



# NEURAL 3D AVATAR–ASSISTED BIDIRECTIONAL COMMUNICATION FRAMEWORK FOR DEAF AND HEARING USERS

Mrs.N.KALAISELVI, M.E<sup>1</sup>, Mavuliya A<sup>2</sup>, Madhavi M<sup>3</sup>, Priyanga R<sup>4</sup>, Sneha S<sup>5</sup>

<sup>1</sup>Assistant Professor, <sup>2</sup>Student,

Department of CSE, Nelliandavar Institute of Technology, Pudhuppalayam, Tamil Nadu, India

**ABSTRACT** - Deaf and mute communities include individuals with partial or complete hearing and speech impairments who primarily depend on sign language for communication. However, interaction with non-signers remains difficult due to limited sign language awareness, creating barriers in education, healthcare, workplaces, and public services. Existing assistive communication systems face challenges such as restricted vocabulary coverage, reliance on predefined gestures, poor real-time performance, limited support for regional sign languages like Indian Sign Language, lack of two-way communication, and inadequate handling of continuous gestures, facial expressions, and environmental noise, which significantly reduces their real-world effectiveness. To overcome these challenges, this project introduces an intelligent, AI-driven communication framework designed to provide seamless two-way interaction between sign-language users and non-signers. The proposed system incorporates a Transformer-Based Gesture Encoder that analyzes continuous hand movements, facial expressions, and body dynamics from live video streams to accurately recognize Indian Sign Language. This enables natural and real-time interpretation of signs without relying on predefined gesture templates, thereby improving recognition accuracy and adaptability in dynamic environments. In addition, the system integrates a Speech Recognition and Synthesis Module utilizing advanced techniques such as RNN-Transducer, Connectionist Temporal Classification, and Deep Neural Networks to convert spoken language into text with high reliability, even under noisy conditions. The interpreted text is further processed by an Avatar Module based on Neural Avatar Synthesis, which generates realistic, human-like sign language animations to visually convey spoken or typed content. Implemented through a web-based interface, the system ensures accessibility, scalability, and ease of use, ultimately promoting inclusive and effective communication across diverse user groups.

**Keywords** : Neural 3D Avatar, Bidirectional Communication, Deaf and Hearing Users, Sign Language Recognition, Sign Language Translation, Gesture Recognition, Computer Vision, Deep Learning, Human-Computer Interaction, Speech-to-Sign Conversion, Sign-to-Speech Conversion, Real-Time Communication, AI-based Avatar, Natural Language Processing, Assistive Technology

## 1. INTRODUCTION

Sign language is manual communication commonly used by people who are deaf. Sign language is not universal; people who are deaf from different countries speak different sign languages. The gestures or symbols in sign language are organized in a linguistic way. Each individual gesture is called a sign. Each sign has three distinct parts: the handshape, the position of the hands, and the movement of the hands. American Sign Language (ASL) is the most commonly used sign language in the India. sign language (also signed language) is a language which uses manual communication, body language, and lip patterns instead of sound to convey meaning simultaneously combining hand shapes, orientation and movement of the hands, arms or body, and facial expressions to fluidly express a speaker's thoughts. Signs often represent complete ideas, not only words. However, in addition to accepted gestures, mime, and hand signs, sign language often includes finger spelling, which involves the use of hand positions to represent the letters of the alphabet. Sign languages have developed in circumstances where groups of people with mutually unintelligible spoken languages found a common base and were able to develop signed forms of communication. Sign languages commonly develop in deaf communities, which include people who are deaf or hard of hearing, friends and families of deaf people, as well as interpreters.

## 2. LITERATURE REVIEW

In this paper [1] The proposed framework integrates a hybrid deep learning model for ASL gesture recognition, followed by text refinement using a prompt-configured Large Language Model (LLM), and finally gesture synthesis into ISL using a real-time video generation network. Initially, ASL alphabet gestures are recognized using a combination of classifiers. The recognized text is refined for grammatical and contextual correctness using an LLM. The corrected text is then converted into smooth ISL gesture videos using RIFE-Net (Real-Time Intermediate Flow Estimation Network). The system addresses gesture variability and linguistic differences between ASL and ISL while maintaining real-time processing.

In this paper [2] The proposed approach extracts hand and body landmark features from signers and models them as visual data streams. A data grading mechanism is introduced to enhance visual embedding through a self-attention mechanism. Textual sign descriptions are incorporated alongside visual embeddings within a Zero-Shot Learning (ZSL) framework. The model is evaluated on two established ZSL benchmarks to validate its effectiveness.

In this paper [3] The proposed framework includes a hand detection module and a correctness discrimination model. Flow-guided feature extraction and key frame detection are used for accurate hand localization. An encoder-decoder architecture combining 3D convolution, 2D deformable convolution, residual connections, and sequence attention mechanisms is employed for correctness discrimination. A new Sign Language Correctness Discrimination (SLCD) dataset is constructed with dual labels for category and standardization. Semi-supervised learning generates pseudo hand position labels.

In this paper [4] The proposed framework includes three components: Dynamic Motion Network (DMN), Accumulative Motion Network (AMN), and Sign Recognition Network (SRN). Key postures are extracted to handle signer variability. A novel accumulative video motion frame is generated by fusing forward and backward key postures to preserve spatial and temporal information. Features from DMN and AMN are fused and fed into SRN for classification.

In this paper [5] Normal people cannot understand the signs used by deaf, as they do not know the meaning of a particular sign. This system uses a camera, which captures various gestures of the hand. Then, processing of the image takes place by using various algorithms. . The implementation of the system is by using OpenCV-Python. The system uses various libraries. RGB is a device dependent model. Different systems produce different values. This method takes 33% of RED, 33% of GREEN and 33% of BLUE i.e. the contribution is equal from all the three colors. Gaussian filters are smoothing filters that reduce this noise by using a Gaussian kernel. As an input, provide the values for height, width and standard deviation in both the directions. If its value is larger than high threshold value, the pixel value is made 255. o detect the changes in intensities and gradient values in horizontal and vertical directions, i.e. (x, y). It returns the first order derivatives for both the directions. This comparison of template image with the images present in the data set uses Sum of Absolute Difference (SAD) method. SAD method computes intensity values. This method computes the values by subtracting the pixel values of both the images and

then adds the result, until no pixel is left. the results obtained after performing the experiment. a sample input image of hand gesture. This image is a three-dimensional image. after template matching using SAD and the matched image from the data set

In this paper [6] Sign language, the non-verbal language used by the people with hearing and speaking disability, used the YOLOv3 algorithm for real-time sign language conversion. the finger-spelling of words is the leading alternative for communication for deaf people. Using our proposed technique, finger-spelling can also be used to write documents. It consists of 10 digits and 36 characters used in standard BdSL along with three proposed signs for generating sentences, which are compound characters, space, and end of sentence. Bangla Sign Language detection system where they converted training samples from RGB color space to HSV color space and then extracted feature using Scale Invariant Feature Transform (SIFT) Algorithm before feeding to the model. extracted features using the Principal Component Analysis (PCA) algorithm to reduce dimensionality and then data fed to the model. proposed a Bangla Sign language to speech generation system using a pair of smart gloves, sensors, and microcontroller. There are multiple models like VGG16 [15], SpineNet [19], CSPResNeXt50 [14], CSPDarknet53 [14] that can be used as backbone model for the network. For our work as the model backbone, CSPDarknet53 is used which significantly reduces computational cost [8] to extract features from the input image. CNN's ability to learn even after reducing the number of layers. The computation distribution property of CSPNet reduces the computational bottleneck and increases the utilization rate of each computational unit.

In this paper [7] The proposed system for ISL interpretation performs two major tasks: (i) Gesture to Text conversion and (ii) Speech to Gesture conversion. Feature extraction is done using the Speeded-Up Robust Feature (SURF) method. Machine Learning algorithms like Convolutional Neural Network (CNN), Recurrent Neural Network (RNN) and Support Vector machine (SVM). The SVM classifier is used along with K-means clustering classifier and Bag of Visual words (BoV) model to achieve better accuracy. For both CNN and RNN, the data in the dataset is divided into three parts. K-Means clustering, BoV, and SVM classifiers has the highest accuracy in recognizing the hand gestures. Hence, it is more reliable for gesture recognition.

In this paper [8] The algorithm used for the implementation automated sign language interpreter is k-nearest neighbour (KNN) algorithm. KNN is simple ML algorithm based on supervised learning technique. It stores all the available data and classifies a new data point based on the similarly. When new data appears then it is easily classified into well suite category by using KNN algorithm.

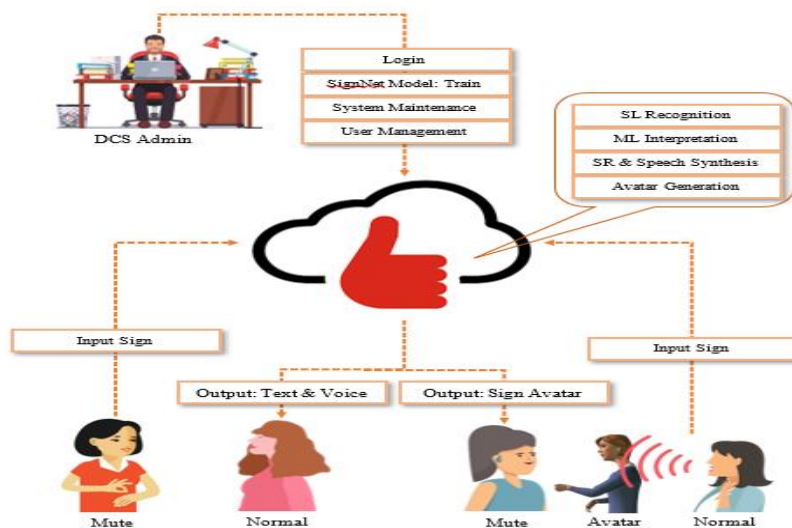
In this paper [9] MPU6050 sensor module is complete 6-axis Motion Tracking Device. It combines 3-axis Gyroscope, 3-axis Accelerometer, on-chip Temperature sensor and Digital Motion Processor all in small package. the trained dataset is fit into a Support Vector Machine (SVM) classifier. Different Languages have different trained datasets. The model is stored in the pkl Format the data received at the computer has to be processed before feeding it into the SVM classifier. to predict both ASL and ISL gestures allowing the user to speak English and some Indian languages by making gestures.

In this paper [10] Sign language recognition (SLR) techniques can be categorized into three groups: computer vision-based models, wearable sensor-based systems, and hybrid systems Exclusion terms such as image, vision, and camera were used to narrow results. Non-English articles and those not published in peer reviewed sources were excluded from the study. Since the primary focus of this review was to examine methods for SLR used in mobile applications, only studies involving wearable devices were included. The VPL Data Glove has two fibre optic sensors on the back of each finger to detect joint flexion, and a tracker on the back of the palm to detect the six degrees of freedom of orientation and position. practical SLR applications require a device that maximizes ease of use and comfort while maintaining a high level of accuracy.

### III. METHODOLOGY

The proposed Neural 3D Avatar–Assisted Bidirectional Communication Framework is designed to enable seamless interaction between deaf and hearing users through advanced artificial intelligence techniques. Initially, input data is captured in two forms: speech signals from hearing users and sign language gestures from deaf users using cameras and microphones. The speech input undergoes preprocessing, including noise reduction and feature extraction, followed by speech-to-text conversion using deep learning-based models. The extracted text is then translated into corresponding sign language sequences using natural language processing techniques. These sequences are rendered through a realistic 3D avatar that performs accurate hand gestures, facial expressions, and body movements for effective visual communication.

#### 3.1. ARCHITECTURE DIAGRAM



FIG(3.1.1) : ARCHITECTURE DIAGRAM

### IV. EXPERIMENTAL RESULTS

The experimental results of the proposed Neural 3D Avatar–Assisted Bidirectional Communication Framework demonstrate its effectiveness in enabling accurate and real-time interaction between deaf and hearing users. The system was evaluated using benchmark sign language datasets and real-time user inputs under varying environmental conditions. The gesture recognition module achieved high accuracy due to the integration of deep learning models, effectively identifying complex hand movements and facial expressions. Similarly, the speech-to-text and text-to-speech components produced reliable outputs with minimal error rates, ensuring smooth communication flow. The 3D avatar rendering system successfully generated natural and synchronized gestures, improving user understanding and engagement. Performance metrics such as accuracy, latency, precision, and recall indicated that the framework operates efficiently with low delay, making it suitable for real-time applications. Comparative analysis with existing systems showed significant improvements in translation accuracy and response time. Overall, the results validate that the proposed framework provides a robust, scalable, and user-friendly solution for inclusive communication between deaf and hearing individuals.

#### 4.1 IMPLEMENTATION RESULT

##### Development Environment

The system is developed using Python for backend logic with Flask as the web framework. The frontend is implemented with HTML, CSS, JavaScript, and Bootstrap to provide a responsive and user-friendly interface. MySQL is used for secure storage of user data, sign-language datasets, audio features, and avatar mappings. AI and ML libraries including OpenCV, MediaPipe/OpenPose, TensorFlow/PyTorch, Librosa, and NumPy are integrated for real-time sign recognition, speech processing, and avatar generation.

##### Database Design

The database is structured to store user profiles, role mappings (Admin, Deaf, Non-Deaf), gesture datasets, audio features, session logs, recognized text, and avatar animation data. Proper relational table design ensures fast retrieval, secure access, and smooth integration across all modules.

## Dataset Preparation and Model Training

Sign gesture images and videos are preprocessed through resizing, grayscale conversion, noise filtering, binarization, and background subtraction. The SignNet model is trained to extract gesture features using convolutional layers and classify gestures accurately. For speech, audio is preprocessed into acoustic features and trained with RNN-Transducer, CTC, and DNN models for robust recognition even in noisy environments.

## Model Integration and Deployment

Trained models are integrated into the Flask application. Live video streams of gestures are captured and processed through the SignNet and Transformer-Based Gesture Encoder for real-time recognition. Audio input is processed through the speech recognition module, and recognized text is mapped to avatar gestures using the Neural Avatar Synthesis engine. The system delivers real-time two-way communication between deaf and hearing users.

## Front-End Development

A web-based interface provides distinct dashboards for Admins, Deaf users, and Non-Deaf users. Deaf users can input signs and view avatar feedback, while non-deaf users can speak or type messages that are converted into animated signs. The interface is designed for ease of use and accessibility.

## Back-End Development

The backend manages gesture recognition, speech processing, avatar animation, session management, user authentication, and secure communication between modules. APIs connect the front-end interface with AI models and the database.

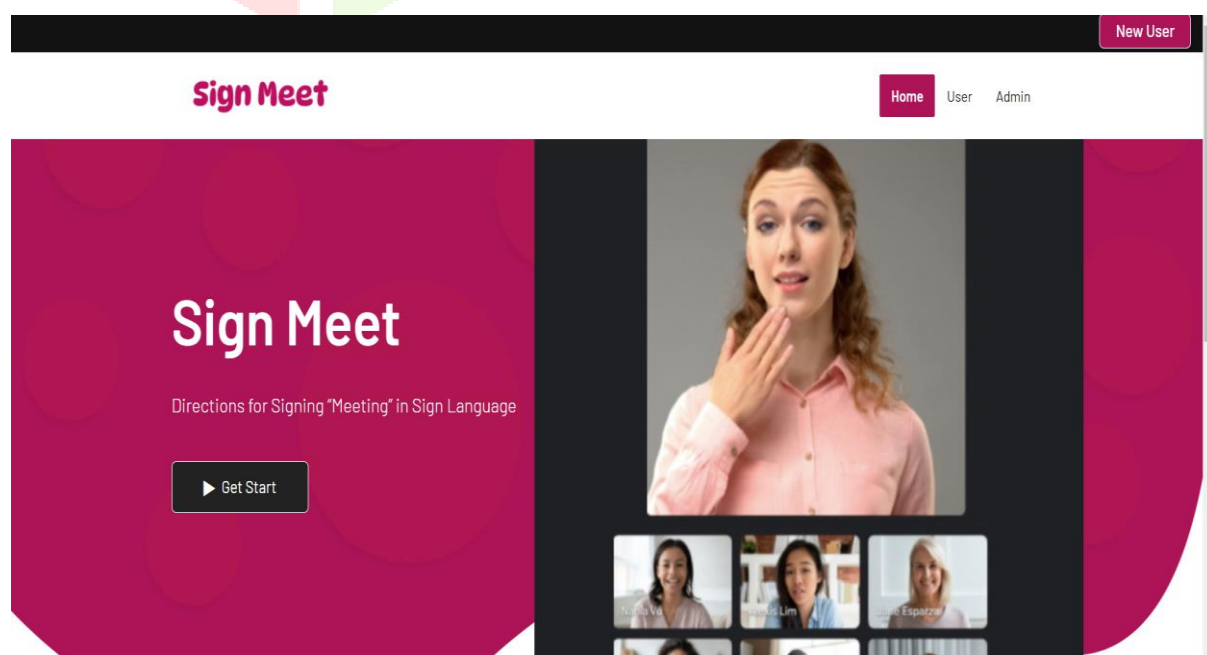
## Module Integration

All modules—SignNet Model, Sign Recognition, Speech Recognition and Synthesis, Neural Avatar Synthesis, and the User Dashboard—are integrated to provide seamless bidirectional communication. Live gesture and speech input are processed in real time, and outputs are delivered to the respective users instantly.

## Testing and Quality Assurance

The system is tested for accuracy of gesture and speech recognition, real-time response performance, avatar animation fidelity, and overall reliability. Stress testing ensures the system handles multiple simultaneous users while maintaining smooth and effective communication.

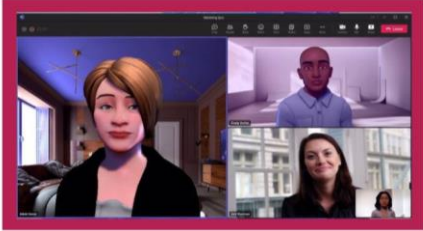
## 4.2 RESULT



**Sign Meet** Home User Admin New User

## Admin

Recognize the Hand Sign and Training



Username:

Password:

**Sign Meet** Home Upload Logout

## Sign Meet

Training Phase  
Hand Sign

Communication For Such People To Convey Their Messages.

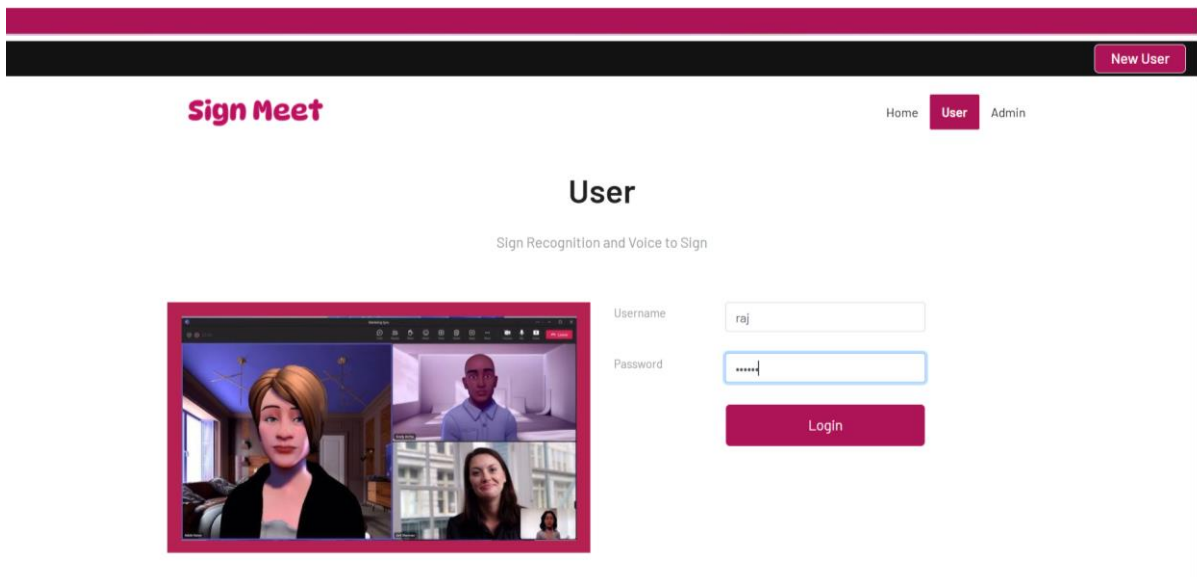
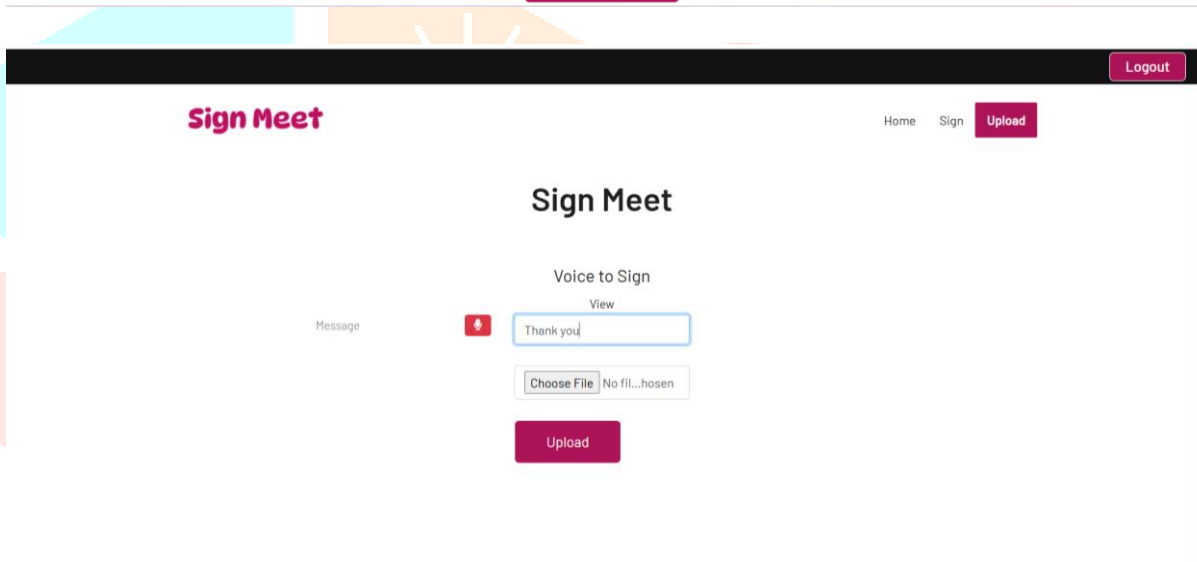
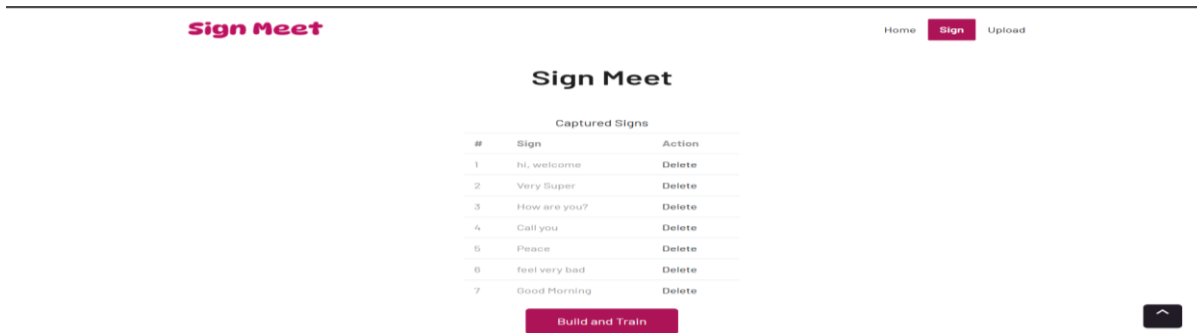
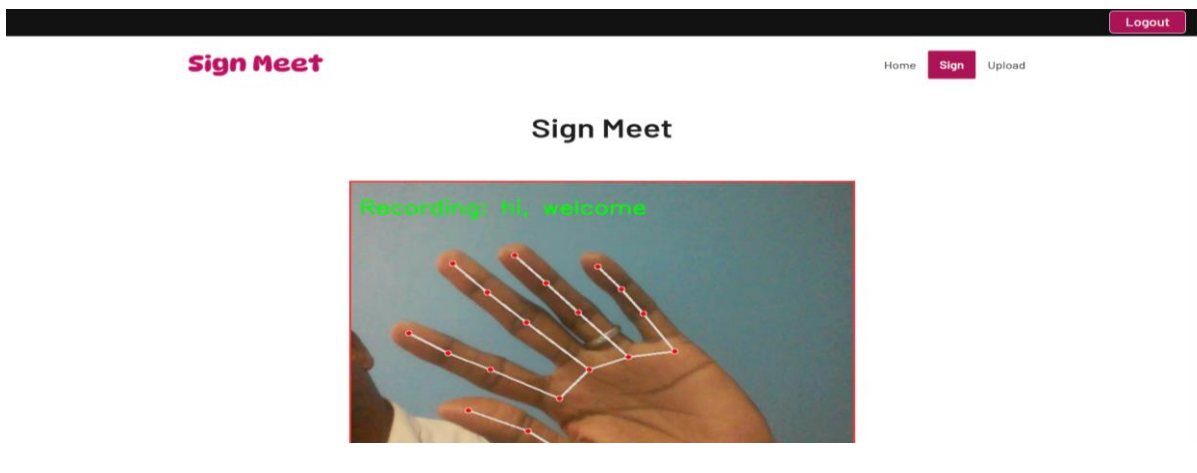
**Sign Meet** Home Upload Logout

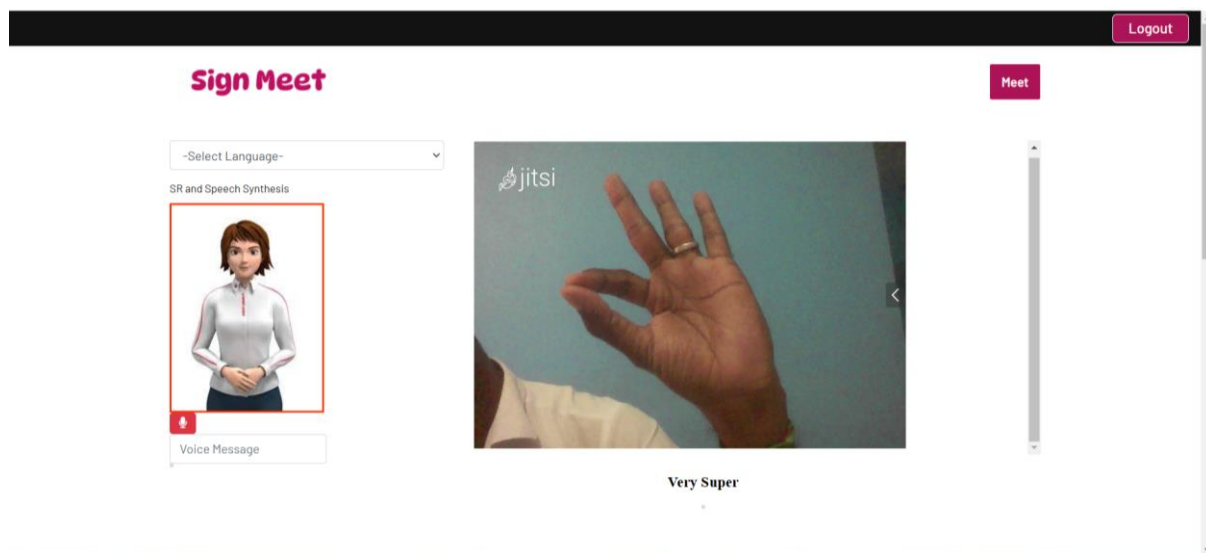
## Sign Meet

Training Phase  
Hand Sign

Enter the Gesture Label:

Communication For Such People To Convey Their Messages.





## V. CONCLUSION

In conclusion, the comprehensive design and development of this project is centered on enhancing communication between deaf and non-deaf users through advanced AI-driven technologies. The system employs Python and Flask for back-end processing, MySQL for structured and secure data management, and TensorFlow/PyTorch with OpenCV and MediaPipe/OpenPose for real-time gesture recognition and speech processing. Core functionalities, including Transformer-Based Sign Recognition, Speech Recognition and Synthesis, and Neural Avatar Sign Generation, work together to provide a seamless, bidirectional communication experience. These modules ensure accurate recognition of gestures and speech, generate expressive avatar animations, support multi-language translation, and facilitate inclusive interactions. The system's responsive web interface, robust AI models, and real-time processing collectively enhance accessibility, independence, and social participation for deaf and mute users, offering a practical and scalable solution for inclusive communication.

## FUTURE ENHANCEMENT

The project has successfully enabled real-time, two-way communication between deaf and hearing users. However, there are several areas for future enhancement to improve accessibility, scalability, and overall user experience:

- **Integration with Smart Devices and IoT** – Connect the system with smart home appliances, hospital systems, and service kiosks so that deaf or mute users can control devices, request assistance, or interact with services using gestures or voice commands.
- **Mobile Application Development** – Develop dedicated Android and iOS applications to allow users to access the system anytime and anywhere, supporting offline features, portable communication, and real-time interaction without relying on a computer.
- **Emotion and Expression Detection** – Integrate emotion-recognition models to interpret facial expressions and emotional tone, enabling the system to convey feelings such as happiness, sadness, urgency, or confusion, making communication more natural and realistic.

**REFERENCE**

1. 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.
2. G. T. R. Chiranji Lal Chowdhary and B. D. Parameshachari, *Computer Vision and Recognition Systems: Research Innovations and Trends*, CRC Press, 2022.
3. 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.
4. 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.
5. P. Likhari, 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.
6. F. Li, K. Shirahama, M. A. Nisar, X. Huang and M. Grzegorzec, "Deep Transfer Learning for Time Series Data Based on Sensor Modality Classification", *Sensors*, vol. 31, no. 20(15), pp. 4271, Jul 2020.
7. 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.
8. S. Sharma, R. Gupta and A. Kumar, "Trbaggbost: 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>.
9. 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.
10. 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.
11. 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.
12. 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.
13. 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.

14. 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.
15. D. Konstantinidis, K. Dimitropoulos and P. Daras, "Sign language recognition based on hand and body skeletal data", *3DTV-CON*, pp. 1-4, 2018.
16. C. Motoche and M. E. Benalcázar, "Real-Time Hand Gesture Recognition Based on Electromyographic Signals and Artificial Neural Networks" in *Artificial Neural Networks and Machine Learning - ICANN 2018 Lecture Notes in Computer Science*, Springer, vol. 11139, 2018.
17. J. Pu, W. Zhou and H. Li, "Dilated convolutional network with iterative optimization for continuous sign language recognition", *IJCAI*, vol. 3, pp. 885-91, 2018.
18. R. Cui, H. Liu and C. Zhang, "Recurrent convolutional neural networks for continuous sign language recognition by staged optimization", *CVPR*, pp. 7361-9, 2017.
19. S. Y. Kim, H. G. Han, J. W. Kim, S. Lee, and T. W. Kim, "A hand gesture recognition sensor using reflected impulses," *IEEE Sensors Journal*, vol. 17, no. 10, pp. 2975-2976, 2017.

