEdgeLVM transformer model for enhanced sign language translation on edge

devices,” Discover Computing, 2025, doi: 10.1007/s10791-025-09509-1.

[12] R. Damdoo and P. Kumar, “Enhanced sign language translation using three stream multimodal occlusion resilient transformer,” Discover Computing, 2025, doi: 10.1007/s10791-025-09772-2.

[13] P. A. Dal Bianco, O. A. Stanchi, F. M. Quiroga, and F. Ronchetti, “Multilingual sign language translation with unified datasets and pose-based transformers,” in Proc. Workshop on Sign Language Processing, 2025.

[14] A. R. Chowdhury and D. K. Sanyal, “Enhancing Indian sign language translation via motion-aware modeling,” in Proc. Workshop on Sign Language Processing, 2025.

[15] U. Rastogi, R. P. Mahapatra, and S. Kumar, “Advanced gesture recognition in Indian sign language using a synergistic combination of YOLOv10 with Swin Transformer model,” Scientific Reports, 2025.

[16] M. Maashi, H. G. Iskandar, and M. Rizwanullah, “IoT-driven smart assistive communication system for the hearing

impaired with hybrid deep learning models for sign language recognition,” Scientific Reports, 2025.

[17] A. Almjally, S. A. Algamdi, N. Aljohani, M. K. Nour, et al., “Harnessing attention-driven hybrid deep learning with combined feature representation for precise sign language recognition to aid deaf and speech-impaired people,” Scientific Reports, 2025.

[18] T. H. Dhrubo, M. T. Reza, S. Sahariar, W. A. Mukto, et al., “Transformer based sign-to-text translation for Bangladeshi sign language,” Scientific Reports, 2025.

[19] G. Guan, et al., “MVDT: Multiview distillation transformer for view-invariant sign language translation,” IET Computer Vision, 2025.

[20] A. R. Chowdhury and D. K. Sanyal, “Enhancing Indian sign language translation via motion-aware modeling,” in Proc. Workshop on Sign Language Processing, 2025.