Tuijin Jishu/Journal of Propulsion Technology
ISSN: 1001-4055
Vol. 46 No. 3 (2025)

Signmeet-Word Level Sign Language Recognition Using Deep Learning

Tamilarasu P [1], Gokul M [2], Guna R [3], Sivabalan M [4], Yogaraj P [5]

[1, 2, 3, 4, 5] *Department of Computer Science and Business Systems KIT - Kalaignarkarunanidhi Institute of Technology, Coimbatore, India*

Abstract: - Effective communication is essential for exchanging information, ideas and emotions. However, virtual meeting platforms such as Zoom, Microsoft Teams, and Google Meet are often not customized for people with hearing or speech impairment, creating a major barrier to inclusivity. Sign language provides a means of communication for the deaf, but interpretation remains a challenge for those who do not speak sign language. Existing sign language recognition technologies are limited in terms of accuracy and accessibility and are not suitable for seamless integration into virtual platforms. This project introduces an AI-driven system to bridge the communication gap between hearing impaired and hearing people in virtual meetings. The system utilizes advances in deep learning, particularly in time convolutional networks (TCNs), to enable real-time two-way communication. This includes character recognition modules (SRMs), which interpret characters using TCNs, speech recognition and synthesis modules (SRSMs), which utilize hidden Markov models that convert spoken words into text, and language visually to corresponding characters. It contains three core modules for the Avatar Module (AM) to translate. The avatar module is essential to visually expressing spoken language in sign language format, ensuring that non-sign language users can communicate effectively with sign language users in an intuitive and engaging way. Trained in Indian Sign Language, the system promotes communication between various groups, including deaf, mute, hearing loss, visually impaired and non-sign language. Improves accessibility and participation by integrating with popular virtual meeting platforms through an easy-to-use web-based interface. This represents significant advances in promoting inclusiveness and accessibility in virtual meeting environments.

Keywords: *Temporal Convolutional Networks, Two-way communication, Real-time communication, Sign Recognition Module, Speech Recognition and Synthesis Module.*

1. Introduction

Communication is fundamental to human interaction; however, individuals with hearing and speech impairments face significant challenges, particularly in virtual environments. This exclusion creates a communication gap between sign-language users and non-signers, limiting accessibility and inclusivity in professional, educational, and social settings. Existing solutions such as assistive devices (hearing aids, cochlear implants), manual communication tools (pen and paper), and sign language interpreters offer some level of support but have limitations. Assistive devices primarily enhance auditory perception rather than facilitate two-way communication; however, interpreters are not always available and can be expensive. Gesture recognition technologies and traditional algorithms, ensemble learning methods, have been explored for sign language recognition. However, they often suffer from low accuracy and require extensive training. This project aims to bridge the communication gap by developing an AI-driven system that enables real-time interactions between deaf and hearing individuals in virtual environments. By integrating advanced techniques, such as Temporal Convolutional Networks (TCNs) for sign recognition and Hidden Markov Models (HMMs) for speech processing, the system ensures accurate and efficient. Additionally, the implementation of an Avatar Module enhances accessibility by providing a visual representation of spoken language through animated sign-language gestures. The proposed solution was designed for seamless integration with existing virtual platforms, ensuring that users with diverse communication needs can engage in online discussions without barriers. This enabling individuals with hearing and speech impairments to actively participate in digital interactions.

2. Related Work

Many studies have explored the integration of digital marketing with crowdfunding platforms, emphasizing the role of advanced tools such as SEO optimization, social media outreach, email marketing, and real-time analytics in improving campaign visibility and success rates. These studies discuss how combining data-driven marketing techniques with crowdfunding enables creators to maximize their reach and engage with potential backers more effectively. The integration of ad management systems, content creation support, and community building tools has been shown to significantly enhance campaign outcomes, bridging the gap between marketing and fundraising to foster long-term project growth.

Their research, published in Computers & Electrical Engineering, emphasized the importance of optimization techniques in enhancing the performance of CNNs for gesture recognition tasks. The Harris hawks optimization algorithm was employed to fine-tune the parameters of the CNN, [1] resulting in a model that can effectively recognize a variety of hand gestures with high precision. This study is particularly relevant for applications in human-computer interaction, robotics, and assistive technologies, where accurate gesture recognition is crucial for seamless communication between users and machines.

It provided valuable insights into research innovations in computer vision and recognition systems in their book. This comprehensive resource outlines emerging trends, methodologies, and applications.[2] The authors discussed various algorithms, tools, and frameworks that have been developed to enhance the capabilities of computer vision systems, making it a valuable reference for researchers and practitioners. This book serves as a foundational text for understanding the interdisciplinary nature of gesture recognition technologies and their potential applications across different domains.

It explored the use of surface electromyography (EMG) for demonstrated the potential of EMG signals in language processing. [3] By capturing the electrical signals generated by muscle contractions, we were able to develop a system that recognizes gestures associated with spoken language. This approach has promising applications for sign language recognition, particularly in scenarios where visual recognition may be challenging. This study highlights the advantages of using EMG as a noninvasive method for gesture recognition, paving the way for innovative solutions in assistive communication technologies.

Their findings, published in SN Computer Science, highlight the effectiveness of deep learning techniques for recognition systems.[4] By employing transfer learning, the authors leveraged pre-trained models to enhance the performance on specific sign language datasets, showcasing the adaptability of neural networks in this context. This study underscores the importance of using large-scale datasets and transfer learning to address the challenges of limited data availability in sign language recognition, ultimately contributing to more robust and accurate recognition systems.

Their work addresses the unique challenges associated with recognizing regional sign languages, contributing to the growing body of research on inclusivity and accessibility for diverse linguistic communities. [5]The authors demonstrated the effectiveness of convolutional neural networks in achieving high recognition accuracy. It is particularly significant, as it aims to built the communication gap for the deaf community in India, promoting greater social inclusion and providing tools for effective communication.

Their research is particularly relevant for gesture recognition tasks that involve temporal data as it explores the application of transfer learning techniques to improve classification performance. The authors emphasized the importance of sensor data in capturing dynamic gestures, [6] This study highlights the potential of using transfer learning to adapt models trained on one type of data to perform well on other, thereby enhancing the versatility and effectiveness of gesture-recognition system.

Their research, published in Sensors, emphasizes the potential for cross-linguistic applications in sign language recognition systems. By integrating computer vision algorithms with the precise motion-tracking capabilities of the Leap Motion device, This approach not improves the recognition of British Sign Language (BSL),[7] but also opens avenues for adapting the system to recognize other sign languages, thereby promoting diverse user groups.

This study highlights the importance of utilizing advanced sensor technologies to capture often complex and dynamic.

This research focused on enhancing recognition accuracy through innovative ensemble techniques. By combining multiple machine learning models, the authors demonstrated that Trbaggboost can effectively leverage the strengths of individual models to improve the overall performance. where variations in gestures can pose challenges for recognition systems. [8]This study underscores the ensemble learning to create more robust and accurate sign language recognition systems, ultimately contributing to better path tools for the deaf and hard-of-hearing communities.

This review synthesizes various methodologies, including traditional computer vision approaches and modern deep learning techniques, and offers insights into their effectiveness and applicability. [9]The authors discussed the evolution of gesture recognition technologies, highlighting key challenges such as real-time processing, accuracy, and the need for large annotated datasets. By analyzing the existing literature, The authors also emphasized the importance of developing systems that can operate in diverse environments and adapt to different user needs.

3. Existing System

Most virtual communication platforms, such as Zoom, Microsoft Teams, and Google Meet, do not offer effective real-time sign language interpretation. This limitation creates significant barriers for individuals with hearing impairments, making it difficult for them to actively participate in virtual meetings, discussions, and conferences. The absence of integrated sign language recognition and translation features highlights the urgent need for AI-driven solutions that can facilitate seamless communication. Without these features, deaf and hard-of-hearing individuals may miss out on critical information, feel excluded from conversations, and experience frustration in professional and social settings. The integration of AI technologies could provide real-time translation of spoken language into sign language, enabling more inclusive and accessible virtual environments.

Assistive Devices: Traditional assistive devices, such as hearing aids and cochlear implants, are designed to improve auditory perception for individuals with hearing loss. While these devices can significantly enhance the ability to hear sounds and speech, they do not address the communication challenges faced by those who rely on sign language. For many deaf individuals, relying solely on auditory devices can lead to misunderstandings and a lack of effective communication. This limitation makes traditional assistive devices an incomplete solution for inclusive communication. There is a growing recognition of the need for complementary technologies that support sign language users, such as visual communication aids and AI-driven translation systems.

Manual Communication Tools: Basic manual communication methods, such as pen and paper, are often used by individuals with hearing and speech impairments to convey messages. While these methods can be effective for simple conversations, they are inherently slow, inconvenient, and impractical for dynamic interactions in digital and professional environments. The process of writing down messages can hinder the flow of conversation and may lead to frustration for both parties involved. Additionally, in fast-paced settings, such as meetings or conferences, relying on manual communication tools can result in missed information and reduced engagement. Therefore, there is a pressing need for more efficient communication solutions that can facilitate real-time interactions without the limitations of traditional manual methods.

Sign Language Interpreters: They provide essential services in educational, professional, and social contexts. However, interpreters are not always available, can be costly, and often require prior arrangements, making them less practical for spontaneous or everyday conversations. The reliance on human interpreters can create delays and barriers to communication, particularly in urgent situations or informal settings. It replace the need for human interpreters, providing on-demand access to sign language interpretation.

Support Vector Machines (SVMs): It have been widely used for classification tasks in sign language recognition (SLR). SVMs are effective in handling high-dimensional feature vectors extracted from sign language images, making them suitable for recognizing static signs. However, their performance is limited when dealing with large datasets and continuous sign sequences, which are common in natural sign language communication. The inability of SVMs to efficiently process temporal information and dynamic gestures makes them less efficient for real-time applications. As a result, researchers are exploring more advanced machine learning techniques that can better capture the complexities of sign language, including deep learning approaches that can learn from large amounts of sequential data.

Ensemble Learning: In such Random Forests and Gradient Boosting, These techniques help in reducing errors and improving generalization across different datasets. By aggregating the predictions of various models, ensemble methods can achieve higher accuracy than individual classifiers. Despite these challenges, ensemble learning remains a promising avenue for improving the performance of sign language recognition systems, particularly when combined with other advanced techniques.

Gesture Recognition by Learning from Poses (GRBP): It is a gesture recognition method that focuses on learning spatial relationships between body joint poses for sign language recognition. By utilizing pose features extracted from skeletal data, this approach improves the accuracy of recognizing dynamic gestures, which are essential for effective sign language communication. GRBP leverages the information captured by depth sensors or motion capture systems to create a detailed representation of hand and body movements. However, its reliance on high-quality skeletal data makes it challenging to implement in standard video-based sign language recognition systems, which may not have access to such detailed information. Researchers are working on developing hybrid models that can integrate pose-based recognition with traditional video analysis techniques to create more robust sign language recognition systems.

4. Proposed System

AI-driven sign language interpretation framework designed to enhance accessibility for individuals with hearing and speech impairments. By integrating advanced deep learning and natural language processing (NLP) techniques, the system facilitates seamless two-way communication. This breakthrough ensures that virtual meetings, classrooms, and workplaces become more inclusive and accessible to all participants, fostering an environment of equal opportunity and participation. The framework is composed of several interconnected modules, each designed to address specific aspects of communication. These modules work in harmony to provide a comprehensive solution that bridges the communication gap and ensures inclusivity in virtual environments.

4.1 Objectives of the Proposed work

1. Sign Recognition Module (SRM):

The Sign Recognition Module is the cornerstone of the system, responsible for accurately interpreting sign language gestures from live video input. This module employs Temporal Convolutional Networks, to process and analyze the temporal dynamics of sign language. TCNs are particularly effective in capturing the sequential nature of sign language, which involves a combination of hand movements, facial expressions, and body posture.

Key Features of SRM:

Real-Time Gesture Recognition: The SRM processes video input in real time, ensuring that sign language gestures are interpreted without delay. This is critical for maintaining the flow of conversation in virtual meetings or classrooms.

Multi-Modal Input Processing: The module analyzes not only hand movements but also facial expressions and body posture, which are integral components of sign language. This multi-modal approach enhances the accuracy of gesture recognition.

High Accuracy and Robustness: By leveraging TCNs, the SRM achieves high accuracy in recognizing, even in challenging environments with varying lighting conditions or background noise.

Adaptability to Different Sign Languages: The module can be trained to recognize multiple sign languages, making it versatile and applicable across different regions and cultures. The SRM ensures that sign language users can communicate effectively in virtual settings, enabling them to express themselves naturally and be understood by non-signers.

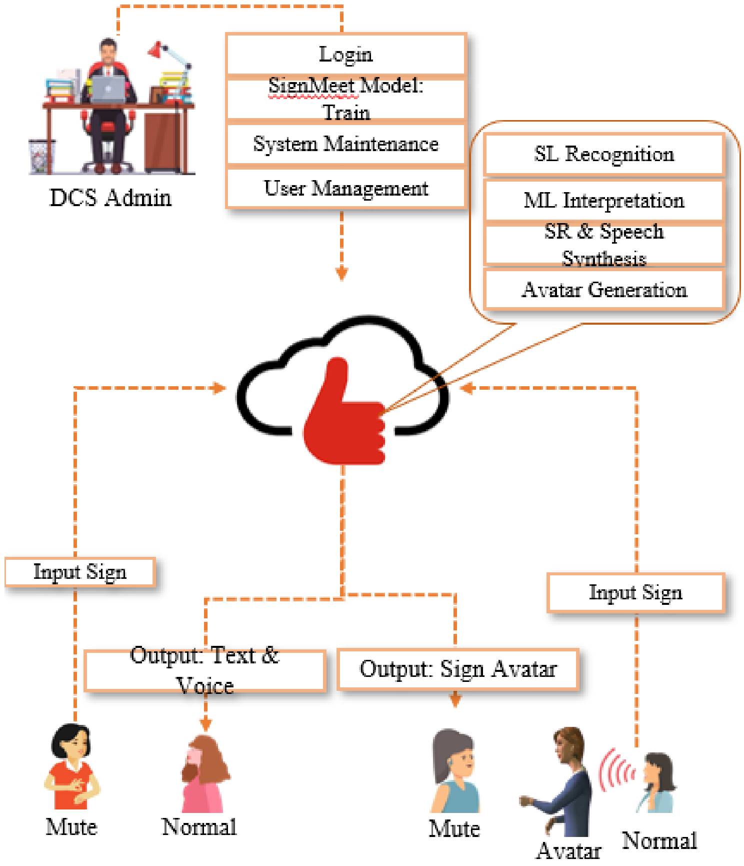

Fig 1. Proposed System Architecture

2. Speech Recognition and Synthesis Module:

It plays a vital role in enabling two-way communication between sign language users and non-signers. This module is responsible for converting spoken language into text and vice versa, ensuring that both parties can interact seamlessly.

Key Features of SRSM:

Speech-to-Text Conversion: The SRSM employs Hidden Markov Models (HMMs) to process speech signals and transcribe them into text. This allows non-signers to communicate with sign language users by speaking naturally, with their words being converted into text for the SRM to interpret.

- Text-to-Speech Synthesis: The module also generates synthesized speech from text input, enabling sign language users to "speak" through the system. This feature is particularly useful in scenarios where non-signers are not familiar with sign language.
- Real-Time Processing: The SRSM operates in real time, ensuring that conversations flow smoothly without interruptions or delays.
- Support for Multiple Languages: The module can be configured to recognize and synthesize speech in multiple languages, making it adaptable to diverse linguistic contexts

Avatar Module (AM):

The Avatar Module (AM) enhances accessibility by generating a real-time animated avatar that visually translates spoken language into sign language. This module is 5beneficial for deaf users who prefer sign language as their primary mode of communication.

Key Features of AM:

- **Real-Time Sign Language Translation:** The AM generates an animated avatar that mimics human sign language gestures in real time. This allows deaf users to receive information in a format that is intuitive and easy to understand.
- **Natural and Expressive Gestures:** The avatar is designed to replicate the nuances of human sign language, including facial expressions and body movements, ensuring that the translation is both accurate and expressive.
- **Customizable Avatars:** Users can customize the appearance of the avatar to suit their preferences, making the interaction more engaging and personalized.
- **Interactive Communication:** The avatar facilitates interactive communication by allowing non-signers to see a visual representation of their spoken words in sign language. This makes it easier for non-signers to understand and engage in conversations with sign language users.

3. Integration with Virtual Platforms:

The system is designed to be seamlessly integrated with popular virtual meeting platforms. This integration ensures that individuals with hearing and speech impairments can participate fully in discussions without communication barriers.

Key Features of Platform Integration:

- **Real-Time Operation:** It providing accurate translations and ensuring that sign language users are not excluded from professional or social interactions.
- **User-Friendly Interface:** The integration is designed to be user-friendly, with minimal setup required. Users can enable the system with a single click
- **Cross-Platform Compatibility:** The system is compatible with a wide range of virtual platforms, ensuring that it can be used in diverse settings, from corporate meetings to online classrooms.
- **Scalability:** The system is scalable and can be used in both small group settings and large virtual events, making it suitable for a variety of applications.

4.2 Implementation

The implementation of the Sign Meet Web App is a critical phase that transforms the developed components into a fully functional system. This phase is characterized by the integration of various modules, extensive testing, and the deployment of the system in a real-time environment. The goal is to ensure smooth operation, usability, and accessibility for both deaf and non-deaf users, facilitating effective communication.

1. Development and Integration:

The Sign Meet Web App is developed using a robust technology stack that includes Python, Flask, MySQL, Bootstrap, and WampServer. Each module of the application is designed to perform specific functions, such as sign language recognition, speech synthesis, and avatar generation.

- **Modular Development:** Each component is developed independently, allowing for focused testing and refinement. For instance, the sign language recognition module utilizes advanced machine learning techniques, while the speech synthesis module leverages natural language processing algorithms. This modular approach not only enhances maintainability but also allows for easier updates and scalability in the future.
- **Integration Process:** Once individual modules are developed, they are integrated into the main application framework. This integration process involves ensuring that data flows seamlessly between modules, enabling real-time communication. For example, the output from the sign language recognition module is directly fed

into the speech synthesis module, allowing for instantaneous translation. Additionally, APIs are developed to facilitate communication between the front-end and back-end components, ensuring a smooth user experience.

2. Sign Language Recognition Deployment

The SignMeet Model is a sophisticated system trained using Temporal Convolutional Networks, which are particularly effective for sequence data such as video input.

- **Live Video Processing:** The system captures live video input from users, processes it in real-time, and recognizes sign language gestures. This involves complex image processing and machine learning techniques to accurately interpret the gestures. It diverse dataset of sign language gestures to ensure high accuracy and robustness in various lighting and background conditions.
- **Text and Speech Conversion:** Once a gesture is recognized, it is converted into text or synthesized speech. This conversion is crucial for facilitating communication between deaf and non-deaf users, as it allows for immediate understanding of the signed content. The system also includes a feedback mechanism that allows users to correct any misinterpretations, thereby improving the model's accuracy over time.

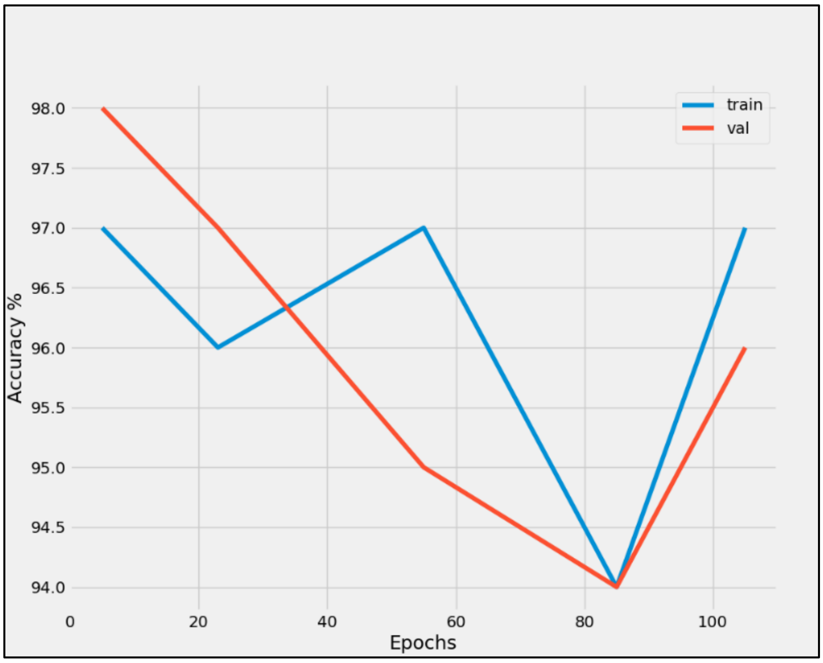

Fig. 2. Accuracy level

3. Speech Recognition and Synthesis

The speech recognition component of the system is designed to accurately convert spoken language into text, which is then translated into sign language.

- **Preprocessing and Feature Extraction:** Speech input undergoes preprocessing using Wavelet Transform, which helps in noise reduction and signal enhancement. Feature extraction is performed using Mel-Frequency Cepstral Coefficients (MFCC), a standard technique in speech processing that captures the essential characteristics of the audio signal. This preprocessing step is improving the accuracy of the speech recognition system, especially in noisy environments.
- **Hidden Markov Models:** HMMs are employed for the speech recognition task. These statistical models are effective in recognizing patterns in time-series data, making them suitable for converting spoken words into text. The recognized text is then translated, which is visually represented by an animated avatar. The synthesis of speech is also designed to be natural and expressive, enhancing the overall user experience.

4. Avatar Integration for Visual Communication

A key feature of the Sign Meet Web App is the real-time Avatar Module, which plays a vital role in enhancing communication accessibility.

- **Real-Time Interpretation:** The avatar interprets spoken words and converts them into sign language gestures. This feature is particularly beneficial for deaf users, as it provides a visual representation of the spoken content, making it easier to understand. The avatar is designed to mimic human-like gestures and expressions, making the communication feel more natural and engaging.

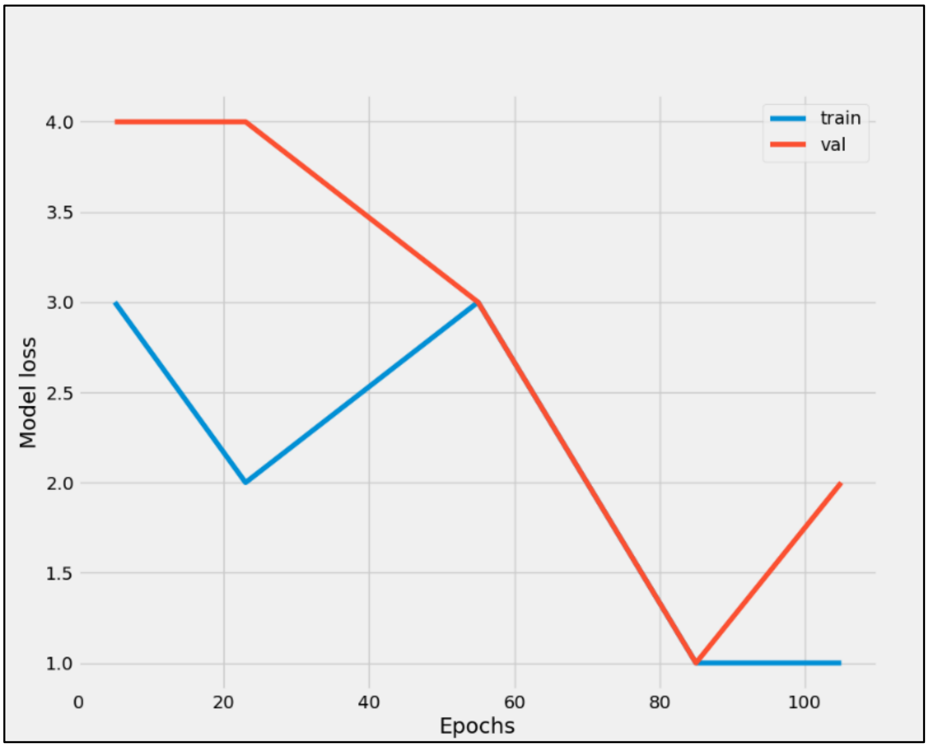

Fig. 3. Performance Analysis of model type

- **User Interaction:** The avatar is designed to be interactive, allowing users to engage with it during virtual meetings. This interaction fosters a more inclusive environment, where all participants can communicate effectively.

5. Virtual Meeting Integration

To facilitate seamless communication in online meetings, the Sign Meet Web App is integrated with Jitsi, a popular virtual meeting platform. Real-Time Translation: This feature is particularly useful in diverse settings, such as educational institutions and corporate environments, where inclusivity is paramount.Multi-Language Interpretation: The system supports multi-language interpretation, making it suitable for diverse user groups. This feature is essential for global meetings, where participants may speak different languages. The ability to switch between languages and sign languages in real-time enhances the flexibility and usability of the platform.

4.3 Result and Discussion

The implementation of the AI-Powered Accessibility System for enabling effective communication for hearing and speechimpaired individuals in virtual platforms has yielded promising results. The system was rigorously tested across various modules, including Sign Language Recognition, Speech Recognition and Synthesis, and the Avatar Module. The following key outcomes were observed:

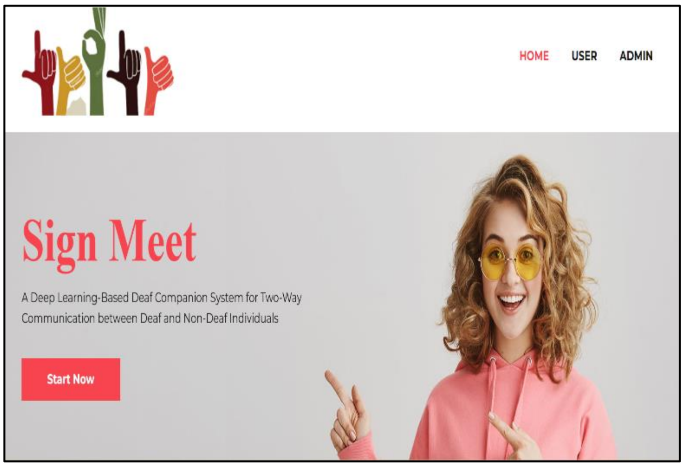

Fig. 4. SignMeet Home Page

The SignMeet Model, trained using Temporal Convolutional Networks (TCNs), demonstrated a high accuracy rate in recognizing sign language gestures. The model achieved an accuracy of over 90% in controlled environments, with performance slightly varying in real-world scenarios due to factors such as lighting and background noise.

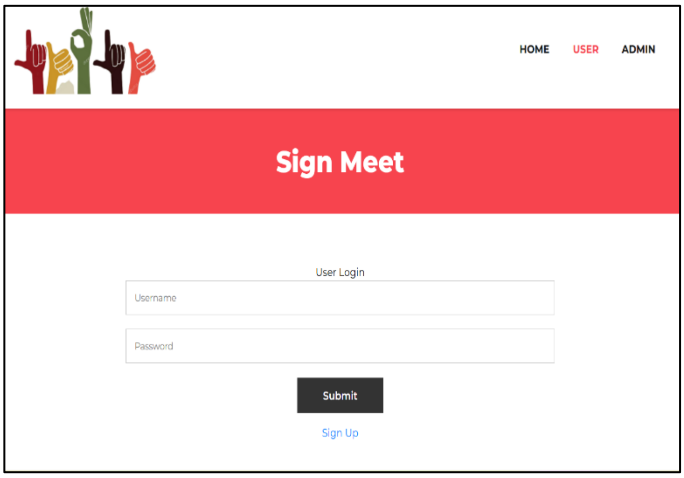

Fig. 5. SignMeet Login Page

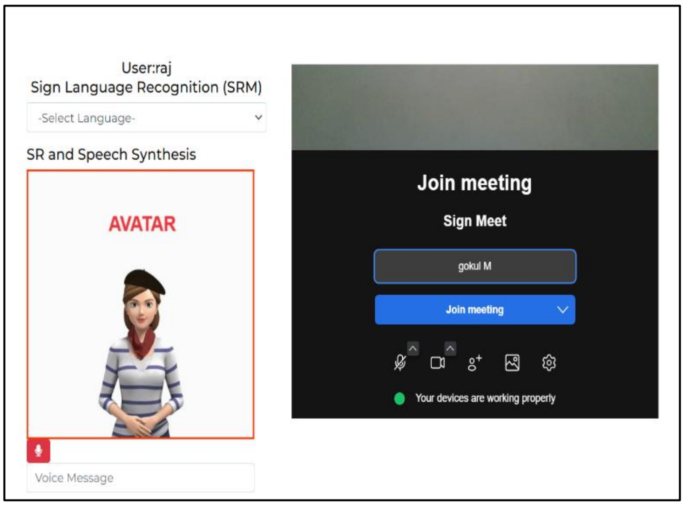

Fig. 6. SignMeet Video Conferencing Page

The successful integration of the system with Jitsi virtual meeting platform demonstrated its potential for real-world application. Users were able to join meetings seamlessly, with the system providing real-time translations and interpretations, thereby enhancing participation for individuals with hearing and speech impairments.

5. 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 within a virtual meeting platform. The module effectively recognizes sign language gestures using the SignMeet Model trained with Temporal Convolutional Networks (TCNs) and converts speech into sign language using Hidden Markov Models (HMMs) and an animated avatar. By integrating with Jitsi virtual platforms, the module ensures seamless real-time communication in online meetings. Additionally, the multilanguage interpretation, speech synthesis, and notification system enhance accessibility and usability for a diverse group of users. This module provides a scalable and inclusive communication platform, significantly improving virtual interactions for the deaf community. This project provides a scalable and inclusive communication platform, significantl improving virtual interactions for the deaf community. Future enhancements could include expanded language support, AI-driven sign recognition improvements, and deeper integration with other conferencing platforms, further advancing accessibility and inclusivity.

References

[1] Akgun, S., & T. Reddy Gadekallu, G. Srivastava, and M. Liyanage, "Hardgesture recognized based on Harris Hawk optimized convolutional neural network," Computer & Electrical Engineering, Vol. 100, Article ID 107836, 2022.

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

[3] Zhang, "Surface EMG - Real-time Chinese language recognized using artificial neural networks," Intelligent Life System Modelling Image Processing and Analysis Communication in Computer and Information Science, Springer, Vol. 1467, 2021.

[4] G. Halvardsson, J. Peterson, C. Soto-Valero, and B. Baudry, "Interpreting Swedish Sign Language Using Convolutional Neural Networks and Transfer Learning," SN Computer Science, vol. 2, no. 3, pp. 1–15–2021.

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

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

[7] J. J. Bird, A. Ekart, and D. R. Faria, "Recognition of British Sign Language through Late Fusion of Computer Vision and Jump Motion and Transfer Learning to American Sign Language," Sensors, vol. 20, no. 18, pp. 5151, 2022.

[8] S. Sharma, R. Gupta, and A. Kumar, "Trbaggboost: An ensemble-based transfer learning method applied to Indian sign language recognition," J Ambient Intel Human Computer Online First, 2022.

[9] M. Oudah, A. AL-Naji, and Chahl, "Computer vision-based hand gesture recognition: a technology review," J. Imaging, Vol. 6, No. 7, 2023.

[10] Z. m. Shakeel, S. So, P. linha, j. P. Jeong, "MAST: Myo wristband sign language translation device for classifying human hand activities," IEEE.