

1. Introduction

Even though communication is a basic human need, millions of deaf and hard-of-hearing people around the world continue to encounter obstacles in their everyday lives because of accessibility issues. Although machine translation between sign language and spoken language is still a major technical issue, sign language is an essential form of communication among deaf communities. In India, regional sign languages, such as Assamese Sign Language, are underrepresented in research and development, creating a pressing need for intelligent systems specifically designed for these linguistic contexts.

Natural language processing, speech recognition, and computer vision have all been transformed by recent developments in deep learning. Long-term memory networks, transformer-based designs, convolutional neural networks, and recurrent neural networks have all shown exceptional performance in identifying intricate visual and sequential patterns. In sign language recognition, where precise capture and interpretation of hand movements, face expressions, and movement dynamics are required, these models are especially well-suited.

By converting spoken signs into spoken language, a deep learning-based speech recognition system for Assamese Sign Language seeks to close the communication gap. Typically, this system combines speech synthesis, sequence modeling, gesture detection, feature extraction, and video capture. Deep learning allows the system to automatically extract temporal and geographical information from gesture data with little need for human guidelines.

This system's development represents both a technological and a social contribution. It helps preserve and digitize regional sign languages, encourages diversity, and supports assistive

communication devices. Additionally, creating a dataset and model for Assamese Sign Language recognition establishes the groundwork for upcoming studies in resource-constrained language technology (Kagalkar, 2025).

To achieve high accuracy, instant performance, and usability in real-world communication scenarios, this work focuses on the design and implementation of an intelligent voice recognition system based on deep learning that is specifically tailored for Assamese Sign Language.

1.1. Problem definition

Most current systems concentrate on widely used languages like American or British Sign Language, despite notable breakthroughs in sign language recognition technologies. There is currently little technological assistance for Assamese Sign Language, and it is still mainly underdeveloped.

Deaf people in Assam face communication challenges due to the lack of automated Assamese Sign Language recognition systems, which restrict their access to public services, work, and education. Thus, an intelligent, effective, and real-time system that uses deep learning techniques to convert Assamese sign movements into speech is required.

1.2.Thesis Objectives

Main Objective

To design and develop a deep learning-based speech recognition system for Assamese Sign Language.

Specific Objectives

1. To collect and preprocess a dataset of Assamese sign gestures.

2. To develop a deep learning model for sign gesture recognition.
3. To implement sequence modeling for dynamic sign interpretation.
4. To integrate a speech synthesis module for real-time audio output.
5. To evaluate system performance using standard metrics (accuracy, precision, recall, F1-score).

1.3.Thesis questions

- How can Assamese Sign Language identification accuracy be increased using deep learning techniques?
- Which neural network design works best for recognizing dynamic gestures?
- Is it possible for the system to translate in real time?
- What difficulties arise when using local sign language with limited resources?

1.4.Thesis contributions

Significant technological, scholarly, and social contributions to assistive technology and sign language recognition are made by this thesis, especially for regional languages like Assamese Sign Language that have few resources:

- **Creation of a New Dataset for Assamese Sign Language:** The development of an organized and indexed dataset for Assamese Sign Language is a significant contribution to this study. Because there aren't many publicly accessible datasets for Assamese Sign Language, this work helps create a foundational resource that will aid in further study and advancement in the area.

- **Developing a Framework for Deep Learning-Based Recognition:** In this thesis, a deep learning architecture that combines long-term memory networks (LSTMs) for chronological modeling and convolutional neural networks (CNNs) for spatial feature extraction is proposed and implemented. The accuracy of dynamic gesture detection is increased by this hybrid approach, which makes it possible to capture hand shape and movement dynamics efficiently.
- **Real-Time Sign-to-Speech Conversion System:** In contrast to numerous theoretical models, this research creates a working prototype that can instantly translate recognized Assamese Sign English signs into spoken English. Combining speech synthesis with sign recognition offers a complete assistive communication solution.
- **Technology Support for Languages with Limited Resources:** Most sign language recognition systems in use today concentrate on languages that are widely spoken. By extending the use of deep learning to resource-constrained regional sign language, this thesis advances inclusive artificial intelligence and lessens technological inequality.
- **Comparative analysis and performance evaluation:** Using common indicators including accuracy, precision, recall, and the F1 score, the study offers a thorough performance evaluation. To identify the most effective architecture for Assamese Sign Language identification, it also contrasts several model configurations.
- **Accessibility and Social Impact:** Apart from its technical achievements, this initiative fosters social inclusion by offering a tool that helps the hearing and deaf populations in Assam communicate with one another. The system promotes accessibility in public services, employment, and education.

1.5. Thesis organization

The study challenge, research methodology, application, and findings of the suggested deep learning-based speech recognition system for Assamese Sign Language are presented in many chapters of this thesis.

Chapter 1: Introduction

The study's background is given in this chapter, along with the difficulties the deaf community faces in communicating and the significance of creating an automated system for Assamese Sign Language recognition. The study's organization, scope, constraints, problem statement, research aims, and research questions are also described.

Chapter 2: Literature Review

This chapter reviews current research in sign language recognition, speech recognition systems, and deep learning techniques such as convolutional neural networks (CNNs), recurrent neural networks (RNNs), long-term memory networks (LSTMs), and transformer-based models. It highlights research gaps, talks about recent advancements, and argues for the necessity of an Assamese Sign Language-specific system.

Chapter 3: Methodology

The general system design and research technique are covered in this chapter. The methods for gathering data, preparing the dataset, preprocessing, feature extraction, model design, training,

and text-to-speech module integration are all covered in this chapter. The workflow and system architecture are also presented.

Chapter 4: Execution and Experimental Outcomes

This chapter describes how the suggested system is implemented using TensorFlow/ByTorch, and Python. It presents training settings, hyperparameter settings, and evaluation metrics. To evaluate system performance, experimental findings are assessed using the accuracy measure, the fine-tuning metric, the recall metric, the F1 metric, and confusion matrices.

Chapter 5: Analysis and Discussion

The experimental data are interpreted, model performance is compared, and the system's advantages and disadvantages are discussed in this chapter. It also discusses the difficulties encountered during development and assesses the efficacy of the suggested technique in practical implementations.

Chapter 6: Final Thoughts and Upcoming Projects

This chapter provides an overview of the research results, highlights the main contributions of this thesis, and makes recommendations for future developments, including database enlargement, mobile deployment, and continuous sentence recognition.

2. Literature review

Unquestionably, communication is difficult for those who are deaf and mute. They use sign language to interact with others and with each other. However, a sizable portion of the population that communicates verbally and in writing is not familiar with sign language. Thus, the development of technical instruments for sign language interpretation is obviously necessary. For most of the world's languages, numerous studies have been done to use technology to recognize sign language. However, there is still need for the development of methods and instruments tailored to local dialects of sign language. In addition to over 19,000 languages spoken locally as mother tongues, India is home to 22 modern languages. The goal of this project is to provide a technical method for identifying Assamese Sign Language, one of India's 22 contemporary languages. This effort aimed at developing a system for recognizing hand movements in Assamese Sign Language using machine learning techniques. The dataset was prepared using both two-dimensional and three-dimensional photographs of Assamese gestures. To find features in the photos, the MediaPipe framework was used. Nine static motions with vowel and consonant sounds were included in a dataset of 2094 data points for Assamese Sign Language (অ, ই, ঈ, উ, এ, ও, য়, ঙ, ঠ). A feedback forward neural network was trained using this dataset, and the model's accuracy was 99%. The findings show that the approach used in this study is successful in identifying letters and other Assamese Sign Language motions. The recognition of signs and gestures in numerous additional Indian regional languages can also be evaluated and implemented using this technique (Bora, 2023). Communicating with hearing people is extremely difficult for deaf people. The shortage of qualified interpreters makes hiring a human interpreter—someone who converts sign language into a language that the hearing person can understand—an ineffective solution.

Consequently, the best and most effective solution is a Sign Language Recognition System (SLRS), which lowers the communication barrier between hearing and deaf people by automatically translating sign language into text or speech without the need for an interpreter. The use of computer vision technologies for Indian Sign Language (ISL) word recognition is covered in this research. Excessive preprocessing, such as collecting characteristics from a wide sequence of frames, causes inefficiencies in the current computer vision-based methods for ISL word recognition. To address the issue of excessive preprocessing, (Das, A deep sign language recognition system for Indian sign language, 2023) present the CNN-BiLSTM Hybrid Language Recognition System (HCBSLR), a novel computer vision-based ISL recognition system. The suggested model eliminates superfluous or redundant frames using a keyframe extraction technique based on histogram difference (HD), which increases system accuracy and efficiency. The HCBSLR system extracts spatial features using the VGG-19 algorithm and temporal features using a Bidirectional Long Term Memory Network (BiLSTM). Comparing the suggested system to some current SLRS systems, the average accuracy was 87.67%. The suggested HCBSLR system outperforms the current SLRS systems in terms of accuracy and efficiency, according to experimental results (Das, 2023).

Deaf and hard-of-hearing people utilize sign language, which is a visual-kinetic language. It uses a variety of body postures, face expressions, hand gestures, and movements. Because of its distinct grammatical and morphological norms, sign language is an essential mode of communication. One of the biggest challenges is the communication gap between sign language users and non-users. In India, there are over 63 million persons who are deaf or hard of hearing. Therefore, research in sign language recognition demonstrates enormous potential for improving the quality of life of individuals with hearing impairments and promoting

communication and social inclusion. Due to its critical role in human-machine interaction, accessibility, interpretation, educational tools, and assistive communication, sign language recognition has attracted a lot of interest in recent years. This paper examines the most recent advancements in Indian sign language recognition systems, going into detail about standard datasets, popular methods, and system performance features. Lastly, it draws attention to the difficulties and potential of these new technologies (Hussain, 2023).

According to (Chetia, 2025) a sophisticated deep learning approach—specifically, Convolutional Neural Networks (CNNs) for Assamese Sign Language (ASL) recognition is presented. The three primary stages of the suggested framework are real-time gesture recognition, model construction, and data preparation. The OpenCV frame capture library and MediaPipe hand feature recognition were used to create a dataset of ten static motions that represented various Assamese alphabet letters. The hand gesture photos were separated into training and test sets after the data was preprocessed. The CNN model was then trained to categorize the motions, attaining remarkable accuracy, recall, and F1 scores in all ten categories in addition to 100% flawless correctness. The outcomes show how well the CNN model extracts and learns the essential characteristics of Assamese Sign Language motions, allowing the system to precisely identify and translate signs into text in real time. Communication between the hearing and deaf people in Assam could be greatly improved by this real-time recognition capabilities. Future research will concentrate on enhancing the system for real-time deployment in mobile applications and broadening the dataset to incorporate dynamic gestures.

An advanced program called the Interactive Sign Language Learning System was created to make learning sign language easier. Among the many important characteristics of this all-

inclusive system are built-in audio output, text-to-gesture conversion, multilingual support, and the ability to recognize letters and words in sign language. To precisely interpret hand gestures and movements, the system makes use of sophisticated sign language recognition algorithms as well as methods like image processing and machine learning. A variety of computer vision methods, such as recurrent neural networks (RNNs) or convolutional neural networks (CNNs), can be used to evaluate and categorize input images or videos into the proper sign language words or symbols for letter and word identification. In text-to-gesture conversion, natural language processing (NLP) algorithms are used to analyze the meaning of the input text and produce appropriate sign language representations by correlating it with relevant gestures or movements. The efficiency of the algorithms employed, and the caliber of the training data determine the system's accuracy, which is determined by how many sign language symbols or words are successfully identified. The system seeks to attain high levels of accuracy in text-to-gesture conversion and sign language recognition tasks through meticulous development and training (Ravikumar, 2024).

The best communication method for those who are hard of hearing is sign language. There is still a communication gap between the hearing and hearing communities in spite of technological improvements. By creating an automated system that uses avatars to translate speech into Hindi Sign Language, this research seeks to close this gap (SISLA). There are three stages to the system's operation: (a) Phase 1: In a speaker-independent setting, speech recognition of single words in Punjabi, Hindi, and English. (b) Phase 2: Hindi Sign Language translation from the original language. (c) Using a 3D avatar built on the HamNoSys technology to represent motions in Hindi Sign Language. Requirements analysis, data collecting, technological development, and evaluation are the four primary SISLA system

implementation modules. The efficiency of the system is enhanced by its multilingual capacity. The speech recognition models were trained and tested using speech sample files in Punjabi (12600, 4193), Hindi (12610, 4211), and English (12660, 4218). The suggested trained models obtained accuracy of at least 91%, 89%, and 89% for English, Punjabi, and Hindi, respectively, according to experimental machine translation findings. In order to assess the sign error rate through feedback, sign language specialists were also consulted. It was suggested that the system be improved in the future by combining sentence-level translation with non-manual sign language elements. Based on survey findings, usability tests verify that the suggested SISLA system is appropriate for communication and education for individuals with hearing impairments (Dhanjal, 2022).

3. Conclusion

Particularly for regional and resource-constrained sign languages like Assamese Sign Language, communication barriers between the deaf and hearing communities continue to be a major social and technological problem. Assamese Sign Language has not gotten enough attention in research and system development, despite significant advancements in sign language identification for commonly spoken languages. This disparity emphasizes the need for technical solutions that are intelligent, effective, and easy to use and that are specific to this linguistic setting.

To translate sign language signals into spoken language in real time, this thesis suggests a deep learning-based speech recognition system for Assamese Sign Language. The system combines long-term memory networks (LSTMs) for chronological modeling, convolutional neural networks (CNNs) for spatial feature extraction, and computer vision algorithms.

Combining these techniques allows the system to precisely identify two crucial components of sign language recognition: hand form and movement dynamics. Comprehensive sign-to-speech conversion is made possible by the inclusion of a text-to-speech module, which makes the system useful and applicable in everyday communication situations. According to a survey of the research, convolutional neural network (CNN)-based architectures, keyframe extraction methods, and hybrid deep learning models have all demonstrated good accuracy in a variety of sign language recognition tasks. Assamese Sign Language, however, receives less attention in the majority of existing systems, which concentrate on Indian Sign Language or other important languages. By offering a structured dataset, putting in place a tailored deep learning architecture, and assessing the system using common performance metrics including accuracy, fine-tuning, recall, and F1, this study fills this gap.

The experimental findings show that deep learning methods are quite successful at identifying motions in Assamese Sign Language, especially when both spatial and temporal aspects are modeled. The suggested approach has great real-time application potential and scalability despite obstacles such restricted dataset availability, user gesture variability, and environmental limitations.

In summary, the contribution of this thesis goes beyond the technical aspects of creating a deep learning-based functional speech recognition framework. Additionally, it has a social component that improves the Assamese deaf community's accessibility and inclusion. Future advancements like continuous sentence recognition, mobile apps, bigger datasets, and integration with multilingual platforms are all made possible by this system. In the end, our study contributes to the larger goal of inclusive AI and fair access to technology for underrepresented linguistic groups.

4. References

- Bora, J. (2023). Real-time Assamese Sign Language Recognition using MediaPipe and Deep Learning. 1384-1393. Retrieved from <https://www.sciencedirect.com/science/article/pii/S1877050923001175>
- Chetia, H. (2025). Advancing Communication: A CNN-Powered Framework for Assamese Sign Language Recognition. 5(2). Retrieved from <https://journals.dbuniversity.ac.in/ojs/index.php/AJEEE/article/view/4354>
- Das, S. (2023). A deep sign language recognition system for Indian sign language. 1469–1481. Retrieved from <https://link.springer.com/article/10.1007/s00521-022-07840-y>
- Dhanjal, A. S. (2022). An automatic machine translation system for multi-lingual speech to Indian sign language. Retrieved from <https://link.springer.com/article/10.1007/s11042-021-11706-1>
- Hussain, A. (2023). Advancements in Indian Sign Language Recognition Systems: Enhancing Communication and Accessibility for the Deaf and Hearing Impaired. 12(2). Retrieved from <https://www.ajesjournal.org/index.php/ajes/article/view/4132>
- Kagalkar, R. M. (2025). Enhancing Communication for the Deaf: A Machine Learning Approach to Real-Time Kannada Sign Language Translation. Retrieved from <https://link.springer.com/article/10.1007/s40031-025-01279-0>
- Ravikumar, S. (2024). Interactive Sign Language Learning System using Deep Learning. Retrieved from <https://ieeexplore.ieee.org/abstract/document/10760668>