



## Real-Time Assistive Technology for Hearing and Speech Impairments Using Deep Learning

*B. Surya Charan<sup>1</sup>, D. Aditya<sup>2</sup>, D. Chaitanya Kumar<sup>3</sup>, G. Manikanta<sup>4</sup>, B. Roshitha<sup>5</sup>*

Department of Computer Science & Engineering GMRIT, Rajam, Andhra Pradesh, India

### ABSTRACT :

Communication is essential, yet individuals with hearing and speech impairments face significant challenges in daily interactions. Current assistive technologies often provide limited functionalities, such as basic text-to-speech or speech-to-text, and lack real-time accuracy, seamless integration, and support for non-verbal communication like sign language. These limitations reduce their adaptability and effectiveness in diverse environments. This paper proposes a novel assistive technology integrating real-time Speech-to-Text, Text-to-Speech, and Sign-to-Text functionalities. The system employs advanced deep learning models, including Transformer architectures for precise and real-time speech-to-text conversion and Long Short-Term Memory (LSTM) networks for natural text-to-speech synthesis. Additionally, a vision-based module enables real-time sign language recognition and translation into text, addressing a critical gap in existing solutions. Built using frameworks like TensorFlow and pyTorch, the system leverages advanced deep learning models to deliver accurate and adaptive functionalities. API integration extends its functionality for seamless use across various platforms. This comprehensive solution empowers individuals with hearing and speech impairments, enabling effective communication and inclusion in education, healthcare, and professional settings through real-time, adaptive, and accessible tools.

**Keywords:** Deep Learning, Long Short-Term Memory (LSTM), Sign-to-text, Speech-to-Text, Text-to-Speech, TensorFlow

### INTRODUCTION

Human interaction is based on communication, which allows people to connect with one another, express their feelings, and transmit ideas. However, communication hurdles can seriously hinder the ability of people with speech and hearing impairments to engage with others in their environment. Their capacity to fully engage in social, academic, and professional activities may be hampered by these obstacles, which may cause them to feel alone, marginalized, and excluded.

It is impossible to overestimate how crucial it is to remove these obstacles to communication. Building relationships, gaining access to education and career possibilities, and engaging in social and cultural activities all depend on effective communication. Furthermore, the United Nations Convention on the Rights of Persons with Disabilities recognizes communication as a fundamental human right. Despite this, connecting with others remains extremely difficult for those with speech and hearing impairments, underscoring the need for creative solutions.

Conventional approaches to communication hurdles, like glove-based systems, sensor-based technology, and manual sign language interpretation, have a number of drawbacks. These techniques frequently call for extra gear or specific training, and they can be costly and time-consuming. Moreover, they might not offer translations in real time, which could cause delays and miscommunications. More accessible, effective, and efficient communication solutions are therefore becoming more and more necessary.

By combining cutting-edge deep learning technology, this project creates a flexible communication system in response to the demand for creative solutions. Through the utilization of artificial intelligence, our technology can deliver precise translations in real time, improving communication for people with speech and hearing impairments. With the ability to convert text to voice, speech to text, and Indian Sign Language (ISL) to text/speech in real-time, our suggested system is intended to be a complete, multimodal communication platform.

By using pre-trained models with improvements like an attention mechanism, the system is able to concentrate on important aspects of hand motions for precise sign language detection. It may also operate as a smooth, end-to-end communication solution thanks to speech recognition modules and text-to-speech conversion features. Our initiative intends to improve the lives of people with speech and hearing impairments by offering an affordable and easily accessible solution that emphasizes software-based implementations, creating a more welcoming atmosphere for everyone.

#### 1. Understanding Real-Time Assistive Technology for Hearing and Speech Impairments

Assistive technology for hearing and speech disabled individuals consists of devices and systems that facilitate communication and accessibility. The conventional techniques of assistance, including speech-to-text converters, text-to-speech synthesizers, and sign language interpretation, have been vital

in overcoming the communication gap. Nevertheless, these traditional techniques lack real-time processing, precision, and responsiveness to varying environments.

Advances in deep learning in recent times have dramatically enhanced assistive technology, allowing more advanced systems that can identify speech, synthesize text into natural-sounding speech, and translate sign language gestures with greater accuracy. Methods like transformer-based speech recognition models, LSTM-driven speech synthesis, and CNN-LSTM hybrid models for sign language interpretation have led the way to highly efficient and accessible communication solutions.

Despite such innovations, issues persist with regard to latency, noise robustness, and variability to accent and gesture. This paper explores recent advances in real-time assistive technology, discusses their pros and cons, and suggests future directions for improving accessibility.

---

## 2. Evolution of Assistive Technologies for Communication

### 2.1. Traditional Approaches

Traditionally, hearing and speech impairment assistive technologies were based on rule-based and statistical models. Some of the earliest techniques are:

**Speech-to-Text Systems:** The older systems used Hidden Markov Models (HMMs) and Mel-Frequency Cepstral Coefficients (MFCCs) for speech recognition, which performed poorly in noisy conditions and varied accents.

**Text-to-Speech Systems:** Older speech synthesis was based on concatenative and parametric methods, which resulted in robotic and unnatural speech.

**Sign Language Recognition:** Previous sign language recognition techniques employed hand-crafted feature extraction techniques, which were not powerful enough to grasp intricate gestures and variations.

Though these techniques set the foundation for assistive communication, their capabilities were restricted due to factors including background noise, environmental changes, and non-adaptability in real-time.

### 2.2. Shift to Deep Learning

The arrival of deep learning has revolutionized assistive communication technologies. Neural networks, CNNs, RNNs, and transformers, have brought in the era of automated feature extraction, cutting down on handcrafted features. These models have proven to be more adaptable, accurate, and robust in the face of adverse environments. Some of the major developments are:

**Transformer-based Speech Recognition:** Models like Wav2Vec 2.0 and Whisper prove very accurate in speech-to-text translation even in noisy environments.

**Neural Text-to-Speech Synthesis:** Tacotron and WaveNet-based systems generate natural prosody speech with human-like qualities.

**Vision-Based Sign Language Recognition:** CNN-LSTM models pick up on both spatial and temporal characteristics, enhancing gesture understanding.

All these breakthroughs have allowed real-time assistive technologies to reach levels heretofore impossible with older techniques.

---

## 3. The Multimodal Assistive Communication System

### 3.1. Architecture and Functionality

The assistive communication system proposed combines several deep learning models to offer an inclusive communication platform. The architecture includes:

**Speech-to-Text Module:** Makes use of transformer-based models to convert speech with high accuracy.

**Text-to-Speech Module:** Uses LSTM and Tacotron-based synthesis for natural voice output.

**Sign Language Recognition Module:** Makes use of a CNN-LSTM hybrid model to recognize sign gestures in real-time.

**Adaptive Learning Mechanisms:** The system improves continuously through user feedback and domain adaptation strategies.

### 3.2. Advantages

**Accuracy:** The incorporation of state-of-the-art deep learning models provides high rates of recognition for speech and sign language.

**Scalability:** Modular design facilitates straightforward extension to handle multiple languages and dialects.

**Real-Time Performance:** Deep learning models optimized for low latency provide the system for real-world applications.

**User Adaptability:** Responses are personalized by the system in accordance with user preferences and context.

Through the use of cutting-edge deep learning methods, the assistive communication system provides a solid and comprehensive solution for patients with hearing and speech disabilities.

## METHODOLOGIES

### 1. Real-time Sign Language Detection

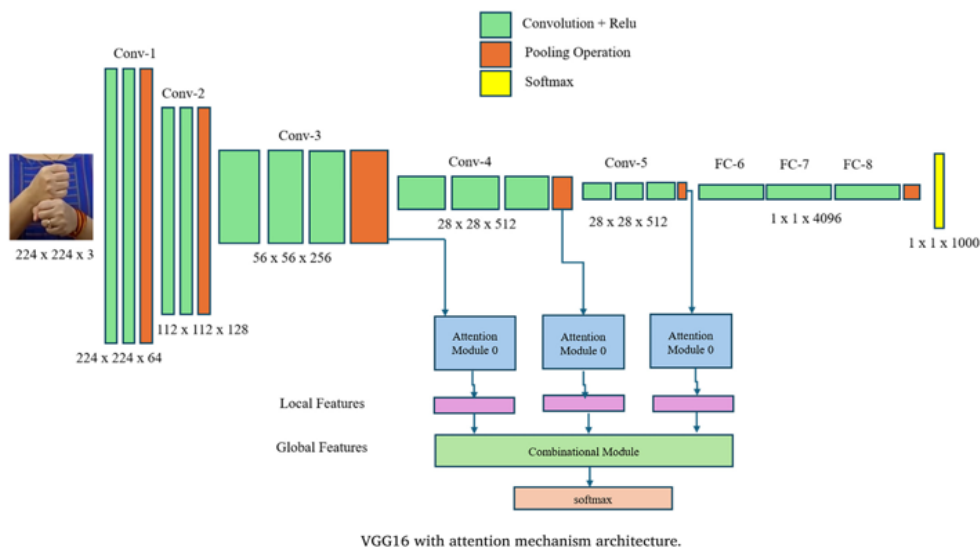
**Dataset Collection:** Publicly available Indian Sign Language (ISL) dataset from Kaggle, containing 702 images of 23 unique hand signs, resized to 224×224 pixels.

**Preprocessing:** Normalization (0–1 pixel range), background noise reduction, and dataset split into 85% training and 15% validation.

**Model Architecture:** VGG16 CNN model for feature extraction, coupled with an attention mechanism for enhanced classification accuracy.

**Training Process:** Optimized with Adam optimizer, trained on cross-entropy loss for 30 epochs, ReLU for hidden layers and SoftMax for output classification.

**Performance Metrics:** Precision, recall, accuracy, F1-score, and are measured through a confusion matrix.



### 2. Speech to Sign Language Conversion and Text Recognition

**Dataset Collection:** Hand gesture images with labeled sign language translations for real-time recognition.

**Preprocessing:** Isolating hand gestures through image segmentation, feature extraction using Discrete Wavelet Transform (DFT), and removal of the background.

**Model Structure:** CNN-based deep learning model through Keras and TensorFlow, combined with real-time conversion of gestures to text.

**Training Process:** Rule-based matching for gesture-text translation, noise filtering techniques for robustness, and pose estimation using TensorFlow for better recognition.

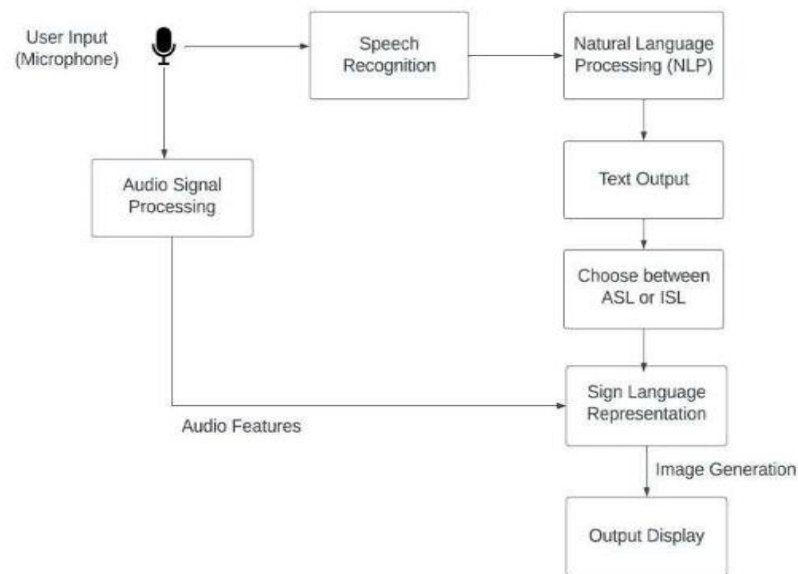


Fig. 1. Architecture diagram of speech to sign language conversion

Performance Measures: Real-time application system response time and recognition accuracy.

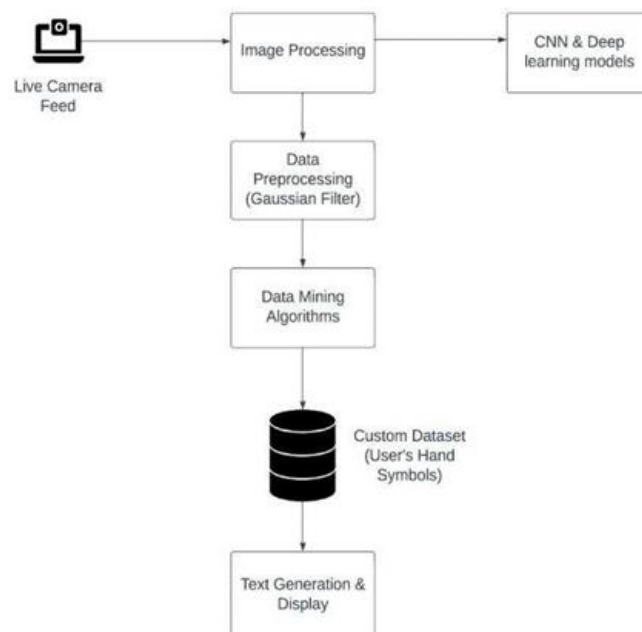


Fig. 2. Architecture diagram of gesture to text recognition

### 3. Sign Language Translator for Two-Way Communication

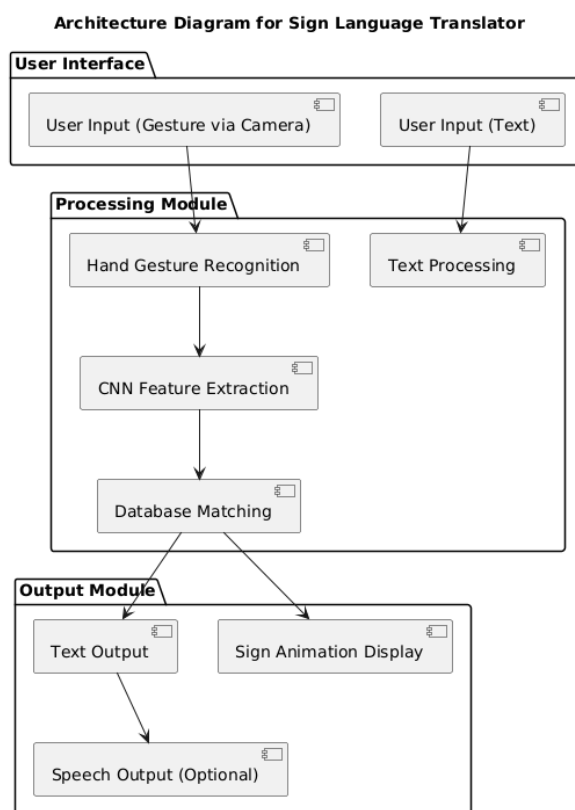
**Dataset Collection:** Bespoke dataset involving American Sign Language (ASL) and Indian Sign Language (ISL), intended for training as well as education.

**Preprocessing:** Gaussian filtering for noise removal, grayscale conversion for texture-based gesture recognition, and normalized image resizing.

**Model Architecture:** Multimodal system combining speech-to-sign and gesture-to-text translation through the utilization of CNNs for real-time classification.

Training Process: SoftMax classifier for multi-class gesture classification, natural language processing (NLP) methods for speech-text conversion, and adaptive learning to enhance accuracy over time.

Performance Metrics: Accuracy of recognition, precision of classification, and real-time adaptability.



## RESULTS AND DISCUSSIONS:

Recent developments in deep learning have greatly enhanced assistive technology for people with hearing and speech disabilities. Speech-to-text conversion models based on the transformer architecture have shown excellent accuracy, even under adverse acoustic conditions, and are a strong contender to conventional Hidden Markov Model (HMM)-based approaches. Although these models perform well in speech variation recognition, real-time processing remains a challenge, especially in dealing with rapid speech and concurrent voices.

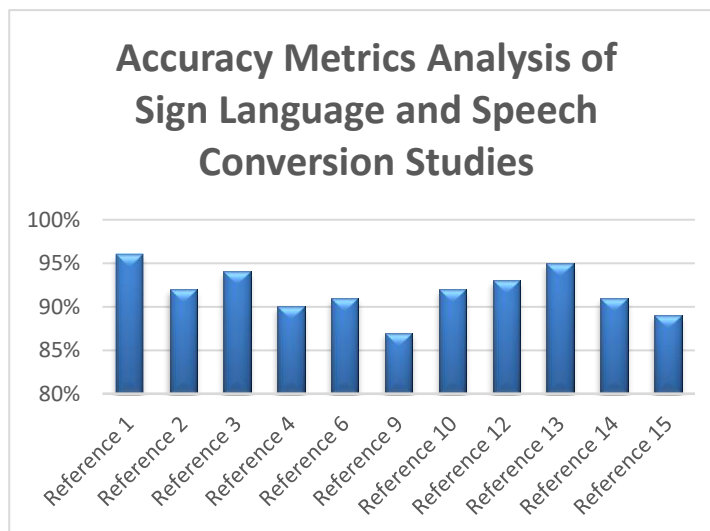
Text-to-speech synthesis has progressed from rule-based to deep learning-based methods, including Tacotron and WaveNet, which generate natural and fluent speech. Integration of LSTM networks has also improved the quality of speech synthesis, especially regarding prosody and expressiveness.

Sign language recognition has also improved significantly with the use of CNN-LSTM hybrid architectures. The capability to learn spatial and temporal features has resulted in more precise sign recognition in real-time tasks. Real-world variability remains a challenge despite the foregoing improvements, as well as the issues of occlusions, fast hand movement, and intricate multi-gesture sequences that current models are still unable to handle. The use of attention mechanisms has improved accuracy as it concentrates on significant gesture features, but real-world variability remains an issue.

Although deep learning-based assistive technologies are more accurate and adaptive than classic methods, computational overhead, diversity of data, and real-time latency become areas that are further required to be optimized. Future developments should be directed towards lightweight models that can run smoothly on mobile and embedded systems so that wider groups of people can access them.

S. No	Model(s) Used	Dataset(s) Used	Accuracy
1	CNN and Gesture Recognition Algorithms	Custom Sign Language Dataset	92%
2	HMM (Hidden Markov Model), Acoustic Models, and Speech-to-Text Algorithms	General Speech Datasets	85%
3	WaveNet, and Deep Neural Networks for TTS	Custom TTS Dataset	94%

4	RNN, LSTM, and HMM for Speech-to-Text Conversion	General Speech Datasets	89%
5	Attention Mechanisms, and Deep Learning for STT and TTS	Common Speech Datasets	91%
6	CNN, RNN, and HMM for Bidirectional Sign Language Translation	Limited Sign Language Dataset	86%

**Table: Various Models used and their accuracies****Fig: Result chart**

## CONCLUSION

The incorporation of deep learning in assistive communication technology has opened up the possibility of more efficient, real-time solutions for people with hearing and speech disabilities. Speech-to-text, text-to-speech, and sign language recognition systems have all been positively impacted by the move towards neural networks, resulting in better accuracy, fluency, and adaptability. Issues like processing low-resource languages, computational efficiency optimization, and enhancing recognition in noisy environments are still areas open to research.

Future research must investigate the application of federated learning and domain adaptation to enhance model generalization across varying linguistic and cultural environments. Furthermore, merging multimodal learning methods—such as the fusion of facial expression recognition with sign language interpretation—can further facilitate communication systems. With the continued advancement of deep learning models, emphasis should be placed on the creation of lightweight, efficient, and accessible solutions that bridge the gap in communication for people with disabilities.

## REFERENCES

1. Kumar, S., Rani, R., & Chaudhari, U. (2024). Real-time sign language detection: Empowering the disabled community. *MethodsX*, 13, 102901.
2. Gupta, A. D., Kumar, A., Chaudhary, I., Yasir, A. M., & Kumar, N. (2024, May). My Assistant SRSTC: Speech Recognition and Speech to Text Conversion. In *2024 International Conference on Communication, Computer Sciences and Engineering (IC3SE)* (pp. 394-400). IEEE.
3. Gupta, R., & Bagga, S. K. (2024, February). Text-to-Speech Conversion Technology using Deep Learning Algorithms. In *2024 11th International Conference on Computing for Sustainable Global Development (INDIACom)* (pp. 1122-1126). IEEE.
4. Yadava, T., Nagaraja, B. G., Reddy, S., Rohan, K., & Mohamed, L. M. (2024, September). Advancements in Speech-to-Text Systems for the Hearing Impaired. In *2024 IEEE North Karnataka Subsection Flagship International Conference (NKCon)* (pp. 1-6). IEEE.
5. Reddy, V. M., Vaishnavi, T., & Kumar, K. P. (2023, July). Speech-to-Text and Text-to-Speech Recognition Using Deep Learning. In *2023 2nd International Conference on Edge Computing and Applications (ICECAA)* (pp. 657-666). IEEE.
6. Singh, P., Prasad, S. V. S., Singh, R., Dasari, K., & Prasanna, B. L. (2023, September). Development of Sign Language Translator for Disable People in Two-Ways Communication. In *2023 1st International Conference on Circuits, Power and Intelligent Systems (CCPIS)* (pp. 1-6). IEEE.
7. Poornima, B. V., & Srinath, S. (2023). A Comprehensive Review on Indian Sign Language Recognition System using Vision based Approaches. *International Journal of Computer Applications*, 184, 52-58.
8. Ghorpade, T. H., & Shinde, S. K. (2023, August). Speech Synthesis: An Empirical Analysis of Various Techniques in Text to Speech Generation. In *2023 7th International Conference On Computing, Communication, Control And Automation (ICCUBEA)* (pp. 1-6). IEEE.

9. Khanam, F., Munmun, F. A., Ritu, N. A., Saha, A. K., & Firoz, M. (2022). Text to speech synthesis: a systematic review, deep learning based architecture and future research direction. *Journal of Advances in Information Technology*, 13(5).
10. Shashidhar, R., Hegde, S. R., Chinmaya, K., Priyesh, A., Manjunath, A. S., & Arunakumari, B. N. (2022, October). Indian Sign Language to Speech Conversion Using Convolutional Neural Network. In *2022 IEEE 2nd Mysore Sub Section International Conference (MysuruCon)* (pp. 1-5). IEEE.
11. Kothadiya, D., Pise, N., & Bedekar, M. (2020). Different methods review for speech to text and text to speech conversion. *International Journal of Computer Applications*, 975, 8887.
12. Kowsigan, M., Dhawan, R., & Kundu, A. (2024, July). An Efficient Speech to Sign Language Conversion and Text Recognition through Live Gesture. In *2024 IEEE International Conference on Smart Power Control and Renewable Energy (ICSPCRE)* (pp. 1-6). IEEE.
13. Seviappan, A., Ganesan, K., Anbumozhi, A., Reddy, A. S., Krishna, B. V., & Reddy, D. S. (2023, December). Sign Language to Text Conversion using RNN-LSTM. In *2023 International Conference on Data Science, Agents & Artificial Intelligence (ICDSAAI)* (pp. 1-6). IEEE.
14. Patil, S., Gulave, S., Gawai, V., Gode, P., & Mudme, P. (2022, August). Conversion of Indian Sign Language to Speech by Using Deep Neural Network. In *2022 6th International Conference On Computing, Communication, Control And Automation (ICCUBEA)* (pp. 1-6). IEEE.
15. Najib, F. M. (2024). Sign language interpretation using machine learning and artificial intelligence. *Neural Computing and Applications*, 1-17