Indian sign Language to Text/speech translation using computer vision and NLP.

Allan D'souza
Computer Science and Technology Department
Dayananda Sagar University
Bengaluru, India
Email: allanferns007@gmail.com

Swaroop K R
Computer Science and Technology Department
Dayananda Sagar University
Bengaluru, India
Email: swaroopkr1628@gmail.com

Sarah Catherine
Computer Science and Technology Department
Dayananda Sagar University
Bengaluru, India
Email: eng21ct0034@dsu.edu.in

Swathi S
Computer Science and Technology Department
Dayananda Sagar University
Bengaluru, India
Email: swathisadashiva03@gmail.com

Abstract—**Indian Sign Language (ISL) is a vital communication medium for the Deaf and Hard of Hearing (DHH) community in India, enabling individuals to express emotions, share ideas, and engage in everyday interactions. However, a significant communication barrier exists between the DHH community, which primarily uses ISL, and individuals who rely on spoken language. This barrier impacts accessibility to education, healthcare, and other essential services, particularly due to the limited availability of datasets and technological solutions tailored to Indian regional languages, such as Kannada.**

In this research, we address these challenges by building a novel dataset for Indian Regional Kannada Sign Language (IRKSL), encompassing character, word, and sentence-level gestures. Furthermore, we propose a comprehensive solution for bidirectional communication between ISL users and non-signers. Our approach combines speech recognition, natural language processing, and computer vision techniques to enable real-time audio-to-ISL conversion and ISL-to-Kannada translation. The proposed system prioritizes accuracy, user-friendliness, and adaptability to regional variations, fostering effective communication and inclusivity. This work aims to bridge the gap between the DHH community and the wider society, empowering individuals and promoting seamless interaction across diverse linguistic contexts.

Keywords—Indian Sign Language (ISL), speech recognition, natural language processing, computer vision, bidirectional communication, Indian Regional Kannada Sign Language (IRKSL), real-time translation.

I. Introduction

Communication is one of the most essential aspects of daily life, enabling individuals to express ideas, share opinions, and integrate into society. However, for millions of Deaf and Hard of Hearing (DHH) individuals, the inability to hear and speak poses significant challenges to effective communication, often isolating them from mainstream society. In India, this challenge is particularly pronounced due to the linguistic diversity and the limited awareness of Indian Sign Language (ISL), a natural visual-gestural language used by the DHH community. According to the World Health Organization (WHO), India is home to approximately 63 million people with hearing impairments, including 3.3 million in Karnataka. Despite this substantial population, communication barriers persist, necessitating technological solutions to bridge the gap between the DHH community and the hearing population.

Sign language, including ISL, is a natural language with its own grammar and structure, differing significantly from spoken and written languages. Unlike commonly held misconceptions, sign language is not universal. ISL has distinct features such as non-manual expressions (e.g., facial expressions, eye gaze), hand gestures for number representation, and unique sentence structures. For example, WH-questions are placed at the end of interrogative sentences, and adjectives follow nouns. These linguistic nuances make the development of automated translation systems complex yet essential for promoting inclusivity and accessibility.

Facts About Sign Language

Some key facts about sign language, which highlight its uniqueness as a natural language, include:

1. Sign languages are not the same across the world; they vary by region and culture.
2. Sign languages have their own grammar, distinct from spoken languages.
3. The dictionary of sign language is smaller compared to other languages, and finger-spelling is used for unknown words.
4. Adjectives are often placed after the noun.

5. Sign languages do not use suffixes and always operate in the present tense.
6. Articles such as "a" or "the" are not used, and personal pronouns like "I" are replaced with "me."
7. Sign languages lack gerunds and rely heavily on non-manual expressions, such as facial movements and body postures.
8. They were not invented by hearing people but developed naturally within the Deaf community.

Indian Sign Language Grammar

1. Numbers are represented through specific hand gestures for each hand.
2. Signs for family relationships are preceded by male or female identifiers.
3. WH-questions in ISL are placed at the end of the sentence.
4. ISL incorporates non-manual gestures such as mouth patterns, body postures, head positions, and eye gaze.
5. Temporal concepts like past, present, and future are expressed through signs for "before," "then," and "after."

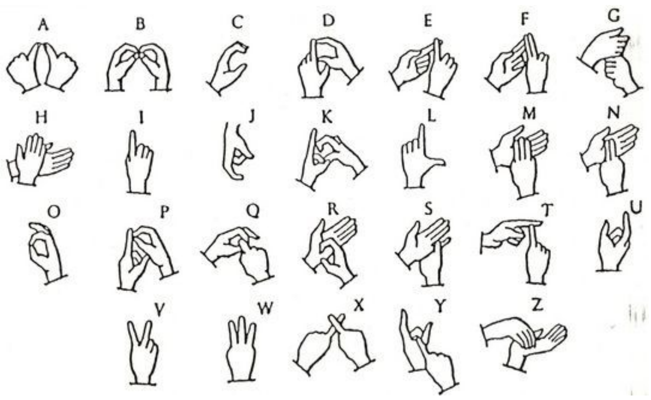


fig. 1 Indian Sign Language Alphabets

Existing Text-to-Sign Language Translation Systems

To facilitate communication, several text-to-sign language translation approaches have been explored:

1. **Direct Translation System** This approach performs word-by-word conversion without considering the sentence's context, meaning, or grammatical structure. In cases like English to ISL translation, this method falls short as ISL's word order often differs from the source language. Direct translation systems lack syntactic analysis, making them inadequate for conveying accurate meaning in ISL.
2. **Transfer-Based Translation System** This system incorporates syntactic and semantic transformations, converting the source language into intermediate text before applying linguistic rules to generate the target language. Known as rule-based translation, it provides improved accuracy over direct translation by considering the linguistic structure of both the source and target languages.

3. **Interlingua-Based Translation System** This approach generates a language-independent semantic structure (Interlingua) by performing semantic analysis on the input text. The Interlingua is then used to produce the target language. By bypassing direct and transfer-based methods, this system offers a more flexible and adaptable solution for multilingual translation needs.

Challenges in ISL Recognition

Despite advancements, several challenges persist in developing ISL recognition systems:

- **Dynamic Hand Movements:** ISL involves complex and fluid hand gestures, making it difficult for conventional recognition systems to capture and interpret accurately.
- **Variations in Signing Styles:** Individual differences in signing styles add complexity to creating universally applicable models.
- **Background Disturbances:** Environmental factors, such as cluttered backgrounds, can interfere with gesture recognition.
- **Limited Datasets:** Publicly available datasets for ISL, particularly regional variants like Kannada Sign Language, are scarce, hindering the development of robust recognition models.

Description of the Problem

The need for an audio-to-ISL conversion system arises from the persistent communication gap between the DHH community and the hearing population. While ISL is a structured and expressive language, it remains inaccessible to the majority of people who do not know sign language. This lack of mutual understanding impedes inclusivity and equity, leaving DHH individuals unable to engage in everyday interactions, comprehend spoken information, or express their needs effectively in real-time settings. Furthermore, the diversity of sign languages worldwide, coupled with regional variations within ISL, adds an additional layer of complexity to developing a universally adaptable solution.

Existing technological solutions for sign language recognition face several limitations. Many systems are constrained by the lack of publicly available datasets, especially for regional variants like Kannada Sign Language, which are crucial for sign recognition and translation tasks. Without these datasets, the development of accurate and efficient systems remains a significant challenge. Additionally, gesture-based sign language recognition systems must contend with issues such as dynamic hand movements, individual variations in signing styles, and environmental distractions, all of which complicate accurate recognition.

To address these challenges, this research focuses on developing a comprehensive system for recognizing ISL and translating it into Kannada, a widely spoken regional language. The solution utilizes Long Short-Term Memory (LSTM) networks, a type of recurrent neural network known for its ability to model sequential data effectively. By combining LSTM networks with computer vision techniques, the system can accurately recognize and classify ISL gestures in real-time, ensuring quick and reliable translation.

A key component of this work is the creation of a novel dataset for Indian Regional Kannada Sign Language (IRKSL). This dataset includes character, word, and sentence-level gestures, addressing the scarcity of resources for regional sign languages. The system also incorporates speech recognition and natural language processing to enable bidirectional communication between signers and non-signers, facilitating both audio-to-ISL and ISL-to-Kannada translation.

This research contributes to bridging the communication gap between the DHH community and the wider society. By leveraging advanced machine learning techniques and linguistic insights, the proposed system prioritizes accuracy, user-friendliness, and adaptability to regional variations. The development of such technology empowers DHH individuals to participate more fully in education, healthcare, and social activities, fostering inclusivity and enhancing their quality of life. Furthermore, the focus on Kannada Sign Language highlights the importance of regional representation, ensuring that technological solutions cater to the linguistic diversity of India.

II. LITERATURE SURVEY

Li et al. propose a detailed and innovative framework for overcoming the complexities inherent in translating American Sign Language (ASL). They combine Natural Language Processing (NLP) techniques with Computer Vision to address the nuances of ASL, such as its limited vocabulary and polysemous signs. The methodology includes context-aware autocorrect to enhance classification accuracy and machine translation techniques to improve language conversion. Their study reports a BLEU-4 score of 31.758, showcasing the effectiveness of this multi-modal approach. Future work suggests refining autocorrect and exploring hybrid models that merge Neural and Statistical Machine Translation (NMT and SMT) to further enhance performance. Gupta et al. focus their research on Indian Sign Language (ISL) translation, which presents unique grammatical and contextual challenges. They employ video-based methods to convey ISL more effectively and preprocess English text into ISL grammar through techniques such as parsing, lemmatization, and animated video rendering. This approach not only aligns with ISL's grammar rules but also eliminates unnecessary elements like stop words, resulting in improved translation accuracy and reduced processing overhead. Kanvinde et al. propose a bidirectional translation system for sign language and speech using a combination of CNN and RNN models. Their system achieves real-time gesture recognition with high accuracy for Argentinian Sign Language datasets, eliminating the need for additional hardware such as gloves. The integration of temporal and spatial features allows the model to operate effectively in real-time scenarios.

Alaftekin et al. take a different approach by optimizing the YOLOv4-CSP model for Turkish Sign Language (TSL) recognition. This optimization incorporates CSPNet, Mish activation function, CIoU loss, and Transformer blocks, which significantly enhance both accuracy and inference speed. The model achieves precision of 98.95%, recall of 98.15%, and a mean average precision (mAP) of 99.49%. By eliminating preprocessing and introducing advanced data augmentation techniques, the study demonstrates the model's robustness in diverse environments. Future directions include expanding the dataset and integrating hybrid approaches that combine vision-based recognition with Natural Language Processing for gesture-to-text translation. Saini et al. explore deep learning techniques for ISL recognition, emphasizing the importance of diverse datasets and robust model architectures. Their evaluation highlights ResNet-50 as the most effective model, achieving an accuracy of 99.26% and excelling in handling variations in lighting, angles, and backgrounds. The authors advocate for integrating deep learning with NLP to enable real-time translation and propose expanding datasets to include more gestures and phrases for broader applicability.

Sincan et al. present Spotter+GPT, a novel hybrid Sign Language Translation (SLT) system that combines a sign spotter with a Large Language Model (LLM). This approach bypasses traditional gloss-based methods, reducing the dependency on manual annotations while maintaining high semantic accuracy. The system leverages datasets for German Sign Language and demonstrates improved BLEU and BLEURT scores, offering a scalable solution for diverse SLT applications. Fang et al. introduce SignLLM, a multilingual Sign Language Production (SLP) model built on the Prompt2Sign dataset. This dataset standardizes pose information from videos to enable efficient training of sequence-to-sequence models. SignLLM incorporates reinforcement learning to optimize training processes and supports multilingual functionality across eight sign languages. The model's state-of-the-art performance addresses challenges such as dataset diversity and computational efficiency, paving the way for broader adoption in multilingual sign language translation. [1-7]

Research gap

The research described in the related work addresses the need for a more precise and effective Indian Sign Language (ISL) recognition system. Despite the increasing number of studies on sign language recognition, ISL recognition research remains scarce. A review of the existing literature reveals that the majority of studies have focused on American Sign Language (ASL) or other sign languages, such as British Sign Language (BSL), Australian Sign Language (Auslan), and Chinese Sign Language (CSL). Few studies have examined the recognition of ISL, and the performance of existing systems is frequently unsatisfactory due to the complexity of ISL and the lack of a comprehensive dataset. Traditional methods rely on high computational data by processing videos which require huge storage of datasets and even more advanced systems for model training. A simpler approach and easy-to-train models are thus required to enhance the real-time detection of Indian Sign Language (ISL). The methodology proposed in this system can be used by anyone to train on their own facial expressions and hand gestures thus reducing the chances of inaccurate classification and providing a robust system for the detection and translation of sign language. As proposed, our system extracts only the key points and saves them in a numpy array, thus reducing the size of the dataset and faster processing of models. We eliminate the use of image operations and storage of data-intensive videos, making the model much simpler and easy to recognize sign language. The research void in this field is therefore the development of a more robust and accurate ISL recognition system that can effectively recognize both static and gesture sign languages and that can be tested and evaluated using a comprehensive dataset of ISL. This would enhance the communication skills of deaf and hard-of-hearing individuals who rely on ISL and promote social inclusion within this community.

III. METHODOLOGY

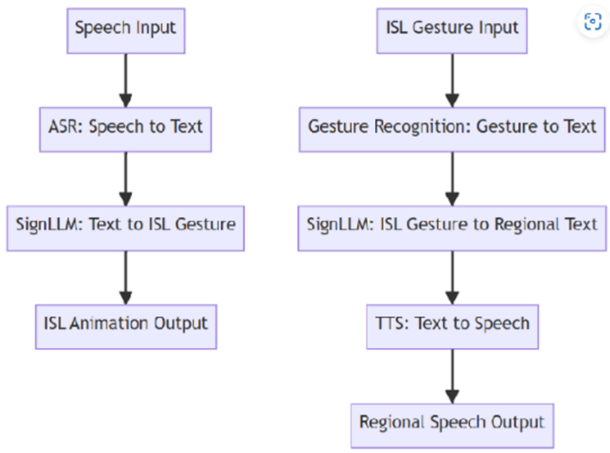


Fig 1. Framework of the model

A. Methodology

The proposed bidirectional Indian Sign Language (ISL) translation system is designed as a modular framework to facilitate effective two-way communication between ISL users and individuals who rely on spoken or written language. The system leverages advanced technologies in **Automatic Speech Recognition (ASR)**, **Natural Language Processing (NLP)**, and **Computer Vision**, ensuring accuracy and inclusivity. Below is a step-by-step breakdown of the system architecture and processes:

1. Speech-to-Sign Translation: This module translates spoken language into ISL, enabling effective communication from non-ISL users to ISL users. It consists of the following components:

a. Speech Input

Users provide spoken input through a microphone connected to the system. The system accepts inputs in English or supported regional languages, such as Kannada, ensuring linguistic flexibility.

b. Automatic Speech Recognition (ASR)

The ASR module transcribes spoken language into text with high accuracy.Regional accents, pronunciation variations, and noise suppression are handled using deep learning-based ASR models. The output text serves as the input for further processing.

c. SignLLM (Text-to-ISL Gesture Conversion)

The text from the ASR module is processed using a custom **Sign Language Language Model (SignLLM)**.

The SignLLM:

- Transforms the text into ISL-compliant grammar, which differs significantly from spoken or written grammar.
- Converts the text into a sequence of ISL gestures, including the integration of non-manual signals (e.g., facial expressions).
- Ensures semantic and contextual accuracy during the conversion process.

d. ISL Animation Output

The ISL gesture sequence is rendered as animated signs on a screen.This output enables ISL users to understand the spoken input visually and interact naturally.Animations are designed to be clear, dynamic, and expressive, incorporating non-manual features where necessary.

2. Sign-to-Speech Translation: This module allows ISL users to communicate with individuals who rely on spoken language. It involves the following stages:

a. ISL Gesture Input

ISL users provide input via gestures, captured in real-time using a camera or wearable sensor.The input includes both **manual signs** (hand gestures) and **non-manual signals** (facial expressions, body movements).

b. Gesture Recognition

Using advanced computer vision algorithms and deep learning models like **Long Short-Term Memory (LSTM)** networks, the system identifies and processes the gestures. Dynamic gestures (sequences of motions over time) and static gestures (single hand positions) are recognized. The system also analyzes non-manual signals to provide contextual and emotional accuracy.

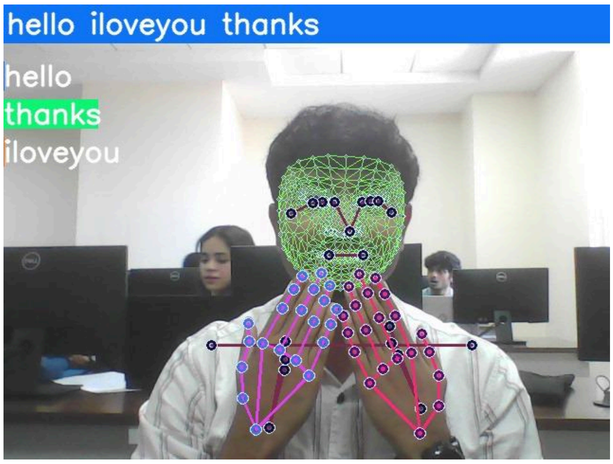
c. SignLLM (ISL Gesture to Regional Text)

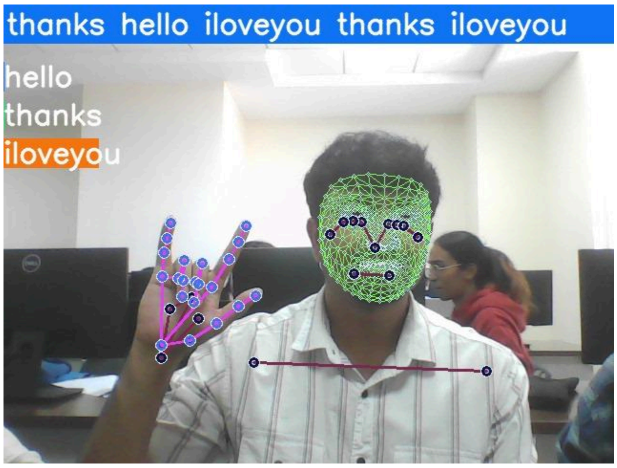
The recognized gestures are translated into text using the SignLLM. The model applies ISL-specific grammatical rules to ensure the text output accurately represents the ISL input. The output is translated into English or regional languages, such as Kannada, depending on the context.

d. Text-to-Speech (TTS)

The translated text is passed to a **Text-to-Speech (TTS)** engine, which converts it into natural-sounding speech. The TTS module supports multiple languages and accents, ensuring clarity and adaptability.

IV. Results





V. Conclusion

Our sign language recognition and translation system marks a significant step toward bridging communication gaps for the hearing-impaired community. By integrating advanced technologies like computer vision, deep learning, and NLP, it delivers real-time, bidirectional translation between ISL, text, and speech while adhering to ISL grammar and supporting regional languages like Kannada.Despite challenges such as dataset limitations and real-world variability, our solution demonstrates strong potential for inclusivity and accessibility. Future efforts will focus on improving datasets, accuracy, and multilingual support. This project lays the groundwork for transformative assistive technology, fostering a more inclusive and connected society.

B. References

[1]. Jacky Li, Jaren Gerdes, James Gojit, Austin Tao, Samyak Katke, Kate Nguyen, and Benyamin Ahmadnia. 2023. Sign Language Recognition and Translation: A Multi-Modal Approach Using Computer Vision and Natural Language Processing. In Proceedings of the 14th International Conference on Recent Advances in Natural Language Processing, pages 658–665, Varna, Bulgaria. INCOMA Ltd., Shoumen, Bulgaria.

[2]. Vardan Gupta, Saumya Sinha, Puneet Bhushan, and Minal Shettigar. "English Text to Indian Sign Language Translator." published in 2021.

[3]. Anjali Kanvinde, Abhishek Revadekar, Mahesh Tamse, and Dhananjay R. K. "Bidirectional Sign Language Translation." Presented at the IEEE International Conference on Communication, Information, and Computing Technology (ICCICT), June 25-27, 2021, Mumbai, India.

[4]. Alaftekin, M., Pacal, I. & Cicek, K. Real-time sign language recognition based on YOLO algorithm. Neural Comput & Applic 36, 7609–7624 (2024). https://doi.org/10.1007/s00521-024-09503-6 Melek Alaftekin, Ishak Pacal, and Kenan Cicek. "Real-Time Sign Language Recognition Based on YOLO Algorithm." published in 2024.

[5]. Bunny Saini, Divya Venkatesh, Nikita Chaudhari, Tanaya Shelake, Shilpa Gite, and Biswajeet Pradhan. "A Comparative Analysis of Indian Sign Language Recognition Using Deep Learning Models." Published in 2024.[https://www.researchgate.net/publication/372950169_A_comparative_analysis_of_Indian_sign_language_recognition_using_deep_learning_models]

[6]. Ozge Mercanoglu Sincan, Necati Cihan Camgoz, and Richard Bowden. "Using an LLM to Turn Sign Spottings into Spoken Language Sentences." Published in 2024.[[2403.10434] Using an LLM to Turn Sign Spottings into Spoken Language Sentences]

[7]. Sen Fang, Lei Wang, Ce Zheng, Yapeng Tian, and Chen Chen. "SignLLM: Sign Languages Production Using Large Language Models." Published in 2024.[[2405.10718] SignLLM: Sign Language Production Large Language Models]