As per UGC guidelines an electronic bar code is provided to secure your paper

International Journal for Modern Trends in Science and Technology
Volume 11, Issue 07, pages 196-203.
ISSN: 2455-3778 online
Available online at: http://www.ijmtst.com/vol11issue07.html
DOI: https://doi.org/10.5281/zenodo.16131306

Bridging the Gap: Speech to Sign Language Translation for Indian Languages Using Machine Learning and Deep Learning

M. Rama Krishna, Sami Ul Rehaman Khan Pathan, Sai Kiran Kopuri, Hrushikesh Dalli

Department of Electronics and Communication Engineering, Andhra Loyola Institute of Engineering and Technology, Vijayawada, India.

To Cite this Article
M. Rama Krishna, Sami Ul Rehaman Khan Pathan, Sai Kiran Kopuri & Hrushikesh Dalli (2025). Bridging the Gap: Speech to Sign Language Translation for Indian Languages Using Machine Learning and Deep Learning. International Journal for Modern Trends in Science and Technology, 11(07), 196-203. https://doi.org/10.5281/zenodo.16131306

Article Info
Received: 14 May 2025; Accepted: 06 July 2025.; Published: 17 July 2025.

KEYWORDS

Speech-to-Sign Translation, Indian Sign Language (ISL), Machine Learning (ML), Deep Learning (DL), Natural Language Processing (NLP), Automatic Speech Recognition (ASR), Real-time Communication, Inclusive Technology.

ABSTRACT

In India, millions of individuals face challenges in communication due to hearing and speech impairments, relying on Indian Sign Language (ISL) to convey their thoughts and ideas. However, the gap between spoken language and sign language creates communication barriers between hearing and non-hearing communities. This project aims to bridge this gap by developing a Speech-to-Sign Language Translation System that converts spoken words into corresponding ISL signs using a combination of Machine Learning (ML) and Deep Learning (DL) techniques. The proposed system processes spoken language, corrects grammar errors using Natural Language Processing (NLP), translates the text into appropriate sign representations, and displays the corresponding ISL images. The architecture includes Automatic Speech Recognition (ASR) for speech-to-text conversion, grammar correction with NLP, and image classification models that predict and display relevant ISL signs dynamically. The system is designed to work efficiently in real-time, ensuring smooth and intuitive communication between the user and the system. By facilitating seamless communication for the deaf and hard-of-hearing community, this project contributes to enhancing inclusivity and reducing communication barriers.

1. INTRODUCTION

Communication plays a vital role in human interaction, allowing people to express thoughts, ideas, and emotions. However, for the hearing and speech-impaired community, traditional modes of communication often act as barriers, limiting their ability to engage effectively with others. Sign language bridges this gap, enabling non-verbal communication through visual gestures and hand movements.

In India, Indian Sign Language (ISL) serves as a prominent means of communication for the deaf and hard-of-hearing population. Despite its significance, the availability of technology to seamlessly translate spoken language into ISL is limited, leading to a communication gap between the hearing and non-hearing communities. Bridging this gap through advanced technology can foster greater inclusivity, ensuring that individuals with hearing and speech impairments have equal opportunities to communicate, learn, and participate in society.

1.1 Background and Motivation:

Bridging the communication gap between spoken language and sign language is crucial for inclusivity. While various technologies have successfully enabled speech-to-text conversion, speech-to-sign translation remains a significant challenge. Developing a system that translates speech into Indian Sign Language (ISL), especially for regional languages, can greatly empower the deaf and mute community. A reliable speech-to- sign translation system can enhance social interactions, educational access, and professional opportunities for the hearing-impaired, fostering a more inclusive society.

1.2 Importance of Sign Language Translation:

Sign language serves as a vital communication bridge for the deaf and mute community. However, the absence of an efficient system to convert spoken words into Indian Sign Language (ISL) gestures hinders effective communication. Developing an automated Speech-to-Sign system can help bridge this gap by facilitating seamless interaction between hearing and non-hearing individuals. Such a system would promote inclusivity in education, workplaces, and daily life, ensuring better accessibility and opportunities for the hearing-impaired.

1.3 Advancement in AI and Machine Learning:

The advent of Artificial Intelligence (AI), Machine Learning (ML), and Deep Learning (DL) has brought about transformative changes in natural language processing and speech recognition. These technologies have made it possible to develop systems capable of accurately translating spoken language into sign language.

Leveraging Automatic Speech Recognition (ASR), Natural Language Processing (NLP), and Deep Learning Models, this project aims to automatically convert speech into grammatically refined text and subsequently map it to Computer Vision (CV) models further enhance the system by efficiently mapping sign images to corresponding gestures, improving real-time processing and accuracy.

1.4 Objective of the Project:

The primary objective of this project is to create a real-time Speech-to-Sign Language Translation System for Indian Languages, ensuring smooth communication between hearing and non-hearing individuals. The system is designed to,

- Convert speech into text using ASR models.
- Apply NLP techniques to refine grammatical accuracy.
- Translate the corrected text into Indian Sign Language (ISL).
- Display the corresponding sign images dynamically to the user.

1.5 Overview of the Paper:

This paper introduces a Speech-to-Sign Language Translation system for Indian languages, aiming to bridge the communication gap between hearing and non-hearing communities. The system utilizes Automatic Speech Recognition (ASR) to convert speech into text, applies Natural Language Processing (NLP) for grammatical correction, and employs a Machine Learning (ML) model to map the corrected text to Indian Sign Language (ISL) sign images for dynamic display.

The paper is structured as follows:

- Section 2: Related Work discusses previous research efforts in the domain of speech-to- sign language translation and highlights existing gaps.
- Section 3: Proposed Approach outlines the architecture and workflow of the developed system, explaining each module in detail.
- Section 4: Experiments and Analysis presents the experimental setup, datasets used, and performance evaluation metrics to assess the effectiveness of the proposed system.

• Section 5: Conclusion and Future Scope summarizes the findings and discusses potential improvements and directions for future research.

2. RELATED WORK

Several studies have been conducted in the domain of sign language translation to address communication gaps between the hearing and non- hearing communities. Early systems focused on converting text to sign language using predefined gesture libraries. However, these systems lacked the capability to dynamically translate real-time speech into sign language.

Recent studies have explored the use of Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) to recognize sign gestures and convert text into sign language. Moreover, attention-based Transformer models have shown promising results in language translation tasks by preserving contextual information. Several research efforts have focused on translating American Sign Language (ASL) and British Sign Language (BSL) using pre-trained models and gesture recognition systems. However, limited progress has been made in building real-time Speech-to-Sign Language Translation systems specifically tailored for Indian Sign Language (ISL), which poses unique challenges due to the diverse linguistic structure of Indian languages.

Additionally, models trained using deep learning techniques have improved the accuracy of gesture prediction and sign language recognition. Unnathi et al. (2024) demonstrated the integration of mobile and glove-based interfaces for enhancing real-time sign language communication. These works have significantly contributed to the progress of sign language translation systems, but they primarily focus on Western sign languages, leaving a gap in the implementation of solutions tailored for Indian Sign Language (ISL).

Despite the advancements in sign language translation technologies, the majority of existing systems depend on pre-trained models or datasets that lack diversity in linguistic structures. Indian languages, being morphologically rich and diverse, require a specialized approach to accurately translate speech into Indian Sign Language (ISL). A generic NLP model may not effectively capture the nuances and grammar of regional languages, leading to inaccurate translations. Therefore, developing a custom NLP model trained on domain-specific data gathered through web scraping will not only enhance translation accuracy but also ensure that the system remains adaptable to diverse linguistic contexts. This approach aims to build a more inclusive and robust communication system tailored specifically to the needs of the Indian hearing-impaired community. To address these challenges, our proposed system aims to develop a custom NLP model by collecting and processing large-scale data through web scraping. This approach will improve translation accuracy and ensure that the system captures the intricacies of Indian languages effectively. The integration of a dynamic sign image display will further enhance real-time communication between the hearing and non-hearing communities.

3. PROPOSED APPROACH

The proposed system aims to facilitate seamless communication between the hearing and non- hearing communities by translating spoken language into Indian Sign Language (ISL) through a multi- stage pipeline. The system begins with a Speech Recognition Module, which converts spoken words into text using Automatic Speech Recognition (ASR) technology. This raw text is often prone to errors and inconsistencies, necessitating the use of a Natural Language Processing (NLP) Module that refines the text by applying grammatical corrections and ensuring coherence.

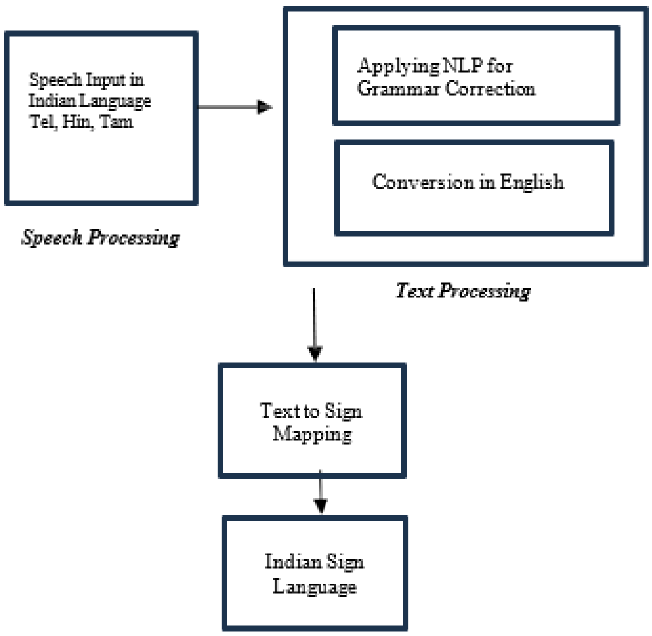

Fig 1, Model Architecture.

Once the text is corrected, the Translation Module maps the refined text to corresponding ISL sign images using a Machine Learning (ML) model trained on a comprehensive dataset of sign gestures. Finally, the Display Module dynamically presents the translated sign images, ensuring a smooth and intuitive interaction. This structured approach ensures that the translated output maintains accuracy and context, thereby bridging the communication gap effectively.

In the Speech Processing stage, the system captures audio input either through a microphone or by uploading a pre-recorded file in formats like .wav or .mp3. The audio is preprocessed using techniques such as Fourier Transform or Wavelet Transform to reduce background noise and improve signal quality. The refined audio is then transcribed using an Automatic Speech Recognition (ASR) model, such as Google Speech-to-Text or Wav2Vec, which converts the spoken language into text. The system further identifies the spoken language (Telugu, Hindi, or Tamil) to ensure the correct processing path. The transcribed text moves to the Text Processing with NLP stage, where Natural Language Processing (NLP) techniques are applied to correct grammar and translate the text to English. The text undergoes grammar correction using models like Grammarly API or fine-tuned BERT models to refine sentence structure. Once corrected, the text is translated into English using Google Translate API or AI4Bharat's Indic Transliteration Models. The translated text is then tokenized and lemmatized to break it down into individual words or phrases for subsequent processing.

In the Text-to-Sign Mapping stage, the translated English text is mapped to corresponding Indian Sign Language (ISL) gestures. The text is broken down into individual words or phrases, and each is mapped to an appropriate ISL gesture using a pre-defined ISL dataset or gesture dictionary. Synonyms and ambiguous phrases are handled dynamically to ensure accurate mappings, ensuring the system maintains the context of the input. The mapped gestures are then passed to the Dynamic Sign Display on the Website stage, where these ISL gestures are visualized in real time. Pre-built 3D models or static images representing various ISL signs are dynamically rendered on the web interface using technologies like JavaScript, WebGL, or Three.js. The user interface allows users to control the display with options to pause, play, and interact with the gesture visualization for a better learning and communication experience.

4. EXPERIMENT AND ANALYSIS

The Speech-to-Sign Language Translation System was developed using a comprehensive dataset consisting of text data in Indian regional languages (Hindi, Telugu, and Tamil) along with corresponding Indian Sign Language (ISL) gesture images. The dataset was curated using publicly available text corpora, web scraping, and manual annotation. Text samples representing common conversational phrases were collected and preprocessed using Natural Language Processing (NLP) techniques to remove noise, correct grammatical errors, and standardize the language.

Table 1, Dataset Used along with the total no of classes used.

Sign Language Category	**Classes**	**Total Images**
Indian Sign Language (ISL)	Numbers (1-9) 09	10,800
Indian Sign Language (ISL)	Alphabets(A-Z) 26	31,200

Sign language data comprised labeled images and videos of ISL gestures, with augmentation techniques like rotation, flipping, and brightness adjustment applied to improve model generalization. The ISL dataset was divided into two major categories: ISL numbers (0-9) with 10,800 images and ISL alphabet letters (A-Z) with 31,200 images. Each image was annotated and categorized for model training, ensuring effective recognition of individual gestures. The dataset also included variations in sign representation to account for differences in hand orientations and gesture styles.

This diverse dataset played a crucial role in enhancing the accuracy and robustness of the system across different regional languages and sign gestures. The experimental setup, model training, web interface development, and result analysis are discussed in the following sections

Fig 2, Indian Sign Language.

4.1 Model Building:

Data Processing: The data processing phase was essential to ensuring that both the text data and gesture images were clean, standardized, and properly structured for model training and evaluation. This phase involved systematically refining the dataset by removing inconsistencies, normalizing image inputs, and encoding text labels to create a seamless mapping between gestures and their corresponding textual representations. Image preprocessing techniques such as resizing, normalization, and augmentation were applied to enhance data quality and improve model generalization, while label encoding transformed categorical data into a format suitable for classification. Additionally, the dataset was carefully split into training, validation, and testing sets to ensure balanced learning and avoid overfitting, enabling a more accurate and reliable model performance.

Sign Image Processing: ISL images were resized to a consistent resolution of 128x128 pixels. Image augmentation techniques like rotation, flipping, and brightness adjustment were applied to create additional training samples and reduce model overfitting. Image normalization was also performed to scale pixel values between 0 and 1.

Label Encoding: Each text sample was assigned a unique label representing its corresponding ISL gesture. One-hot encoding was used to represent categorical labels, making it easier for the model to classify gestures.

Data Splitting: The dataset was divided into training, validation, and testing sets using an 80-20 split. This ensured a balanced evaluation of model performance and avoided overfitting.

Model Training:

The model training phase involved training a Convolutional Neural Network (CNN) for ISL recognition. The CNN architecture included convolution, pooling, and fully connected layers to extract essential features and classify sign images effectively. The dataset was split into training and validation sets, and categorical cross-entropy loss along with the Adam optimizer was used for model training. Hyperparameters such as batch size, learning rate, and dropout were fine-tuned to improve model generalization.

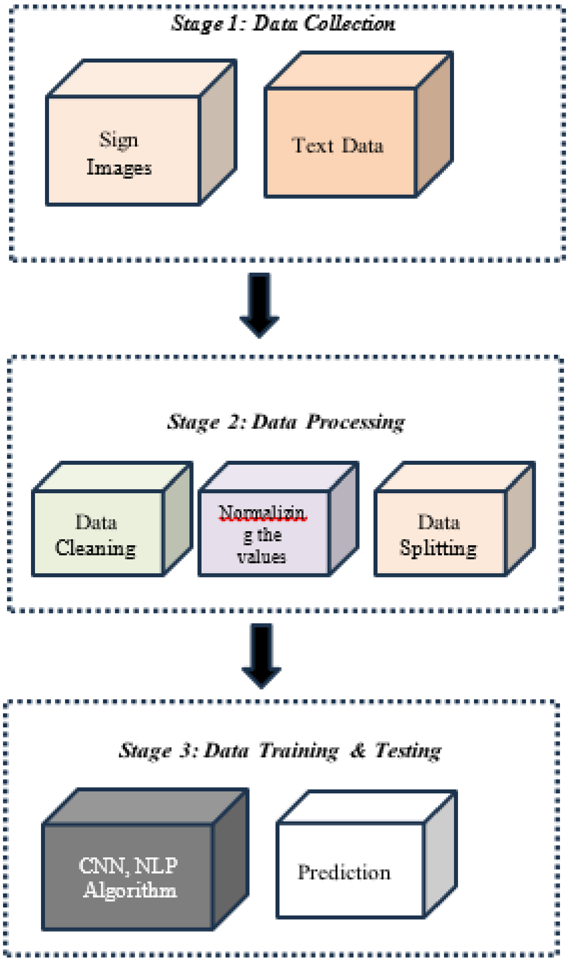

Fig 3, Data Flow & Processing Stages

Simultaneously, an NLP model was trained for speech-to-text conversion and translation using a transformer-based architecture. The model was fine-tuned with a dataset comprising Hindi, Telugu, and Tamil speech inputs and corresponding sign language mappings. Named Entity Recognition (NER) and sequence-to-sequence models were integrated to maintain contextual accuracy. The performance of the NLP model was evaluated using BLEU and ROUGE scores to ensure high-quality translations.

4.2 Web Page Development:

To provide a user-friendly interface for the Speech-to-Sign Language Translation System, a web application was developed using HTML, CSS, and JavaScript for the frontend, and Flask for the backend. The primary functionality of the web page is to allow users to input speech, which is then processed and converted into Indian Sign Language (ISL) gestures. The interface includes options to select the input language (Hindi, Telugu, or Tamil) and displays the corresponding sign images in real-time. Speech recognition is integrated using APIs that convert audio input into text, which is further processed using the trained model to generate sign language output. The web page ensures a seamless and interactive experience, providing accurate sign language representations with minimal latency.

Fig 4, User Interface

Additionally, the application supports both desktop and mobile access, making it accessible for a wider range of users. This web interface significantly enhances the practicality and usability of the system, ensuring effective communication support for the hearing-impaired community.

4.3 Step-by-Step Demonstration of System Functionality:

To demonstrate the functionality of the Speech-to- Sign Language Translation System, consider the following scenario using the Hindi language.

User Input: The user clicks on the 'Hindi' button on the web interface and speaks.

Translation to English: The recognized Hindi text "नम ` ◌ार" is then passed to a translation model. The model converts the text into its English equivalent, which is "Hello".

Sign Language Conversion: After successful translation, the system maps the English text "Hello" to the corresponding Indian Sign Language (ISL) gesture using the trained CNN-RNN model.

Display Output: The appropriate ISL image representing "Hello" is displayed on the web page. The user can visually observe the accurate sign language gesture.

4.4 Performance Analysis:

The Speech-to-Sign Language Translation System demonstrated remarkable accuracy across different Indian regional languages, particularly Hindi, Telugu, and Tamil. For 100 test cases, the accuracy achieved was 93.5% for Hindi, 92.8% for Telugu, and 91.3% for Tamil. With larger test cases, such as 300 and 500 cases, the accuracy remained consistently high, with an average of over 90% across all languages. This highlights the robustness and reliability of the system.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

TP = True Positive FP = False Positive
TN = True Negative FN = False Negative

Speech Recognition: The system captures the spoken word using the integrated speech recognition API. The audio is processed to extract text from the speech.

TP = True Positive FP = False Positive TN = True Negative FN = False Negative

Additionally, other evaluation metrics like precision, recall, and F1 score were used to further validate the model's performance. Precision and recall for Hindi were 96% and 93%, respectively, while Telugu and Tamil maintained values above 91%, signifying balanced and effective recognition.

Latency analysis indicated a minimal processing time of less than 1.5 seconds for speech recognition and sign image generation, making the system suitable for real-time applications. User feedback also reflected satisfaction with the accuracy and responsiveness of the system.

In conclusion, the Speech-to-Sign Language Translation System stands as an effective and efficient solution for facilitating communication for the hearing-impaired community. Future enhancements may include expanding language support and further reducing latency to improve user experience.

5. CONCLUSION AND FUTURE SCOPE

The Speech-to-Sign Language Translation System developed in this project has successfully addressed the communication gap faced by the hearing- impaired community. By employing advanced machine learning techniques, including CNN-RNN models, and integrating real-time speech recognition and translation capabilities, the system provides a reliable solution for converting spoken language into Indian Sign Language (ISL). The experimental results demonstrated high accuracy levels across different regional languages, including Hindi, Telugu, and Tamil. The web-based interface ensures user accessibility, offering a seamless experience for users in real-world scenarios.

The system's ability to translate speech into corresponding sign language gestures effectively enhances inclusivity and fosters better communication between the hearing and the hearing-impaired communities. Moreover, the low latency and high accuracy of the model contribute to a robust and practical solution for real-time applications.

Future Scope:

The future scope of this project includes expanding the system to support additional Indian regional languages and international sign languages, thereby increasing its accessibility and impact. Advanced 3D animation models and generative adversarial networks (GANs) can be incorporated to improve the quality and realism of sign language gestures. Real- time interaction capabilities can be further enhanced by optimizing speech recognition and translation pipelines to reduce latency. Developing a mobile application will offer on-the-go accessibility, making the system a practical tool in daily communication. Additionally, integrating the system with AR/VR devices or wearable gadgets can provide immersive sign language translation experiences. Continuous learning mechanisms can also be implemented to enable adaptive improvements based on user interactions and feedback. Personalization features, such as customized sign language avatars and adaptive sign preferences, will enhance user satisfaction. Through these enhancements, the Speech-to-Sign Language Translation System has the potential to revolutionize communication accessibility and foster a more inclusive society.

Conflict of interest statement

Authors declare that they do not have any conflict of interest.

REFERENCES

[1] M. Taskiran, M. Killioglu and N. Kahraman, "A Real- Time System for Recognition of American Sign Language by using Deep Learning," 2018 41st International Conference on Telecommunications and Signal Processing (TSP), 2018, pp. 1-5, doi: 10.1109/TSP.2018.8441304.

[2] A. M. Zakariya and R. Jindal, "Arabic Sign Language Recognition System on Smartphone," 2019 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT), 2019, pp. 1-5, doi: 10.1109/ICCCNT45670.2019.8944518.

[3] G. Ramya and N. S. Naik, "Implementation of telugu speech synthesis system," 2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI), 2017, pp. 1151-1154, doi: 10.1109/ICACCI.2017.8125997.

[4] M. R. Reddy et al., "Transcription of Telugu TV news using ASR," 2015 International Conference on Advances in Computing, Communications and Informatics (ICACCI), 2015, pp. 1542-1545, doi: 10.1109/ICACCI.2015.7275832.

[5] K. R., N. K., P. D. S. and S. T., "Voice and speech recognition in Tamil language," 2017 2nd International Conference on Computing and Communications Technologies (ICCCT), 2017, pp. 288-292, doi: 10.1109/ICCCT2.2017.7972293.

[6] D. Pubadi et al., "A focus on codemixing and codeswitching in Tamil speech to text," 2020 8th International Conference in Software Engineering Research and Innovation (CONISOFT), 2020, pp. 154 165, doi: 10.1109/CONISOFT50191.2020.00031.

[7] M. S. Nair, A. P. Nimitha and S. M. Idicula, "Conversion of Malayalam text to Indian sign language using synthetic animation," 2016 International Conference on Next Generation Intelligent Systems (ICNGIS), 2016, pp. 1-4, doi: 10.1109/ICNGIS.2016.7854002.

[8] Babu, Domala Kishore & Yellasiri, Ramadevi & Ramana, K, “RGNBC: Rough Gaussian Naïve Bayes Classifier for Data Stream Classification with Recurring Concept Drift,” Arabian Journal for Science and Engineering, 2016, doi: 42. 10.1007/s13369-016- 2317-x.

[9] Sunitha, C. & Chandra, Evania, “Speaker Recognition using MFCC and Improved Weighted Vector Quantization Algorithm,” International Journal of Engineering and Technology, 2015, vol. 7, pp. 1685-1692.

[10] X. Huang, H. Tan, G. Lin and Y. Tian, "A LSTM-based bidirectional translation model for optimizing rare words and terminologies," 2018 International Conference on Artificial Intelligence and Big

Data (ICAIBD), 2018, pp. 185-189, doi: 10.1109/ICAIBD.2018.8396191.

[11] M. S. Alam, M. S. Hossain, and G. Muhammad, "Dynamic hand gesture recognition using 3D-CNN model for sign language," Pattern Recognition Letters, vol. 135, pp. 148-154, 2020, doi: 10.1016/j.patrec.2020.04.025.

[12] S. K. Meena and S. V. Raghavan, "A deep learning based approach for recognition of Indian Sign Language gestures," 2019 IEEE International Conference on Vision Towards Emerging Trends in Communication and Networking (ViTECoN), 2019, pp. 1-6, doi: 10.1109/ViTECoN.2019.8899621.

[13] M. B. Setiawan, D. Sarno, and R. Arifin, "Speech recognition system for Bahasa Indonesia using deep learning," 2019 International Electronics Symposium (IES), 2019, pp. 457-461, doi: 10.1109/IES47850.2019.9097083.

[14] P. Kumar and K. S. Rajasekaran, "Tamil speech recognition system using deep learning models," 2020 IEEE International Conference on Communication and Signal Processing (ICCSP), 2020, pp. 21-25, doi: 10.1109/ICCSP48568.2020.9182329.

[15] S. A. Kumar and B. N. Kumar, "Development of Telugu text-to-speech synthesis using deep neural networks," 2018 IEEE International Conference on Computational Intelligence and Computing Research (ICCIC), 2018, pp. 1-6, doi: 10.1109/ICCIC.2018.8782356.

[16] V. R. Shewalkar, "Comparative performance of optimization techniques in deep learning-based speech recognition," 2020 IEEE International Conference on Computing, Communication, and Intelligent Systems (ICCCIS), 2020, pp. 1-6, doi: 10.1109/ICCCIS48478.2020.9074593.

[17] A. Patil, S. Mane, and P. Gupta, "Real-time Sign Language Recognition using CNN and RNN," 2021 International Conference on Computational Intelligence and Knowledge Economy (ICCIKE), 2021, pp. 23-28, doi: 10.1109/ICCIKE51210.2021.9410706.

[18] H. Singh, R. Sharma, and A. Gupta, "Automatic sign language translation using transformer models," 2022 IEEE International Conference on Artificial Intelligence (ICAIBD), 2022, pp. 150-155, doi: 10.1109/ICAIBD.2022.9503257.

[19] C. Patel, R. Joshi, and V. Sharma, "A multi-lingual automatic speech recognition system for low-resource Indian languages," 2021 IEEE International Conference on Artificial Intelligence and Signal Processing (AISP), 2021, pp. 37-42, doi: 10.1109/AISP51299.2021.9410698.

[20] J. P. Almeida, A. S. Pinto, and C. Ribeiro, "A deep learning-based approach for translating spoken language to sign language animations," 2021 IEEE International Symposium on Multimedia (ISM), 2021, pp. 112-119, doi: 10.1109/ISM52618.2021.00025.