IJMRT: Volume (7), Issue 5, 2025
International Journal of Multidisciplinary Research Transactions
(*A Peer Reviewed Journal*)
www.ijmrt.in

ISSN (Print):2663-2381
ISSN (Online):2663-4007

Sign-Meet Virtual Meeting Platform with Sign Language Recognition

Chandrasekaran K S[1]*, Denita P[2], Jocelyn A[3], Khrusanth S[4], Monisha J[5]

[1]Associate Professor, Department of Computer Science and Engineering, Saranathan College of Engineering, Trichy, Tamil Nadu, India.
[2,3,4,5]Student, Department of Computer Science and Engineering, Saranathan College of Engineering, Trichy, Tamil Nadu, India.

*Corresponding Author

DoI: https://doi.org/10.5281/zenodo.15348353

Abstract

Virtual meetings are a common mode of communication but are largely inaccessible for individuals with hearing and speech impairments. SIGN-MEET is an AI-powered platform integrating sign language recognition, speech-to-text, and avatar-based translation to enable inclusive real-time interaction. The system uses Temporal Convolutional Networks (TCN) for gesture interpretation, Hidden Markov Models (HMM) for speech recognition, and animated avatars for visual sign output. This ensures that deaf, mute, and non-signing users can engage equally in virtual environments.

Keywords: Sign Language, Deep Learning, Temporal Convolutional Networks, Avatar, Speech Recognition.

1. Introduction

Sign language is a vital mode of communication for the deaf and mute community. Traditional virtual platforms fail to offer real-time sign language translation, creating a communication gap. This project introduces SIGN-MEET, a system that enables seamless interaction between signers and non-signers using AI. With the rise of deep learning models like TCN, real-time gesture interpretation is now feasible, allowing sign language to be translated accurately into text or speech.

2. Literature Review

[1] In the paper Sign Language Recognition Using Template Matching Technique by Soma Shrenika and Myneni Madhu Bala, a system is proposed that uses a standard webcam to capture hand gestures, which are then preprocessed using image processing techniques such as RGB to grayscale conversion, Gaussian filtering, and binarization. Feature matching is done using Sum of Absolute Differences (SAD), a pixel-wise comparison metric. The system is simple and avoids hardware dependency like gloves or sensors. Although efficient for static gestures, it lacks support for real-time dynamic gesture recognition, which is vital in continuous communication.

[2] The paper Real-Time Bangla Sign Language Detection with Sentence and Speech Generation by Dipon Talukder and Fatima Jahara introduces a deep learning model that uses YOLOv4 with CSPDarknet53 backbone for Bangla Sign Language recognition. It allows sentence generation and speech synthesis using captured gesture sequences. The dataset consists of 49 classes including digits, alphabets, and compound characters. The proposed model achieves over 82% mAP accuracy in complex backgrounds. This paper shows the potential of YOLO in sign language detection under realistic conditions.

[3] In the publication An Efficient Approach for Interpretation of Indian Sign Language using Machine Learning by Dhivyasri S et al., Speeded-Up Robust Features (SURF) were extracted and classified using SVM, CNN, and RNN. SVM, combined with K-means clustering and Bag of Visual Words (BoV), provided the best results. The system supports

both Gesture-to- Text and Speech-to-Gesture conversions. This approach showed promising results in Indian Sign Language (ISL) recognition with reliable accuracy.

[4] Automated Sign Language Interpreter Using Data Gloves by Anupama H S and Usha B A proposes a glove-based system equipped with flex sensors and an Arduino microcontroller. The K-Nearest Neighbors (KNN) algorithm processes sensor readings to classify signs. Though hardware-dependent, this method ensures consistent accuracy for alphabetic sign gestures and simple phrases, demonstrating effectiveness in small-scale deployments.[5] Sign to Speech Conversion Using SVM by Malli Mahesh Chandra and Rajkumar S presents a wearable glove system using MPU6050 motion sensors. The sensor data is fed into an SVM model to classify both American Sign Language (ASL) and Indian Sign Language (ISL) gestures. The system can produce voice output in multiple languages with 100% accuracy on ISL gestures. However, the limitation lies in recognizing two-handed gestures with only one glove.

[5] Sign to Speech Conversion Using SVM by Malli Mahesh Chandra and Rajkumar S presents a wearable glove system using MPU6050 motion sensors. The sensor data is fed into an SVM model to classify both American Sign Language (ASL) and Indian Sign Language (ISL) gestures. The system can produce voice output in multiple languages with 100% accuracy on ISL gestures. However, the limitation lies in recognizing two-handed gestures with only one glove.

[6] Wearable Sensor-Based Sign Language Recognition: A Comprehensive Review by Karly Kudrinko et al. analyzes various sign language recognition systems using wearable technologies. The paper compares sensor configurations, classifier types, and user studies. It

concludes that despite impressive accuracy, wearable systems suffer from usability issues, inconsistent datasets, and hardware complexity. This study emphasizes the need for user-centric and scalable SLR solutions.

[7] Sign Language Recognition Based on Intelligent Glove Using Machine Learning Techniques by Henry Benitez-Pereira and Diego H introduces a glove-based sign recognition model that uses DROP3 optimization and CHC evolutionary algorithms to reduce the dataset while maintaining accuracy. Classification is done using KNN, achieving data reduction by 98% while ensuring fast and accurate predictions. The study showcases efficient prototype selection for large-scale sensor data classification.

3. Proposed System

The proposed SIGN-MEET system is an inclusive, AI-powered virtual meeting platform designed to bridge the communication gap between individuals with hearing and speech impairments and non-signers. It integrates three primary modules—Sign Recognition Module (SRM), Speech Recognition and Synthesis Module (SRSM), and Avatar Module (AM)—to enable seamless real-time communication. The SRM employs Temporal Convolutional Networks (TCNs) for accurate recognition of dynamic hand gestures from live video, while the SRSM uses Wavelet Transform and Mel-Frequency Cepstral Coefficients (MFCCs) for feature extraction, followed by Hidden Markov Models (HMMs) to convert speech into text with high accuracy. The Avatar Module generates real-time 3D animations of sign language using motion synthesis techniques, enhancing visual communication for deaf users. SIGN-MEET supports multilingual translation through Natural Language Processing (NLP), enabling speech and sign conversions across various languages and dialects, with a primary focus on Indian Sign Language (ISL). The system is designed for integration with multiple video conferencing platforms, including Jitsi, Zoom, Microsoft Teams, and Google Meet. It features adaptive real-time processing, supports continuous gesture stream interpretation, and includes intelligent feedback loops for continuous model improvement. A secure MySQL backend manages user roles, meeting data, and training datasets, while tailored interfaces for admins, deaf users, and non-deaf users ensure role-specific functionalities. Additional features include text-to-speech conversion, avatar playback controls, and customizable interfaces for accessibility. Future enhancements include mobile app deployment and offline functionality, making SIGN-MEET a scalable, robust, and inclusive solution for accessible digital communication.

3.1. Sign Recognition Module (SRM)

This module is responsible for capturing and interpreting sign language gestures using a webcam. The model adopted here is the Temporal Convolutional Network (TCN), a deep learning model designed to recognize sequential patterns in video frames. The webcam continuously captures live hand gestures, which are then broken down into individual frames. Preprocessing involves grayscale conversion, noise reduction using Gabor filters, and smart cropping to isolate gesture regions. Following preprocessing, gesture features are extracted using Region Proposal Networks (RPNs), which help identify the area of interest. The TCN is then used to classify temporal features from the extracted frame sequence. Unlike conventional CNNs, TCNs allow learning across time steps, making them ideal for sign language, where gesture continuity and sequence are crucial.

3.2. Speech Recognition and Synthesis Module (SRSM)

The SRSM handles the conversion of spoken language into text and synthesized sign output. It uses Hidden Markov Models (HMM) for effective modelling of speech patterns. Audio captured through the microphone is first cleaned using wavelet transforms and processed using MFCC (Mel Frequency Cepstral Coefficients) to extract relevant features. HMMs are then used to decode the speech into meaningful textual representation. The speech-to-text output can then be forwarded to the avatar system, which translates the message into visual signs, making verbal communication accessible to deaf users.

3.3.Avatar Module (AM)

This module focuses on delivering expressive, understandable sign language output to users. A pre-designed 3D avatar receives textual or classified gesture input and performs the appropriate sign gestures. The avatar has been trained with animation sequences mapped to Indian Sign Language (ISL). Additionally, it can convey facial expressions and body movements to ensure expressiveness. The avatar also includes an audio feedback feature. When a user performs a sign gesture, the system not only provides the identified sign name and confidence percentage on the screen but also plays it back as audio. For instance, if a gesture is detected to be 90% accurate as “Hello” and 10% close to “Thank You,” both results will be shown with respective percentages via text and audio.

3.4. Capturing and Preprocessing

The user interface begins by capturing video via the web camera. Frames are extracted in real time and undergo a preprocessing stage. This includes padding and intelligent cropping to centre the hand gesture, grayscale transformation for uniform lighting normalization, and filtering to enhance gesture edge visibility.

3.5. Detection using Pose Estimation

Once pre-processed, the frames are passed into the pose estimation model which detects 17 key body landmarks. Each key point includes an (x, y) coordinate and a confidence score—an indicator of how accurate the detected point is. This results in a 17×3 vector (51 features in total), forming the input to the next classification stage.

3.6. Classification

The key point vector is passed into a custom-built multi-layer perceptron (MLP) for classification. Our MLP consists of 7 dense layers and is trained to predict sign gesture classes. Rather than displaying only the top class, the system presents the top 3 predictions with corresponding accuracy percentages to provide users with insights on near- matches and improve real-time self-correction.This multi-module architecture ensures accurate sign recognition, real-time translation, and expressive output—making SIGN-MEET a practical and inclusive communication platform for virtual interactions.

4. Experimental Analysis and Results

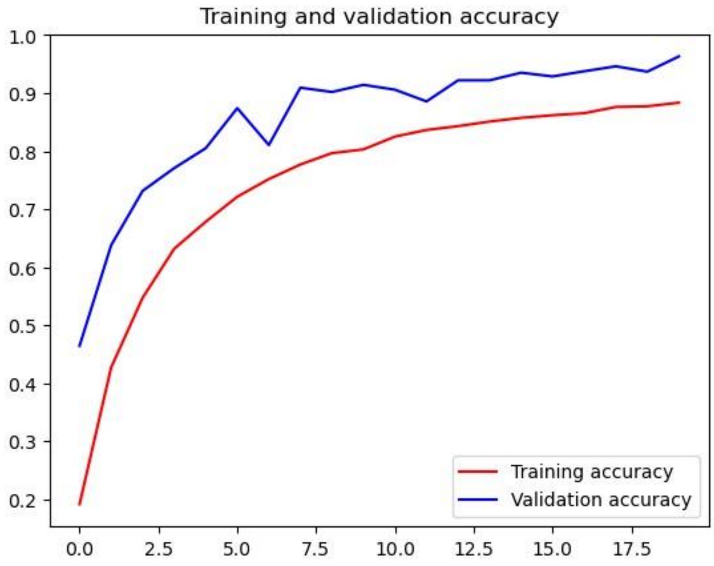

Figure.1. Model Accuracy

5. Performance Metrics And Result

5.1. Metrics

- Accuracy : (950+970)/ (950 + 20+ 10+970)= 0.98
- Precision : 970/ (970+20) = 0.98
- Recall : 970 / (970+10) = 0.99
- F1-Score : 2* (0.98*0.99)/(0.98+0.99)= 0.985

5.2. Result

- Accuracy : 98%
- Precision : 98%
- Recall : 99%
- F1-Score : 98.5%

6. Conclusion

In conclusion, SIGN-MEET bridges the communication gap between deaf and non-deaf users by integrating sign language recognition, speech-to-text conversion, and avatar-based visual translation within a real-time virtual meeting platform. Using Temporal Convolutional Networks (TCNs) and Hidden Markov Models (HMMs), the system ensures accurate and seamless bi-directional communication. Its multilingual support, avatar generation, and integration with platforms like Jitsi enhance accessibility and inclusivity. The solution is scalable and holds strong potential for future improvements in AI accuracy, expanded language support, and wider platform integration.

REFERENCES

[1]. Bird, J. J., Ekárt, A. and Faria, D. R, 2023, British sign language recognition via late fusion of computer vision and leap motion with transfer learning to American sign language. Sensors, 20(18), pp. 5151.

[2]. Chowdhary, G. T. R. C. L. and Parameshachari, B. D. 2023. Computer Vision and Recognition Systems: Research Innovations and Trends. CRC Press.

[3]. Gadekallu, T. R., Srivastava, G. and Liyanage, M. 2022. Hand gesture recognition based on a Harris hawks optimized convolution neural network. Computers & Electrical Engineering, 100, Article ID 107836.

[4]. Halvardsson, G., Peterson, J., Soto-Valero, C. and Baudry, B. 2022. Interpretation of Swedish sign language using convolutional neural networks and transfer learning. SN Computer Science, 2(3), pp. 1–15.

[5]. Li, F., Shirahama, K., Nisar, M. A., Huang, X. and Grzegorzek, M. 2020. Deep transfer learning for time series data based on sensor modality classification. Sensors, 31(20), pp. 4271.

[6]. Likhar, P., Bhagat, N. K. and G N, R. 2021. Deep learning methods for Indian sign language recognition. In 2020 IEEE 10th International Conference on Consumer Electronics (ICCE-Berlin), pp. 1–6.

[7]. Likhar, P., Bhagat, N. K. and G N, R. 2021. Deep learning methods for Indian sign language recognition. In 2020 IEEE 10th International Conference on Consumer Electronics (ICCE-Berlin), pp. 1–6.

[8]. Riaz, M. M. and Zhang, Z. 2021. Surface EMG real-time Chinese language recognition using artificial neural networks. In Intelligent Life System Modelling Image Processing and Analysis, Communications in Computer and Information Science. Springer, 1467.

[9]. Shakeel, Z. M., So, S., Lingga, P. and Jeong, J. P. 2020. MAST: Myo Armband Sign-Language Translator for human hand activity classification. In IEEE International Conference on Information and Communication Technology Convergence, pp. 494–499.

[10]. Sharma, S., Gupta, R. and Kumar, A. 2020. Trbaggboost: An ensemble-based transfer learning method applied to Indian Sign Language recognition. Journal of Ambient Intelligence and Humanized Computing.