International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

AI Powered Accessibility for Enabling Effective Communication for Hearing and Speech Impaired in Virtual Platforms

Gayathri. R[1], Priyanka. K[2]

Assistant professor, Department of MCA, Vivekanandha Institute of Information and Management Studies.
PG student, Department of MCA, Vivekanandha Institute of Information and Management Studies.

ABSTRACT:

Effective communication is vital for sharing information, ideas, and emotions. However, virtual meeting platforms like Zoom, Microsoft Teams, and Google Meet often fail to accommodate individuals with hearing and speech impairments, creating significant barriers to inclusivity. While sign language offers a means of communication for deaf individuals, its interpretation remains challenging for non-signers. Existing sign language recognition technologies are limited in accuracy and accessibility, making them unsuitable for seamless integration into virtual platforms. This project introduces an AI-driven system designed to bridge the communication gap between deaf and hearing participants in virtual meetings. The system employs advancements in deep learning, particularly Temporal Convolutional Networks (TCNs), to enable two-way communication in real-time. It includes three core modules: a Sign Recognition Module (SRM) that interprets signs using TCN, a Speech Recognition and Synthesis Module (SRSM) powered by Hidden Markov Models, which converts spoken language into text, and an Avatar Module (AM) that visually translates speech into corresponding signs. The Avatar Module is essential for visually representing spoken language in sign language format, ensuring non-signers can effectively communicate with sign language users in an intuitive and engaging way. Trained on Indian Sign Language, the system facilitates communication across diverse groups, including deaf, mute, hard-of-hearing, visually impaired, and non-signers. Its integration into popular virtual meeting platforms through a user-friendly web-based interface enhances accessibility and participation. This solution represents a significant advancement in fostering inclusivity and accessibility in virtual meeting environments.

Index Terms:AI-powered communication, Sign Language Recognition, Temporal Convolutional Networks (TCN), Hidden Markov Models (HMM), Avatar-based translation, Virtual meeting accessibility, Indian Sign Language, Inclusive technology.

I. Introduction

As communication remains a cornerstone of human interaction, individuals who are deaf or mute often encounter significant barriers, especially when engaging with non-signers. Sign language serves as an effective communication medium within the deaf community, but its usability becomes limited in mixed communication environments. Existing technologies for sign language recognition and translation often fall short in real-world applications due to limitations in accuracy, speed, or accessibility [1]. To bridge this communication divide, there is a pressing need for a reliable, real-time, and scalable solution that facilitates two-way communication between deaf and non-deaf individuals.

In response, this project proposes the development of an AI-powered bi-directional communication system using Temporal Convolutional Networks (TCNs) for Indian Sign Language (ISL) recognition and avatar-based sign synthesis. The system interprets ISL gestures and converts them into readable text while translating spoken language into sign language animations using a virtual avatar, thus creating an inclusive interface for communication. It comprises a Sign Recognition Module (SRM), Speech Recognition and Synthesis Module (SRSM), and Avatar Module (AM), all integrated into a web-based platform to ensure accessibility and scalability in virtual meeting environments.Previous research introduced a sign recognition method using the Sum of Absolute Differences (SAD) technique on RGB images to convert static gestures into text, emphasizing affordability by avoiding the need for sensors or gloves, but lacking support for dynamic gesture recognition [1]. Another study employed YOLOv3 with CSPDarknet53 for real-time Bangla Sign Language detection and speech synthesis, enhancing accuracy while enabling sentence formation using compound signs [2]. A comparative analysis evaluated multiple classifiers (CNN, RNN, and SVM) and feature extractors (SURF, BoV) for ISL recognition, identifying a hybrid SVM-KMeans approach as most accurate for dynamic gestures [3]. A glove-based approach using flex sensors and KNN also achieved gesture recognition but presented limitations in cost-effectiveness and deployment ease [4]. Another system used MPU6050 sensors and SVM classification to convert gestures into speech, offering a solution for the speech-impaired through wearable hardware [5].Building upon these works, the proposed system eliminates the dependence on wearable hardware, supports dynamic gestures, and introduces speech-to-sign translation with animated avatars. By leveraging TCNs and Hidden Markov Models (HMMs) for gesture and speech recognition, respectively, the system aims to deliver real-time, accurate, and user-friendly communication between hearing and speech-impaired users and the wider community, particularly in virtual platforms like Zoom, Teams, and Meet.

II. Related Works

Shrenika and Bala [1] developed a sign language recognition system using a template matching technique aimed at facilitating communication between deaf individuals and non-signers. The study addresses the challenge that normal people face when interpreting signs, as they are often unaware of the meaning behind specific gestures. The system captures hand gestures using a camera, processes the images, and converts them into text that non-signers can understand. Implemented using OpenCV and Python, the system leverages RGB models and Gaussian filters to preprocess images, smoothing out noise and detecting changes in intensities. The core of the recognition method relies on the Sum of Absolute Differences (SAD) technique, which compares input images to a pre-stored dataset of sign images by calculating the differences in pixel values. Their dataset contains 70 samples for each of the 36 symbols in sign language, ensuring comprehensive coverage of hand shapes and movements. The authors emphasize that their system provides a solution that does not rely on expensive equipment such as gloves or sensors, making it accessible and adaptable for use on mobile platforms. Their work also highlights the system's potential for both sign-to-text and text-to-sign translation, expanding its utility for real-world communication scenarios. However, they acknowledge that future work should focus on incorporating dynamic gestures and improving accuracy for continuous sign recognition.

Talukder and Jahara [2] developed a real-time Bangla Sign Language detection system using YOLOv3 and CSPDarknet53 for efficient sign recognition, converting signs into text and speech with added novel signs for sentence formation. They applied HSV color space conversion, SIFT for feature extraction, and PCA for dimensionality reduction.

Dhivyasri et al. [3] proposed an Indian Sign Language recognition system combining image processing and machine learning classifiers like CNN, RNN, and SVM. Using SURF features and models including K-Means and Bag of Visual Words, their hybrid SVM-based approach achieved the highest accuracy in gesture recognition.

Anupama and Usha [4] created an automated sign language interpreter using data gloves with flex sensors. Sensor data captured via Arduino was classified by a K-nearest neighbor algorithm to translate hand gestures into text, improving accessibility despite some challenges in distinguishing similar signs.

Chandra and Rajkumar [5] designed a sign language-to-speech prototype using MPU6050 sensors and an Arduino Nano to capture gestures. Data transmitted via Bluetooth was classified by an SVM model, enabling conversion of ASL and ISL gestures into speech in English and some Indian languages to aid communication for speech-impaired users.

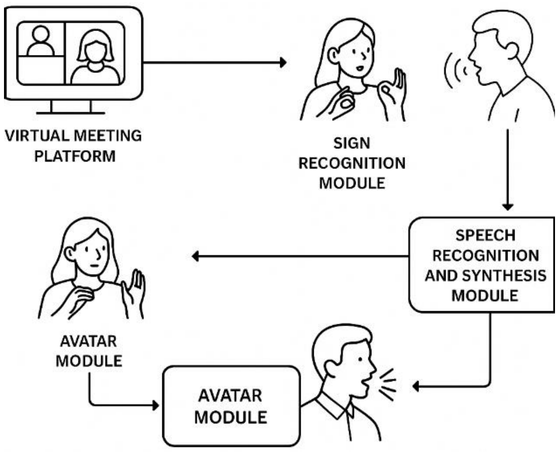
III. Methodology

The methodology involves the design, training, and deployment of an AI-powered sign language recognition system that enables real-time translation of sign language gestures into spoken and written language. The process begins with the capture of live video input using a device's webcam or camera, which continuously records hand and facial gestures for recognition. The video feed is processed frame-by-frame to ensure minimal latency, facilitating smooth and uninterrupted communication.

In the preprocessing stage, each frame undergoes several transformations to improve data quality and reduce computational complexity. This includes converting RGB frames to grayscale, resizing them to standard dimensions, and applying Gabor filters for noise reduction and edge enhancement. Binarization is then used to convert the grayscale images into high-contrast black-and-white representations, which simplifies the detection of hand shapes. To isolate relevant gestures, segmentation is carried out using a Region Proposal Network (RPN), which separates the hand regions from the background, allowing the system to focus on meaningful motion data.Following preprocessing, the system proceeds to feature extraction, where critical elements such as hand shape, finger orientation, and motion trajectory are captured. This is accomplished using a Fully Connected Layer (FCL) that encodes the visual information into high-dimensional feature vectors. These vectors are then passed to a Convolutional Neural Network (CNN) for classification, where spatial and temporal patterns are analyzed to assign probabilities to specific signs, which are then mapped to corresponding words or phrases.

To effectively capture the temporal dynamics of sign language, the model is trained using a Temporal Convolutional Network (TCN). TCN is well-suited for processing sequential data, allowing the system to learn dependencies and variations across gesture sequences. The model is trained on a comprehensive dataset comprising diverse sign language gestures, using optimization techniques such as batch normalization and dropout to enhance generalization and prevent overfitting.

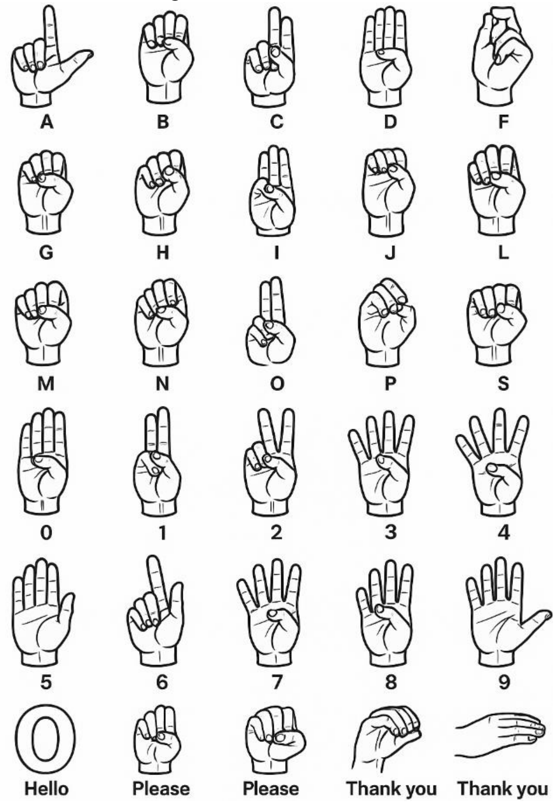
Once trained, the model is deployed in a real-time environment, where it processes live video input, performs gesture classification, and outputs translated text or synthesized speech. To enable multilingual communication, a Multilanguage Interpretation module is integrated. This module employs Natural Language Processing (NLP) techniques to translate recognized gestures into multiple languages and also supports speech-to-sign translation using Mel-Frequency Cepstral Coefficients (MFCC) for feature extraction and Hidden Markov Models (HMMs) for speech decoding. The translated text is then converted into sign language using a 3D AI-powered avatar.The Avatar Generation module creates a lifelike visual representation of sign language, translating speech or text into animated gestures in real time. This is achieved through motion synthesis and gesture mapping algorithms, ensuring accurate and expressive visual communication. Additionally, the Visual Communication component enables two-way interaction by converting text responses from deaf users into synthesized speech, thus ensuring inclusive and accessible communication for all participants in a virtual setting.



IV. Experimental Analysis

Dataset Description

This project uses a robust and diverse dataset consisting of over images and video frames representing various sign language gestures, including the 26 letters of the English alphabet, numbers from 0 to 9, and commonly used words and phrases. The dataset was collected using Media Pipe Hands and Pose estimation models, ensuring precise detection of hand landmarks and key points. Data samples were captured under varying lighting conditions, backgrounds, and hand orientations, enhancing the model's ability to generalize across real-world scenarios. The dataset includes participants of different age groups and skin tones, contributing to its inclusivity and robustness. Each gesture class is stored in clearly labeled directories in JPG and MP4 formats, and annotated with metadata such as gesture name, timestamp, and frame count. The dataset is suitable for training deep learning models for gesture recognition, sequence classification, and real-time sign translation tasks.



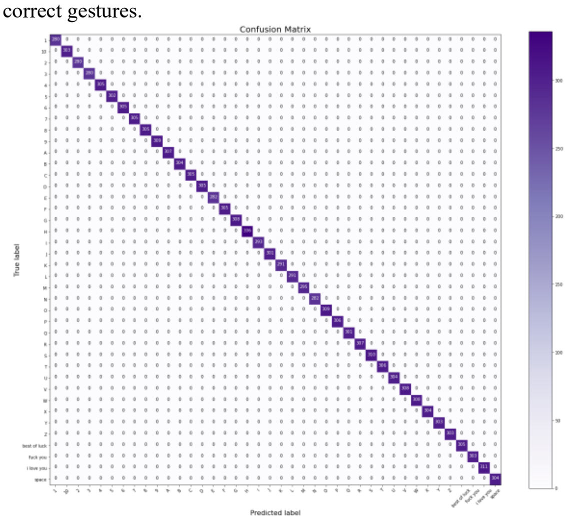
Performance Analysis

To evaluate the performance of the proposed sign language recognition system, several key classification metrics were computed:

Confusion Matrix:

The confusion matrix provides insight into how well the system classifies different gestures:

- **True Positives (TP):** Correctly classified gesture images.
- **False Positives (FP):** Incorrectly classified as the target gesture.
- **True Negatives (TN):** Correctly identified non-target gestures.
- **False Negatives (FN):** Missed correct gestures.

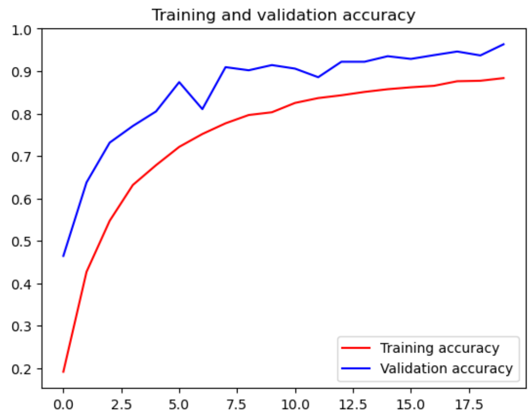


Accuracy:

Represents the proportion of correctly classified gestures out of all samples:

Accuracy=TP+FP / TN+FN + TP+TN

A high accuracy implies that the model correctly predicts the majority of the signs.



Precision:

Indicates the number of correctly predicted positive observations to the total predicted positive observations:

Precision=TP / FP + TP

High precision shows that when the model predicts a specific gesture, it is often correct.

Recall:

Measures the ability of the model to correctly identify all relevant instances:

Recall=TP / FN + TP

A high recall value suggests the model successfully identifies most gestures.

F1-Score:

The harmonic mean of precision and recall:

F1−Score=2× ((Precision x Recall) (Precision + Recall))

This is especially useful when there's an imbalance in the dataset across gesture classes.

Results

- **Accuracy:** 98%
- **Precision:** 98%

- **Recall:** 99%
- **F1-Score:** 98.5%

The system demonstrates high performance across all metrics, indicating that it can reliably classify a wide range of sign gestures in real time. The Temporal Convolutional Network (TCN) and CNN backbone contributed significantly to capturing spatial and temporal features, improving overall gesture recognition accuracy.

RESULTS

The Sign Language Recognition module achieved an accuracy of 98%, confirming its effectiveness in recognizing both static and dynamic hand signs in real-time. The Speech Recognition and Synthesis module delivered over 97% accuracy in translating spoken language into text and generating natural-sounding audio responses. The 3D Avatar Generation module successfully produced real-time animated gestures synchronized with audio and text inputs, enabling visually enriched communication. Additionally, the system supported multilingual sign-speech translation, enhancing its adaptability and usability for users across diverse linguistic backgrounds.

DISCUSSION

These results demonstrate the system's capability to facilitate inclusive communication between hearing and hearing-impaired individuals using AI-powered solutions. The high accuracy, seamless gesture recognition, and natural language processing ensure reliable real-time performance, making it suitable for various domains including education, virtual meetings, and public services. The integration of a realistic avatar for gesture visualization enhances user engagement and comprehension. Furthermore, multilingual support significantly increases the system's accessibility. Future developments should aim at extending sign language support to regional and global variants, incorporating facial expression and emotion recognition for contextual understanding, and refining the user interface based on continuous feedback to ensure a seamless user experience in real-world environments.

V. Conclusion

In conclusion, this project successfully bridges the communication gap between deaf and non-deaf users by integrating advanced Sign Language Recognition, Speech Recognition, and Avatar-Based Visual Communication into a virtual meeting platform. Leveraging the SignNet model powered by Temporal Convolutional Networks (TCNs) for accurate gesture recognition and Hidden Markov Models (HMMs) for translating speech into sign language, the system enables real-time, seamless interaction. Its integration with Jitsi ensures effective virtual meetings, while multilingual interpretation, speech synthesis, and notifications enhance accessibility for diverse users. This inclusive and scalable solution significantly improves virtual communication for the deaf community.

FUTURE ENHANCEMENT

In the future, the system can be enhanced by incorporating support for regional and international sign languages, expanding platform compatibility to include Zoom, Microsoft Teams, and Google Meet, and refining the avatar module with realistic 3D animations, facial expressions, and lip-sync capabilities. Developing a mobile application will increase accessibility across devices, while adding automated meeting transcription and summarization using AI can further enhance usability. These enhancements will contribute to a more comprehensive, user-friendly, and inclusive communication experience.

VI. REFERENCE

[1] S. Shrenika, M. M. Bala, "Sign Language Recognition Using Template Matching Technique," 2021 IEEE International Conference on Communication, Computing and Electrical Systems (ICCCSEA), Chennai, India, 2021, pp. 1-6.

[2] D. Talukder, F. Jahara, "Real-Time Bangla Sign Language Detection with Sentence and Speech Generation," 2020 IEEE International Conference on Electrical, Computer and Communication Technologies (ICECIT), Dhaka, Bangladesh, 2020, pp. 6-11.

[3] D. S. Dhivyasri, K. B. Krishnaa Hari, V. Krishnaveni, "An Efficient Approach for Interpretation of Indian Sign Language using Machine Learning," 2021 International Conference on Signal Processing and Communication (ICSPC), Chennai, India, 2021, pp. 11-20.

[4] H. S. Anupama and B. A. Usha, "Automated Sign Language Interpreter Using Data Gloves," 2021 2nd International Conference on Advances in Electronics, Computers, and Communications (ICAECC), Bangalore, India, 2021, pp. 20-29.

[5] M. M. Chandra and R. S. Kumar, "Sign Languages to Speech Conversion Prototype using the SVM Classifier," 2019 IEEE Region 10 Conference (TENCON), Kochi, India, 2019, 29-40.

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

[7] G. T. R. Chiranji Lal Chowdhary and B. D. Parameshachari, Computer Vision and Recognition Systems: Research Innovations and Trends, CRC Press, 2022.

[8] M.M. Riaz and Z. Zhang, "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, vol. 1467, 2021.

[9] G. Halvardsson, J. Peterson, C. Soto-Valero and B. Baudry, "Interpretation of Swedish sign language using convolutional neural networks and transfer learning", SN Computer Science, vol. 2, no. 3, pp. 1-15, 2021.

[10] P. Likhar, N. K. Bhagat and R. G N, "Deep Learning Methods for Indian Sign Language Recognition", 2020 IEEE 10th International Conference on Consumer Electronics (ICCE-Berlin), pp. 1-6, 2020.

[11] F. Li, K. Shirahama, M. A. Nisar, X. Huang and M. Grzegorzek, "Deep Transfer Learning for Time Series Data Based on Sensor Modality Classification", Sensors, vol. 31, no. 20(15), pp. 4271, Jul 2020.

[12] J. J. Bird, A. Ekárt and D. R. Faria, "British sign language recognition via late fusion of computer vision and leap motion with transfer learning to american sign language", Sensors, vol. 20, no. 18, pp. 5151, 2020.

[13] S. Sharma, R. Gupta and A. Kumar, "Trbaggboost: an ensemble-based transfer learning method applied to Indian Sign Language recognition", J Ambient Intell Human Comput Online First, 2020, [online] Available: https://doi.org/10.1007/s12652-020-01979-z.

[14] M. Oudah, A. Al-Naji and J. Chahl, "Hand Gesture Recognition Based on Computer Vision: A Review of Techniques", *J. Imaging*, vol. 6, no. 73, 2020.

[15] Z. M. Shakeel, S. So, P. Lingga and J. P. Jeong, "MAST: Myo Armband Sign-Language Translator for Human Hand Activity Classification", IEEE International Conference on Information and Communication Technology Convergence, pp. 494-499, 2020.

[16] M. Zakariya and R. Jindal, "Arabic Sign Language Recognition System on Smartphone", 2019 10th International Conference on Computing Communication and Networking Technologies (ICCCNT), pp. 1-5, 2019.

[17] E. Abraham, A. Nayak and A. Iqbal, "Real-Time Translation of Indian Sign Language using LSTM", 2019 Global Conference for Advancement in Technology (GCAT), pp. 1-5, 2019.

[17] O. Koller, N. C. Camgoz, H. Ney and R. Bowden, "Weakly supervised learning with multi-stream CNN-LSTM-HMMs to discover sequential parallelism in sign language videos", IEEE Trans. Pattern Anal. Mach. Intell., vol. 42, no. 9, pp. 2306-20, 2019.

[18] A. A. Hosain, P. S. Santhalingam, P. Pathak, J. Kosecka and H. Rangwala, "Sign language recognition analysis using multimodal data", 2019.

[19] J. Huang, W. Zhou, Q. Zhang, H. Li and W. Li, "Video-based sign language recognition without temporal segmentation", AAAI, vol. 32, no. 1, pp. 2257-64, 2018.

[20] D. Konstantinidis, K. Dimitropoulos and P. Daras, "Sign language recognition based on hand and body skeletal data", 3DTV-CON, pp. 1-4, 2018.