International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 12 | Dec - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

Smart Mern Based Video Conferencing with Sign and Voice Assisstance

Sasikala P- Assistant Professor, Department of Computer Science and Engineering, Sri Shakthi Institute of Engineering and Technology

Agalya E, Anbuselvi M, Bhavishnu S, Hema J, Jayendhran S -UG Students, Department of Computer Science and Engineering, Sri Shakthi Institute of Engineering and Technology

ABSTRACT

An inventive web-based communication tool called the Video Meeting Application using MERN Stack was created to encourage accessibility and inclusivity for both regular and speech/hearing-impaired users. Conventional video conferencing services, like Zoom or Google Meet, frequently require external tools or interpreters because they do not provide integrated support for those with communication challenges. In order to close that gap, this project combines WebRTC-based real-time video conferences with capabilities like text-to- speech voice conversion, speech-to-text subtitles, and sign language detection. The solution, which was developed using the MERN Stack (MongoDB, Express, React, and Node.js), allows users to create or join meeting rooms with secure JWT authentication and easily communicate via voice, text, and gestures. Using Media Pipe and TensorFlow.js, the system can recognize simple hand movements, translate them into readable text and voice for other participants, and use the Web voice API to simultaneously translate spoken audio into live captions for users with hearing impairments.

Keywords

Accessibility, Sign Language Detection, Speech- to-Text, Text-to-Speech, Real-time Communication, Video Conferencing.

I. INTRODUCTION

Virtual conference tools like Zoom, Google Meet, and Microsoft Teams have revolutionized communication in the current digital era, but accessibility for people with hearing and speech impairments is still restricted. A video meeting application utilizing the MERN stack is created as a real-time, intelligent, and inclusive communication platform in order to close this gap. The system incorporates text-to-speech conversion, speech-to-text subtitles, and sign language identification to facilitate seamless communication between people with and without disabilities. It guarantees scalability, security, and responsive performance because it was built with Express.js, React.js, Node.js, and MongoDB.

While Node.js and Express.js handle meetings, authentication, and real-time communication via Socket.IO, React.js offers an easy-to-use interface. Gesture detection is handled by TensorFlow.js or MediaPipe, live subtitles are made possible using the Web Speech API, and low-latency video and audio streaming is guaranteed by WebRTC. The platform provides an affordable browser-based solution that encourages inclusion, accessibility, and smooth digital communication.

By combining all functions into a single platform, the system also reduces reliance on other tools. It improves equitable participation in online social, intellectual, and professional interactions.

II. LITERATURE REVIEW

[2.1] Richard Kennaway; John R. W. Glauert; I. Zwitserlood; "Providing Signed Content on The Internet By Synthesized Animation", ACM TRANS. COMPUT. HUM. INTERACT., 2007.

Written information is often of limited accessibility to deaf people who use sign language. The eSign project was undertaken as a response to the need for technologies enabling efficient production and distribution over the Internet of sign language content. By using an avatar-independent scripting notation for signing gestures and a client-side web browser plug-in to translate this notation into motion data for an avatar, we achieve highly efficient delivery of signing, while avoiding the inflexibility of video or motion capture.

[2.2] Matjaz Debevc; Primoz Kosec; Andreas Holzinger; "Improving Multimodal Web Accessibility for Deaf People: Sign Language Interpreter Module", MULTIMEDIA TOOLS AND APPLICATIONS, 2010.

The World Wide Web is becoming increasingly necessary for everybody regardless of age, gender, culture, health and individual disabilities. Unfortunately, there are evidently still problems for some deaf and hard of hearing people trying to use certain web pages. These people require the translation of existing written information into their first language, which can be one of many sign languages.

[2.3] Ganesh Ananthanarayanan; Paramvir Bahl; Peter Bodík; Krishna Chintalapudi; Lenin Ravindranath; Sudipta Sinha; "Real-Time Video Analytics: The Killer App for Edge Computing", COMPUTER, 2017.

Video analytics will drive a wide range of applications with great potential to impact society. A geographically distributed architecture of public clouds and edges that extend down to the cameras is the only feasible approach to meeting the strict real-time requirements of large-scale live video analytics.

[2.4] Larwan Berke; Khaled Albusays; Matthew Seita; Matt Huenerfauth; "Preferred Appearance of Captions Generated By Automatic Speech Recognition for Deaf and Hard-of-Hearing Viewers", EXTENDED ABSTRACTS OF THE 2019 CHI CONFERENCE ON HUMAN ..., 2019.

As the accuracy of Automatic Speech Recognition (ASR) nears human-level quality, it might become feasible as an accessibility tool for people who are Deaf and Hard of Hearing (DHH) to transcribe spoken language to text. We conducted a study using in-person laboratory methodologies, to investigate requirements and preferences for new ASR-based captioning services when used in a small group meeting context.

[2.5] Raja S. Kushalnagar; "A Classroom Accessibility Analysis App for Deaf Students", THE 21ST INTERNATIONAL ACM SIGACCESS CONFERENCE ON ..., 2019.

Deaf and hard of hearing (DHH) individuals do not have equal access to audio information in most educational settings, even with visual translation accommodations such as sign language interpreters or captioners.

III. METHODOLOGY

3.1 EXISTING SYSTEM

Existing video conferencing systems such as Zoom, Google Meet, and Microsoft Teams provide powerful real-time communication features including video calls, audio conferencing, screen sharing, and instant chat, making them widely adopted for professional, academic, and personal use. However, despite their technical advancements, these platforms fall short when it comes to accessibility for individuals with hearing or speech impairments. Most existing systems are primarily designed for normal users and do not include dedicated features to support sign language communication or gesture-based interaction. As a result, differently-abled users often depend on external assistive tools, human interpreters, or separate speech-to-text applications, which are not only inconvenient but also inefficient in real-time conversations.

Although some platforms provide basic captioning features, they are often limited in accuracy, language support, and real-time responsiveness. Additionally, these systems lack AI-based sign language recognition and do not support gesture-to-text or gesture-to-speech conversion, making it difficult for speech- impaired users to communicate independently. Standalone tools for speech-to-text conversion and sign language translation exist, but they operate separately and are not integrated into mainstream video conferencing applications. This fragmented approach leads to delays, miscommunication, and reduced user engagement, especially during live meetings or interactive discussions.

Furthermore, existing video conferencing platforms do not offer two-way accessibility communication, where both hearing and speech- impaired users can interact seamlessly with normal users in real time.

3.2 PROPOSED SYSTEM

The proposed system, titled "AI-Powered Accessible Video Meeting Application using MERN Stack," is developed to address the accessibility challenges present in conventional video conferencing platforms by offering an inclusive, intelligent, and real-time communication solution for both normal users and individuals with hearing or speech impairments. The primary goal of the system is to create a unified video meeting environment where all users can communicate effectively without relying on external translation tools or manual assistance. By integrating advanced web technologies and artificial intelligence, the proposed system ensures seamless interaction through video, voice, text, and gestures.

A key feature of the proposed system is its AI- powered accessibility module, which supports communication for hearing- and speech- impaired users. Using TensorFlow.js or MediaPipe, the system captures hand gestures through the user's webcam and processes them in real time. The recognized gestures are converted into readable text and further transformed into audible speech using the Text- to-Speech (TTS) functionality of the Web Speech API. This enables speech-impaired users to communicate naturally with normal users during live meetings. Conversely, when a normal user speaks, the system captures the audio input and converts it into real-time

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 12 | Dec - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

subtitles using Speech-to-Text (STT) technology, allowing hearing-impaired users to read and understand the conversation instantly. platform depends.

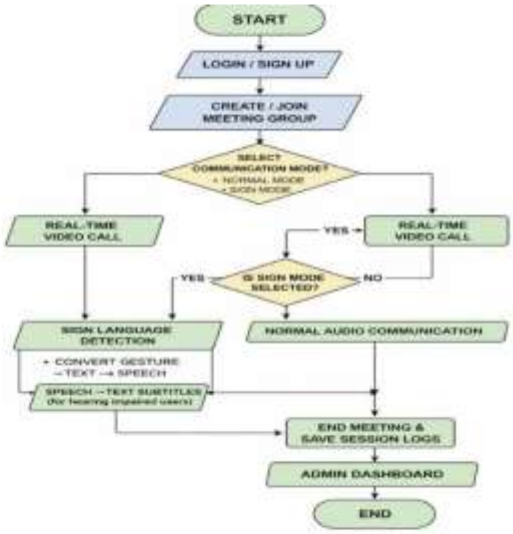


FIG 3.3 FLOW DIAGRAM

IV SYSTEM SPECIFICATION

4.1 SOFTWARE REQUIREMENTS

- MERN Stack

Frontend : React.js, Tailwind CSS, React Router, Tensorflow.js,WwbRTC Backend : Node.js, Express.js Socket.io

Email Services : NodeMailer

- MongoDB

4.2 SOFTWARE DESCRIPTION

4.2.1 MERN STACK

The MERN Stack is a popular set of technologies used to build full-stack web applications. It includes MongoDB, Express.js, React.js, and Node.js. MongoDB serves as the database, Express.js and Node.js handle the backend/server-side logic, while React.js manages the frontend/user interface. All components use JavaScript, which allows for smooth integration and a unified development experience across the entire application stack. The MERN Stack is widely used for developing scalable, dynamic, and high-performance web applications. The MERN Stack is widely used for developing scalable, dynamic, and high-performance web applications. It offers an end-to- end JavaScript environment, improving development efficiency and consistency. And also it has strong community support and a wide range of libraries, making it ideal for modern web development.

4.2.2 REACT. J S

React.js is a popular JavaScript library developed by Facebook for building user interfaces, particularly single-page applications where real- time updates and dynamic content are crucial. It uses a component-based architecture that allows developers to create reusable UI components, enhancing code modularity and maintainability.

4.2.3 TAILWIND CSS

Tailwind CSS is a modern utility-first CSS framework that enables developers to build custom user interfaces quickly and efficiently. Unlike traditional CSS frameworks that provide pre-designed components, Tailwind offers low- level utility classes that can be composed to create any design directly in the markup.

4.2.4 REACT ROUTER

React Router is a standard routing library for React applications that enables navigation between different components and views. It allows developers to implement dynamic routing, where the URL reflects the application's current state.

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 12 | Dec - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

4.2.5 NODE.JS

Node.js is an open-source, cross-platform JavaScript runtime environment that allows developers to execute JavaScript code outside of a web browser. Built on the V8 engine used by Google Chrome, Node.js is known for its speed and efficiency. It uses an event-driven, non-blocking I/O model, making it ideal for building scalable network applications and real-time systems.

4.2.6 EXPRESS.JS

Express.js is a minimal and flexible Node.js web application framework that provides a robust set of features for building web and mobile applications. It simplifies the process of handling HTTP requests, routing, middleware integration, and server-side logic. Express.js enables developers to create APIs and web servers quickly and efficiently.

4.2.7 MONGODB

MongoDB is a popular NoSQL database that stores data in a flexible, JSON-like format called BSON (Binary JSON). Unlike traditional relational databases, MongoDB does not require a fixed schema, allowing developers to store and manage large volumes of unstructured or semi-structured data with ease.

4.2.8 TENSORFLOW.JS

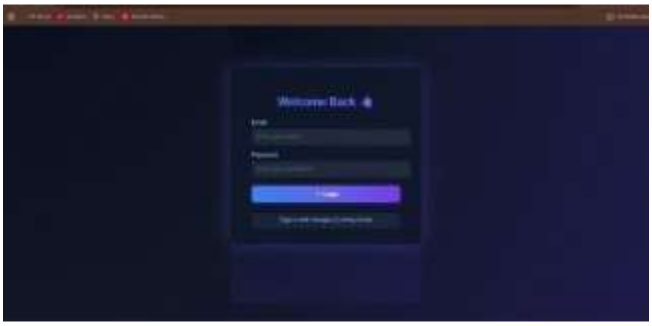
TensorFlow.js is an open-source JavaScript library developed by Google that allows developers to build, train, and run machine learning models directly in the browser or on Node.js.

4.2.9 NODEMAILER

Nodemailer is a module for Node.js that allows applications to send emails easily and efficiently. In this project, Nodemailer is used to automatically send confirmation emails to users after successful event registration and payment.

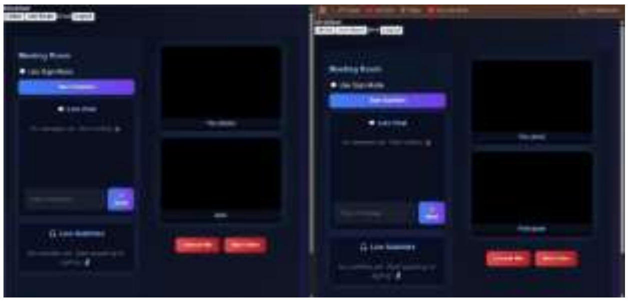
V EXPERIMANTAL RESULT

5.1 TEST CASE 1:

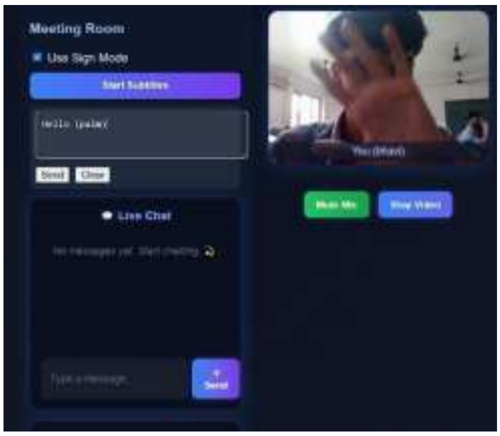


5.2 TEST CASE 2:

5.3 TEST CASE 3:



5.4 TEST CASE 4:



5.5 TEST CASE 5:

VI RESULTS AND DISCUSSION

The result of the AI-Powered Accessible Video Meeting Application using MERN Stack is the successful development of a fully functional, scalable, and inclusive web-based video conferencing platform that effectively bridges the communication gap between normal users and individuals with hearing or speech impairments. The system enables users to securely register, log in, and create or join real- time meeting rooms, ensuring smooth and uninterrupted communication through browser- based video and audio streaming. By integrating WebRTC for peer-to-peer media transmission and Socket.IO for real-time signaling and event handling, the application achieves low-latency, high-quality video conferencing without the need for external plugins or third-party software. One of the most significant outcomes of the project is the successful implementation of AI-driven accessibility features that operate in real time.

These recognized gestures are accurately converted into readable text and further transformed into audible speech using Text-to- Speech (TTS) technology, allowing speech-impaired users to communicate naturally with other participants.

Simultaneously, the application captures spoken input from normal users and converts it into live on-screen subtitles using Speech-to-Text (STT) functionality, enabling hearing-impaired users to follow conversations effortlessly.

The frontend of the application, developed using React.js and Tailwind CSS, provides a clean, modern, and responsive user interface that enhances usability and accessibility across different screen sizes and devices. The backend, powered by Node.js and Express.js, efficiently manages authentication, meeting sessions, and real-time data exchange, while MongoDB ensures reliable storage of user information and session logs. Secure JWT-based authentication safeguards user sessions and prevents unauthorized access, ensuring data privacy and system integrity.

The achieved results validate the feasibility of embedding AI-based gesture recognition and speech translation directly into web applications and highlight the system's potential to significantly enhance digital inclusivity, making virtual meetings more accessible and equitable for differently-abled individuals.

VII CONCLUSION AND FUTURE WORK

7.1 CONCLUSION

In conclusion, the AI-Powered Accessible Video Meeting Application using MERN Stack represents a significant advancement toward creating a more inclusive and accessible digital communication environment. The system successfully integrates artificial intelligence and modern web technologies to bridge the communication gap between normal users and those with hearing or speech impairments. By combining real-time video conferencing using WebRTC with AI-driven features such as sign language detection, speech-to-text subtitle generation, and text-to-speech conversion, the platform ensures smooth, two-way interaction that accommodates the needs of all users.

The implementation of TensorFlow.js and MediaPipe enables real-time gesture recognition directly through the browser, while the Web Speech API ensures accurate voice translation and subtitle generation. The MERN stack architecture (MongoDB, Express.js, React.js, Node.js) provides a robust, scalable, and efficient foundation for the application, ensuring seamless performance and security through JWT authentication. The project not only demonstrates the potential of merging AI with web development but also highlights the social impact of technology in promoting digital inclusivity and accessibility.

7.2 FUTURE SCOPE

The future work for the AI-Powered Accessible Video Meeting Application using MERN Stack focuses on enhancing its accuracy, scalability, and usability to make it a complete and production-ready accessibility platform. One major improvement involves developing a custom-trained deep learning model for full- fledged sign language recognition, capable of understanding complex gestures, facial expressions, and two-hand movements across different regional sign languages such as ASL or ISL. The system can also be expanded to support multi-user group meetings with efficient bandwidth management and participant control features.

VIII REFERENCES

[1] Abid, Muhammad R.; Santiago Melo, Lidia B.; Petriu, Emil M., “Dynamic Sign Language and Voice Recognition for Smart Home Interactive Application”, 2013 IEEE International Symposium on Medical Measurements, 2013.

[2] Ananthanarayanan, Ganesh; Bahl, Paramvir; Bodík, Peter; Chintalapudi, Krishna; Philipose, Matthai; Ravindranath, Lenin; Sinha, Sudipta, “Real- Time Video Analytics: The Killer App for Edge Computing”, Computer, 2017.

[3] Banitsas, Konstantinos A.; Perakis, Konstantinos; Tachakra, Sapal; Koutsouris, Dimitrios, “Use of 3G Mobile Phone Links for Teleconsultation Between a Moving Ambulance and a Hospital Base Station”, Journal of Telemedicine and Telecare, 2006.

[4] Berke, Larwan; Albusays, Khaled; Seita, Matthew; Huenerfauth, Matt, “Preferred Appearance of Captions Generated by Automatic Speech Recognition for Deaf and Hard-of-Hearing Viewers”, Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems, 2019.

[5] Boyes-Braem, Penny, “A Multimedia Bilingual Database for the Lexicon of Swiss German Sign Language”, Sign Language & Linguistics, 2002.

[6] Debevc, Matjaz; Kosec, Primoz; Holzinger, Andreas, “Improving Multimodal Web Accessibility for Deaf People: Sign Language Interpreter Module”, Multimedia Tools and Applications, 2010.

[7] Deshmukh, Anushri; Machindar, Aryan; Lale, Sahil; Kasambe, Prahshant, “Enhancing Communication for the Hearing Impaired: A Real- Time Speech to Sign Language Converter”, 27th International Symposium on Wireless Personal Multimedia Communications, 2024.