International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2025.0173
e ISSN: 2584-2854
Volume: 03
Issue: 04 April 2025
Page No: 1060 - 1064

# Sign Language Conversion to Text and Speech Using Machine Learning

*Dr. M. Laxmaiah[1], V. Harshitha[2], C. Mahesh[3], P. Sumanth[4], CH. Harsha Vardhan[5]*
*[1]Head of Department, Dept. of Data science, CMR Engineering College, Medchal, 501401, Telangana, India.*
*[2,3,4,5]UG Scholar, Dept. of Data science, CMR Engineering College, Medchal, 501401, Telangana, India.*
***Emails ID:*** *datasciencehod@cmrec.ac.in[1], harshithareddyveerelly@gmail.com[2], cmaheshkumarguptha@gmail.com[3], palapalasumanth52@gmail.com[4], harshachelimilla@gmail.com[5]*

## Abstract
*This Study introduces a camera-based sign language detection system that bridges communication barriers for deaf and hard-of-hearing individuals. Unlike existing solutions requiring specialized sensors, our approach uses standard cameras with OpenCV for image capture and Convolutional Neural Networks for gesture recognition. The system processes hand movements in real-time, translating American Sign Language into text and speech through TensorFlow, MediaPipe, and the PYTTSX3 library. Experimental results demonstrate high accuracy across various environmental conditions and user variations. This accessible technology enables seamless communication between signing and non-signing individuals, promoting greater inclusion in educational, workplace, and public settings without requiring specialized equipment.*
***Keywords:*** *CNN; MediaPipe; OpenCV; PYTTSX3; TensorFlow.*

## 1. Introduction
Sign language serves as an essential communication tool for millions of hearing-impaired people globally, allowing them to convey their thoughts and engage with others throughout their daily activities. Even with its significance, a notable communication divide persists between the deaf community and individuals who are unfamiliar with sign language. In India, there are approximately 60 million individuals who are deaf or speech-impaired, creating a significant communication barrier that hinders inclusion and participation in educational, professional, and social settings. Conventional methods for sign language recognition have mainly depended on sensor-based or vision-based techniques, typically needing specialized equipment like data gloves or motion detectors. These solutions, though effective, have various disadvantages such as high expenses, invasiveness, restricted adaptability, and diminished usability in natural settings. These restrictions have obstructed the broad acceptance and availability of sign language translation technology. This study investigates the creation of a sign language detection system that relies solely on cameras, removing the requirement for extra sensors or specialized tools. Utilizing CNNs to analyze real-time video streams, our suggested system seeks to precisely recognize and translate American Sign Language (ASL) gestures into text and speech formats. The system emphasizes identifying the 26 letters of the English alphabet, establishing a base for broader communication. The importance of this effort rests in its ability to foster a more inclusive community by dismantling communication obstacles. Our goal is to create a system that translates sign language into accessible formats for non-signers, enabling smooth communication between deaf people and the wider community. [1] Additionally, our strategy emphasizes accessibility and scalability, guaranteeing that the technology can be broadly implemented with standard cameras available in common devices. Figure 1 shows the American sign language alphabet in which every letter is represented by a distinct hand gesture. These hand signs form the foundation of ASL communication, allowing individuals to spell out words and convey messages through fingerspelling. Figure 2 shows Accuracy of CNN-LSTM. combining spatial and temporal information

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2025.0173
e ISSN: 2584-2854
Volume: 03
Issue: 04 April 2025
Page No: 1060 - 1064

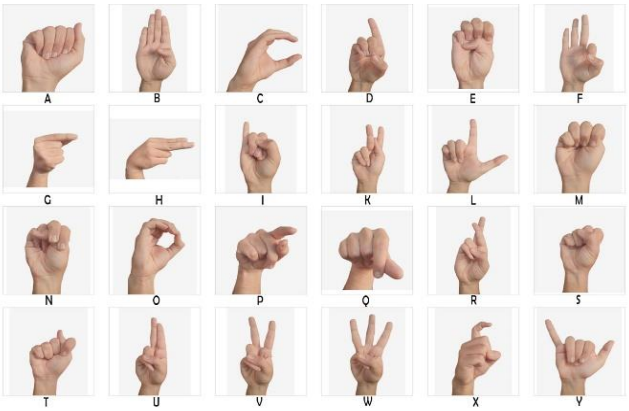


**Figure 1 Dataset**

## 2. Literature Survey

Sign language recognition has gained significant attention due to its potential to bridge communication gaps between the deaf community and non-signers. Traditional methods relied on sensor-based techniques that involved motion detectors and specialized gloves. These approaches had drawbacks, including high cost, limited adaptability, and invasiveness. More effective and user-friendly sign language recognition systems have been made possible by recent developments in deep learning, transfer learning, and computer vision.

### 2.1 Deep Learning for Sign Language Recognition

The accuracy of SLR has been greatly improved by deep learning models, especially Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs). RNNs manage sequential dependencies in gesture recognition, whereas CNNs are frequently used to extract spatial features from images. Research has demonstrated that CNN-based models are highly accurate at recognizing hand gestures. [2&3] A CNN-based architecture for real-time SLR was used in a recent study, which showed increased accuracy using data augmentation methods and optimized hyperparameters. Another method improved recognition rates in real-world situations by combining computer vision and deep learning. The generalizability of deep learning models is still impacted by issues like signer dependency, occlusion, and background noise, though. [4&6] combining spatial and temporal information

### 2.2 Transfer Learning and Multi Model Approaches

Transfer learning has been widely adopted in SLR to leverage pre-trained models. These models, trained on large datasets, help overcome the limitations of small sign language datasets by transferring learned features to new recognition tasks. A comparative analysis of different deep learning architectures for Indian Sign Language recognition demonstrated that transfer learning models outperform traditional CNNs, achieving higher accuracy and robustness. [11] Multi-model approaches, where CNNs are integrated with LSTMs (Long Short-Term Memory networks) or Transformer models, have also been explored. By combining spatial and temporal information, these hybrid models improve recognition performance for complex gestures. Another study highlighted the importance of domain adaptation in transfer learning, showing that fine-tuning pre-trained models on sign language datasets leads to better accuracy. [10]

### 2.3 Object Detection for Sign Language Detection

Object detection techniques play a crucial role in segmenting and identifying hand gestures from video frames. OpenCV-based object detection methods have been utilized to improve feature extraction for SLR. Feature-based detection techniques such as contour detection, edge detection, and key point tracking enhance recognition accuracy by filtering irrelevant background information.[5] To create a reliable sign language recognition system, another study combined real-time tracking algorithms with CNN-based object detection. However, object detection models must be optimized for different lighting conditions and hand variations to improve real-world applicability.

### 2.4 Transformer-Based Models for Gesture Recognition

we utilize CNN-Transformer hybrid architectures for sign language recognition. Transformers are integrated to enhance temporal sequence modelling, improving gesture recognition accuracy compared to traditional CNN-RNN approaches. The self-attention

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2025.0173
e ISSN: 2584-2854
Volume: 03
Issue: 04 April 2025
Page No: 1060 - 1064

mechanism in transformers enables efficient feature extraction from sequential hand movements, making the system robust to variations in gesture speed and signer differences. Our approach combines CNN-based spatial feature extraction with transformers for sequence learning, ensuring better recognition of complex sign gestures. We employ pre-trained transformer models with transfer learning to optimize performance while reducing the need for large training datasets. Additionally, the system is designed for real-time processing, allowing deployment on edge devices with minimal computational overhead.

### 2.5 Directions on Future Sign Language Detection

In future developments, we aim to expand the capabilities of our sign language recognition system by integrating it into an Android application, making it more accessible to a broader audience. This mobile implementation will leverage on-device machine learning for real-time gesture prediction, ensuring efficient performance without requiring an internet connection. By optimizing the system for low-latency processing and energy efficiency, we aim to make the application practical for everyday use. Furthermore, we plan to extend the gesture vocabulary beyond the ASL alphabet by incorporating dynamic hand gestures and full sign language phrases, improving the system's ability to recognize more complex expressions. Future enhancements will include multi-hand gesture detection, context-aware predictions, and user adaptability, allowing the model to learn and refine accuracy based on individual signing styles.

## 3. Proposed Approach

The proposed gesture detection system combines computer vision and machine learning algorithms to offer an accurate and scalable real-time gesture recognition solution. The system obviates the requirement for expensive sensors or gloves as it relies on a camera-only configuration, rendering it more accessible and cost-efficient. The below modules are incorporated into the proposed system.

### 3.1 Vision-Based Gesture Analysis

The system uses deep learning algorithms to recognize hand motion and gestures tracked via an ordinary webcam. Convolutional Neural Networks (CNNs) obtain spatial features, and Long Short-Term Memory (LSTM) networks obtain sequential dependencies for better accuracy. The process is designed to make the system able to identify and classify gestures with minimal latency. The data thus obtained is pre-processed to maximize model performance and generalization across multiple users and varying lighting conditions.

### 3.2 Classification Based on Machine Learning

For improving the accuracy of recognition, the gesture features that are extracted are classified through optimized deep models of learning. The training is done on large-scale gesture datasets with CNNs and LSTMs, ensuring optimal performance through hyperparameter tuning. Transfer learning is used to enhance accuracy and minimize training time by tapping into pre-trained models. The trained model is saved for real-time classification, facilitating efficient and automatic sign language recognition. Table 1 shows Model Performance Comparison.

**Table 1 Model Performance Comparison**

| Model | Accuracy (%) | Precision (%) | Recall (%) |
|---|---|---|---|
| CNN | 94.1 | 93.5 | 94.8 |
| RNN | 92.5 | 91.8 | 93.0 |
| Transformer | 95.3 | 94.7 | 95.9 |

### 3.3 Data Preprocessing and Collection

Sign language gesture data is gathered through live video feed to provide real-world usage. The collected data goes through pre-processing operations like resizing, normalization, and data augmentation to enhance model strength. Feature extraction mechanisms further process the input data to ensure accurate recognition of gestures, considering variations in hand positioning and movement.

### 3.1 Real Time Detection and Translation

The system has real-time gesture recognition, recording live hand gestures through a webcam and converting them into text. The recognized gestures are categorized and shown on an easy-to-use interface, allowing for smooth communication. The model is real-time optimized, ensuring the processing is low-latency, thereby providing fluid and natural sign language interpretation [7-10].

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2025.0173
e ISSN: 2584-2854
Volume: 03
Issue: 04 April 2025
Page No: 1060 - 1064

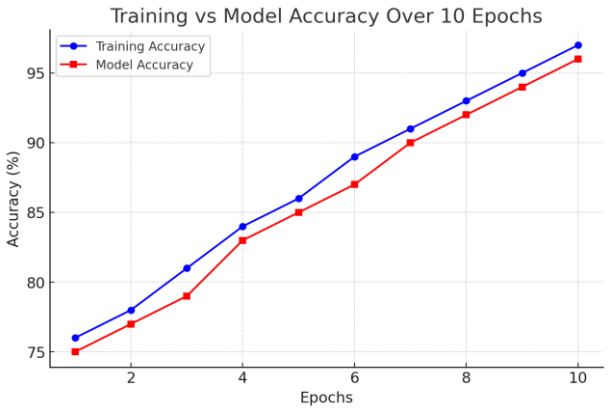


**Figure 2 Accuracy of CNN-LSTM**

### 3.2 Integration to Text and Speech

The Speak functionality in the application is designed to convert recognized sign language text into speech using the pyttsx3 text-to-speech (TTS) engine. When a user performs sign gestures, the application translates them into text, which is stored as a sentence. Upon pressing the Speak button, which is configured to trigger the speak fun method, the system reads the converted text aloud using pyttsx3. To execute the speech output. This feature enhances accessibility for individuals using sign language by allowing them to communicate verbally with non-sign language users.

### 3.3 Word Suggestions and Word Predictions

The suggestion mechanism relies on several key components to provide alternative word recommendations and improve the accuracy of recognized text. The system likely uses an English dictionary library (enchant.Dict("en-US")) or a predictive algorithm to generate possible word alternatives. These suggested words are dynamically updated based on the recognized characters forming words during sign language interpretation. This mechanism significantly enhances the accuracy and usability of the application by reducing errors in sign-to-text conversion and allowing users to actively participate in refining the final text output.

## 4. Results, Experimentation and Discussion

Two experiments were conducted to evaluate the sign language recognition system. The first experiment involved fine-tuning model parameters like convolutional layers, filter sizes, and optimizers to optimize accuracy. The second experiment tested the trained model on colour and grayscale datasets to assess its generalization capability [12]. The model, trained using CNN and Media pipe for hand landmark detection, achieved an accuracy of 97% in challenging conditions and 99% in optimal settings. The system effectively handled varied backgrounds and lighting, improving recognition accuracy compared to traditional methods. A Python GUI was developed using Tkinter to provide a user-friendly interface for real-time gesture recognition. The system also includes a suggestion feature that helps users correct misrecognized signs. Additionally, a text-to-speech module using the pyttsx3 library converts recognized signs into audio output, enhancing accessibility for communication. Future work will focus on dynamic gesture recognition and mobile application integration. Figure 3 shows Gary Image with Gaussian Blur.

### 4.1 Results and Discussion

**Figure 3 Gary Image with Gaussian Blur**

After extracting image features, Mediapipe is used to detect hand landmarks, identify key points, and classify the hand sign based on these landmarks, ultimately generating the corresponding text output Figure 4 shows Working Model Output.

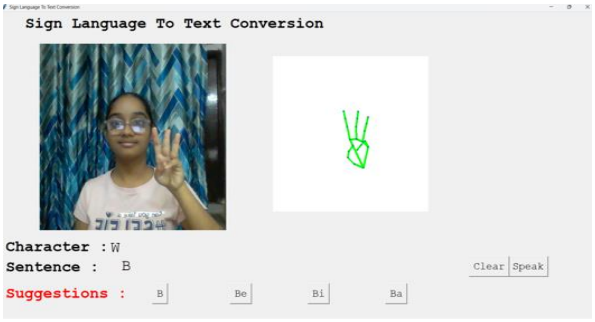


**Figure 4 Working Model Output**

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2025.0173
e ISSN: 2584-2854
Volume: 03
Issue: 04 April 2025
Page No: 1060 - 1064

## Conclusion

We have been able to successfully create a strong gesture recognition system that can precisely predict any alphabet (A–Z) with 97% accuracy in varied conditions, such as different backgrounds and lighting. In ideal conditions like a clear background and adequate illumination the model is an impressive 99% accurate, proving its suitability for sign language recognition. This system is indicative of the promise of computer vision and deep learning for real-time sign language interpretation, and it presents an important step towards closing the communication gap for the deaf and hard-of-hearing population.

## References

[1]. Victoria Adebimpe Akano, Adejoke O Olamiti (2018). Conversion of Sign Language To Text and Speech Using Machine Learning Techniques. DOI: 10.36108/ jrrslasu/ 8102/ 50(0170)

[2]. Felix Zhan (2019). Hand Gesture Recognition with Convolution Neural Networks. Doi: 10.1109/IRI.2019.00054

[3]. R.S. Sabeenian, S. Sai Bharathwaj, and M. Mohamed Aadhil (2020). Sign Language Recognition Using Deep Learning and Computer Vision. Doi: 10.5373/ JARDCS/ V12SP5/20201842

[4]. Muhammad AI-Qurishi, Thariq Khalid, Riad Souissi (2021). Deep Learning for Sign Language Recognition: Current Techniques, Benchmarks, and Open Issues. Doi: 10.1109/ACCESS.2021.3110912

[5]. Ayushi Sharma; Jyotsna Pathak; Muskan Prakash; J N Singh (2021). Object Detection using OpenCV and Python. Doi: 10.1109/ICAC3N53548.2021.9725638

[6]. Md Nafis Saiful, Abdulla AI Isam, Hamim Ahmed Moon, Rifa Tammana (2022). Real-Time Sign Language Detection Using CNN. Doi: 10.1109/ICDABI56818.2022.10041711

[7]. M. J, B. Krishna, S. S, Surendar K (2022). Sign Language Recognition using Machine Learning. Doi: 10.1109/ ICSES55317. 2022.9914155

[8]. Aman Pathak, Avinash Kumar, Gupta (2022). Real Time Sign Language Detection. Doi: 10.46501/IJMTST0801006

[9]. CH. Nanda Kumar, E. Nithin Computer, C. Krishna, Ch. Bindhu Madhavi (2023). Real-Time Face Mask Detection using Computer Vision and Machine Learning. DOI:10.1109/ICEARS56392.2023.10085276

[10]. N. Rajasekhar, M. Yadav, Charitha Vedantam, Karthik Pellakuru, Chaitanya Navapete (2023). Sign Language Recognition using Machine Learning Algorithm. Doi: 10.1109/ICSCSS57650.2023.10169820

[11]. Bunny Saini, Divya Venkatesh, Nikita Chaudhari, Tanaya Shelake, Shilpa Gite, Biswajeet Pradhan (2023). A comparative analysis of Indian sign language recognition using deep learning models. Doi: 10.18063/fls.v5i1.1617

[12]. Sajin Xavier, Vaisakh B, Maya L. Pai (2023). Real-time Hand Gesture Recognition Using Media Pipe and Artificial Neural Networks. Doi: 10.1109/ICCCNT56998.2023.10306439