Enhancing Accessibility: Leveraging Large Language Models for Indian Sign Language Translation

Amrita Soney, Esha Garg, Ishita Garg, Prof M.P.S Bhatia
Computer Science and Engineering Netaji Subhas University of Technology Delhi, India

***Abstract*—This paper explores the use of machine learning algorithms and Large Language Models (LLMs) to enhance the translation of Indian Sign Language (ISL) into spoken languages. A hybrid model is proposed for gesture identification from video inputs, and an LLM is used to translate identified signs into text, ensuring grammatical correctness. The method addresses the lack of annotated ISL datasets and promotes inclusive AI applications that bridge barriers between sign language communities and technology. The results indicate the viability of developing AI systems that respect linguistic diversity while putting underrepresented languages on the map.**

***Index Terms*—large language model, indian sign language, deaf, hard of hearing, linguistic diversity**

I. INTRODUCTION

Sign languages are a vital means of communication for millions of people around the world, especially deaf people. Sign languages are a complex multimodal language system that uses hand shapes, facial expressions, bodily movement, and spatial arrangements to convey messages. Of the hundreds of sign languages used around the world, the Indian Sign Language (ISL) is a valuable cultural and linguistic asset for millions of deaf and hard of hearing individuals in India. However, even though sign languages are widely used, they have been underrepresented in computational linguistics, particularly in the creation of sophisticated models such as large language models (LLM) [1].

Perhaps the most remarkable application of LLMs is the outstanding capabilities they offer in natural language processing. They have dramatically changed the way we interact with languages based on text and speech due to their impressive ability to generate and understand texts [2]. These models face a unique challenge in the case of sign languages: the multimodal nature with temporal dynamics and a lack of standardized written forms [3]. An important move towards making these powerful models accessible for everyone will be the incorporation of sign languages into LLMs.

This study bridges the gap between sign languages and LLMs, particularly Indian Sign Language (ISL). With its cultural and regional richness in variations across the Indian sub- continent, it introduces an extra layer of complexity. This study will help unlock new channels to incorporate sign languages in AI-driven solutions by having LLM process, comprehend, and generate ISL. Meanwhile, this paper addresses the demand for language models that are sensitive to the diversity of sign languages and the distinctive cultural context.

A. Sign Language Recognition

Sign Language Recognition (SLR) is the process of automatically identifying and interpreting gestures, hand signs, facial expressions, and body movements used in sign languages, using computational methods. The goal is to convert these non-verbal communication signals into spoken or written language so that communication between deaf/hard-of-hearing individuals and others becomes more accessible.

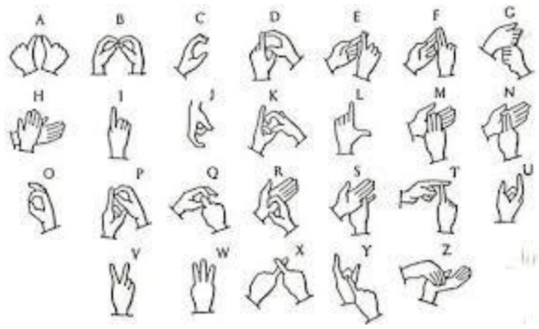

Fig. 1. ISL Alphabets [4]

B. Features of Indian sign language

- ISL is expressed through hand gestures, facial expressions, and body movements.
- It uses spatial positioning and motion to convey grammar and meaning.
- The typical sentence structure follows a

Subject-Object- Verb (SOV) order.

- Facial expressions play a key role in indicating emotions, questions, and negation.
- A single sign can represent multiple spoken words, de- pending on the context.
- There are regional variations in ISL in different parts of India.
- It combines manual gestures with non-manual signals like head tilts and eyebrow movements.
- ISL is still developing, with ongoing efforts to standardize its vocabulary and grammar.

Fig. 2. ISL Words

C. Motivation

The motivation for this research stems from the following reasons:

Social Inclusion and Equality: Of a population of over 7 billion, there are approximately 466 million deaf or hard-of- hearing people in the world, which represents around 6% of the population. Sign language is an important communication tool among the deaf community. Therefore, developing LLMs that are capable of comprehending, translating, and generating sign languages can help promote inclusion and equality for deaf people, fostering a sense of belonging.

Improved Accessibility: LLMs can revolutionize accessibility for the deaf community by enabling various communication tools. Real-time captioning on sign language videos can help viewers who may miss subtle signs or have difficulty understanding the language. There can be sign language interpreters for the deaf that accept sign languages and translate to different sign languages. The use of technologies for Virtual sign language interpreters would enable them to reach remote areas that have limited services in terms of Interpretation and, therefore, make it possible for interpretation to happen beyond the normal working hours. We can leverage these technologies to improve accessibility.

Advances in Natural Language Processing: Creating LLMs with the ability to comprehend, interpret and generate sign languages can help drive progress in NLP. Through their study of the visual, spatial, and temporal dimensions of the sign language, LLMs can acquire an understanding of how language may be conveyed by more than one modality. This can translate to other areas of NLP.

II. LITERATURE REVIEW

This section reviews existing Indian Sign Language (ISL) datasets, sign language recognition models, and English-to- Hindi translation approaches to establish a foundation for multimodal sign-to-text systems.

The paper ”Enhanced Sign Language Translation between American Sign Language (ASL) and Indian Sign Language (ISL) Using LLMs” by Malay Kumar, S. Sarvajit Visagan, Tanish Sarang Mahajan, and Anisha Natarajan introduces an innovative framework for closing the communication gap between ASL and ISL speakers. The work suggests a real-time ASL-to-ISL translation pipeline using Large Language Models (LLMs), where deep learning, NLP, and gesture recognition are integrated for better accuracy. A Random Forest Classifier-CNN hybrid ensures robust gesture recognition, and LLMs enable context-aware translation. RIFE-Net enhances real-time gesture synthesis. Although robust, the study is debatable. The system is restricted to fixed gestures, affecting dynamic sign conversation. Signer motion variability, complexity of the background, and illumination reduce recognition performance. The limited dataset is further restrictive of generalization. Future work is focused on dataset expansion, real-time optimization, and multimodal fusion for more effective translation [5].

The research paper ”Continuous Sign Language Recognition System using Deep Learning with MediaPipe Holistic” by Sharvani Srivastava, Sudhakar Singh, Pooja, and Shiv Prakash is a system that interprets sign language into speech to enable communication for hearing-impaired individuals. The study, however, has a limitation. Interference caused by background noise, signer movement variability, and hand and face overlap lowers the recognition rate. The system is not able to support real-time dynamic gestures either, and this restricts its application in practical natural conversation. The size of the dataset used is relatively small, and this affects the generalization ability of the model. The authors propose enlarging the size of the dataset, including more gestures, and further optimizing the model for improved real-time efficiency [6].

S. Fang et al. (2024) introduce SignLLM, a pioneering method that relies on large language models (LLM)for sign language generation. SignLLM features a multimodal architecture, in which text embeddings are paired with pose estimation modules. A major aspect of SignLLM is the addition of Reinforcement Learning with Efficiency Feedback (RLEF) to balance model size and inference speed without degrading accuracy. The authors demonstrate that SignLLM is superior to state-of-the-art methods on benchmark data. Challenges remain despite the improvements. Data sparsity remains a main challenge since annotated sign language data are sparse. Future studies should aim to develop data augmentation, transfer learning, and unsupervised learning techniques to counter this challenge [7].

J. Gong, L. G. Foo, Y. He, H. Rahmani, and J. Liu, in their research paper entitled ”LLMs are Good Sign Language Translators”, introduced a new method for Sign Language Translation (SLT) based on Large Language Models (LLMs). The paper points out the limitations of standard SLT models, notably their dependency on gloss annotations and non-alignment with natural language representation. To overcome this, the authors present SignLLM, a framework that transforms sign language videos into structured token representations, allowing it to be compatible with LLMs. The results show that LLMs can significantly enhance SLT accuracy, revealing new avenues for multimodal natural language processing in sign languages. The work contributes to sign language processing by showing that LLM-based SLT is feasible without relying on intermediate gloss annotations. Subsequent work can look at applying this method to more sign languages and enhancing real-time translation performance [8].

The article entitled ”An Efficient Sign Language Translation Using Spatial Configuration and Motion Dynamics with LLMs” by E. J. Hwang, S. Cho, J. Lee, and J. C. Park presented SpaMo, a new framework for Gloss-free Sign Language Translation (SLT) utilizing spatial configurations and motion dynamics native to sign language. As opposed to typical approaches that are typically based on domain-specific fine-tuning of visual encoders, SpaMo leverages pre-trained visual encoders to derive spatial and motion features, which are fed into a Large Language Model (LLM) with a language prompt. The framework also included visual-text alignment as an initial step prior to SLT supervision. Experimental results on the PHOENIX14T and How2Sign datasets show that SpaMo attains state-of-the-art performance in gloss-free SLT [9].

P. Zhang et al. introduced novel methods for sign language recognition and translation with event cameras. Generally, conventional RGB video methods are prone to failure because of issues like motion blur caused by rapid hand movement and illumination changes. Event cameras that record asynchronous brightness changes offer a solution by successfully perceiving dynamic hand movements. Comprehensive experiments can be conducted on both simulation (Phoenix14T) and EV-Sign datasets [10].

Mishra et al. described SignSpeak, an Indian Sign Language (ISL) recognition system based on machine learning for sim- plified communication. It identifies ISL letters and numbers with 99.98%

accuracy with Random Forest classifiers and ISL gestures as sentences with 87% accuracy with LSTM networks and LLMs. It simulates the process of acquiring datasets from Kaggle and internal sources, image preprocessing, and gesture recognition with MediaPipe. It has high classification accu- racy, improved user experience through interactive learning modules, and validated real-world usability through feedback from the Jhaveri Thanawala School for the Deaf. Mishra et al. suggest larger datasets, enhancing real-time recognition, and model optimization for better performance at scale. The research mentions the novelty of SignSpeak in uniting sentence formation and gesture identification, introducing inclusivity to the hearing-impaired community [11].

III. EXISTING DATASETS AND MODELS

A. Datasets for Indian Sign Language Recognition

Indian Sign Language (ISL) research has undergone remarkable growth. The following mentioned datasets, also tabulated in ?? meet different requirements of ISL recognition and translation to serve as input for researchers for designing strong machine learning models for isolated gestures, continuous sentences, and multimodal communication systems. Some popular ISL datasets are given below.

FDMSE-ISL Dataset [12]: FDMSE-ISL dataset is an enormous isolated Indian Sign Language dataset with 40,033 videos of 2,002 most common words being used in daily interactions in the Indian deaf society. It is comprised of roughly 36.2 hours and has 7.8 million frames. The data was collected using several cameras and comprises numerous categories, including behavior norms, household objects, and family relationships. The dataset makes FDMSE-ISL quite valuable to study in the realm of ISL recognition research.

Kaggle ISL Dataset by Soumya Kushwaha [13]: This Kaggle-hosted dataset contains Indian Sign Language samples focusing on alphabets (except H, J, Y) and numbers. It provides standardized image-based data that can be used for gesture recognition tasks. The dataset is suitable for training machine learning models for isolated ISL recognition.

IEEE DataPort ISL Dataset [14]: This dataset contains skeletal-point data for each alphabet in Indian Sign Language (excluding ’R’), represented as NumPy arrays with (x, y, z) coordinates for left and right hand movements. It includes 120 sequences per alphabet and supports gesture-based action recognition tasks using MediaPipe landmarks.

ISL-CSLTR Dataset [15]: The ISL-CSLTR dataset is designed for continuous sign language translation and recognition. It includes 700 fully annotated videos, 18,863 sentence-level frames, and 1,036 word-level images for 100 spoken-language sentences performed by seven signers. The dataset focuses on sentence-level ISL translation and is publicly available to support research on deep learning-based SLTR frameworks.

TABLE I: INDIAN SIGN LANGUAGE DATASETS

Dataset Name	Modality	Scope	Size
FDMSE-ISL [12]	Video	Word	40,033 videos
Custom ISL Dataset [13]	Image	Alphabet	Varies
IEEE DataPort ISL [14]	Skeletal Point Data	Alphabet	90,000 .npy files
ISL-CSLTR [15]	Video	Sentence/Word	700 videos
INCLUDE [16]	Video	Word	5,484 videos
ISL Dataset [17]	Video	Word/Phrase	Varies

B. Models for Indian Sign Language Recognition

The following papers highlight key developments in Indian Sign Language (ISL) recognition and translation, focusing on different techniques for gesture classification. These works address challenges such as background variability, intra-class differences. Comparisons of these papers are also provided in II. Below are their summaries:

Vashisth et al. proposed a deep learning-based system for recognizing Indian Sign Language (ISL) gestures using convolutional neural networks (CNNs). It achieved high accuracy by preprocessing gesture data and optimizing the model architecture. The system focused on isolated signs, addressing challenges such as intra-class variability and background noise [18].

Kothadiya et al. introduced Deepsign, which is a framework that combines CNNs and LSTMs for real-time sign language detection. It leveraged pose estimation for feature extraction, achieving around 96% accuracy on ISL datasets [19].

Dhiman et al. explore the classification of ISL gestures using CNNs of diverse backgrounds. The model achieves a testing accuracy of 99.10%, 92.69%, and 95.95% on the Indian sign language dataset with simple backgrounds, complex backgrounds, and in mixed background scenarios, respectively. The study highlights the adaptability to real- world environmental variations [20].

Likhar et al. compared CNN with segmented videos as input to them, transfer learning from American Sign Language to ISL for classification, LSTM model and U-Net with ResNet to give different accuracy results [21].

TABLE II: COMPARISON OF METHODS FOR ISL RECOGNITION

Study	Method	Dataset	Accuracy (%)	Limitations
2023 [18]	CNN	Self made static dataset for A-Z	99	Limited to alphabets
2022 [19]	LSTM+GRU	Self made dynamic dataset	97	Smaller dataset
2021 [20]	Hybrid CNN- RNN	Cas-Talk-ISL dataset, 50 ISL words	96	model's performance de- clines in complex back-ground scenarios
2020 [21]	CNN + resnet + unet	Own dataset with static and dynamic input	98.81	Limited Vocabulary
2019 [22]	Fuzzy C means clustering	Self made dataset	75	Limited to very few vocabulary

C. Models for English-to-Hindi Translation

Furthermore, for English-to-Hindi translation, several datasets have been developed catering to tasks such as machine translation, linguistic alignment, and contextual understanding. These datasets play a crucial role in training and evaluating models for accurate and relevant translations between English and Hindi; they are mentioned in III and are mentioned below.

The Helsinki-NLP/opus-mt-en-hi model is a pre-trained machine translation (MT) model designed for English-to- Hindi translation. It is part of the OPUS-MT project, which provides open-source neural machine translation models based on the Marian NMT framework [23].

Wordvice AI is a general LLM-based translation and language processing tool designed for high-quality English- to-Hindi translation. It utilizes large-scale language models to generate accurate, fluent, and contextually appropriate translations.

OpenHathi [24] is an AI model tailored for Hindi language tasks, outperforming models like GPT-3.5 in translation tasks involving Hindi. It is particularly adept at handling Hindi inputs.

Airavata is an instruction-tuned Hindi-specific language model built upon OpenHathi and further fine-tuned using diverse instruction datasets. It is designed to improve Hindi natural language understanding and generation, supporting Hindi, English, and Indian English [25].

Developed by AI4Bharat, IndicTrans2 [26] is a transformer- based multilingual Neural Machine Translation (NMT) model supporting high-quality translations across all 22 scheduled Indic languages, including Hindi. It employs script unification to enhance lexical sharing between languages.

TABLE III: MACHINE TRANSLATION MODELS FOR ENGLISH-TO-HINDI

Model/Resource Name	Type	Key Features/Description
Helsinki-NLP/opus- mt-en-hi [23]	Pre-trained MT Model	A machine translation model trained for English-to-Hindi translation.
Google Translate	General LLM Tool	Widely used online translation service with support for

		English to Hindi.
Wordvice AI	General LLM Tool	Utilizes comprehensive large language models for accurate and natural English- to-Hindi translations.
OpenHathi [24]	Hindi-Specific LLM	Open-source foundational model for Hindi language processing.
Airavata [25]	Hindi-Specific LLM	Instruction-tuned model for Hindi, built upon fine-tuning OpenHathi with di-verse instruction datasets. Supports Hindi, English, and Indian English.
IndicTrans2 [26]	Multilingual LLM	Open-source transformer-based multi- lingual NMT model supporting high- quality translations.

IV. METHODOLOGY

Indian Sign Language (ISL) recognition has been done through a series of machine learning and deep learning approaches. Various efforts have been made in the past to recognize ISL gestures with hand movement, both static and dynamic, with high accuracy. Yet, there continues to be a lack of translation of recognized words into Hindi to facilitate wider accessibility and communication. This system suggested a Long short-term memory (LSTM) based model for ISL recognition with a Large Language Model (LLM) for English to Hindi language conversion to enhance communication efficiency.

A. Dataset Selection and Preprocessing

To find the best dataset for training our recognition model, we have investigated a number of Indian Sign Language (ISL) datasets, as indicated in Table ??.

After conducting an examination of several datasets about Indian Sign Language (ISL), the ISL-CSLTR dataset has been selected as the primary dataset. Specifically tailored for research endeavors in ISL regarding Continuous Sign Language Recognition (CSLR) and Sign Language Translation (SLT), this dataset encompasses both word-level and sentence-level video samples, thereby rendering it as one of the exhaustive resources for ISL research [27]. To cater to a degree of diversity and generalizability in the training of the model, it comprises recordings of multiple signers executing gestures across varying backgrounds.

Raw video sequences are preprocessed to improve recognition performance and gesture clarity before classification. Every video is broken up into frame sequences, followed by frame resizing. In order to isolate significant sign features, Gesture segmentation is done by background removal, and hand keypoint detection algorithms, like MediaPipe Hand Tracking [28], are used to track the movements of the fingers and palm. Prior to being fed into the classification pipeline, the extracted frames are resized and normalised.

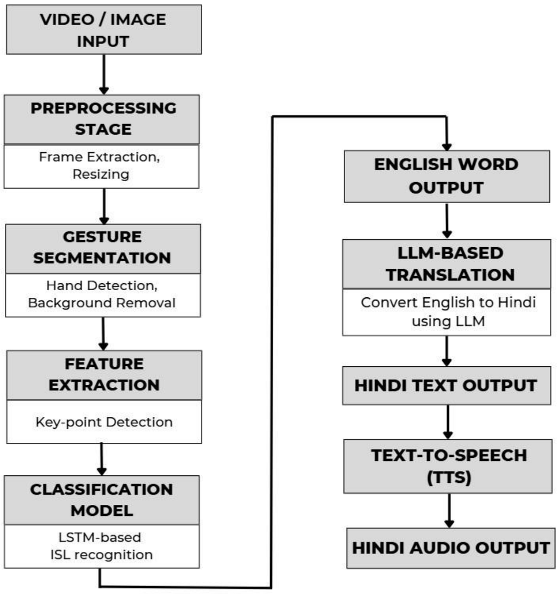

Fig. 3. ISL Recognition and Translation Framework

B. ISL Gesture Recognition Model

This framework uses a Long Short-Term Memory (LSTM) network for ISL gesture classification, which has shown to perform better than conventional CNN-based methods for sequential classification tasks in sign language recognition as seen in Table II. Since LSTMs are effective at learning motion patterns and temporal dependencies in video sequences, they are suited for time-series data.

During the classification process ,in the first step, the feature representations of hand gestures are obtained by passing the preprocessed frames through a feature extractor. Then these features are fed into an LSTM model that classifies the input gesture into one of the predefined English classes by analyzing patterns it observes across frames. Categorical cross-entropy loss is used while training the model and Adam optimizer is applied for better convergence. To evaluate model performance and identify the misclassification patterns, evaluation metrics such confusion matrices, accuracy, and F1-score are used.

C. Machine Translation Models for English-to-Hindi

We suggest using the Helsinki-NLP/opus-mt-en-hi model, a pre-trained neural machine translation model developed as part of the OPUS-MT project, for English-to-Hindi translation in the proposed system. This model is built on the Marian NMT framework and employs a transformer-based architecture that utilizes a sequence-to-sequence learning paradigm and self-attention mechanisms [29].

To adapt the Helsinki-NLP model for sign language interpretation, we propose fine-tuning it on a domain-specific dataset. The IIT Bombay English-Hindi Parallel Corpus is put forward for this purpose, as it provides linguistic data tailored to the language's terminology and grammatical structures. Fine-tuning allows the model to align with the requirements of ISL interpretation, significantly improving translation accuracy. This process ensures that the model captures linguistic subtleties relevant to sign language while maintaining high-quality translations. By fine-tuning the Helsinki-NLP model, we enhance its ability to handle the complexities of ISL-translated English text and generate accurate Hindi output.

D. Text to Speech output

The translated Hindi text, the final output of the system, is displayed on the screen to the user. The system also integrates Google Text-to-Speech (gTTS) to convert the translated Hindi text into speech. This auditory output is particularly beneficial for individuals who rely on spoken communication or have visual impairments, ensuring that the system caters to a broader range of users.

The gTTS module leverages Google's Text-to-Speech API to generate natural-sounding speech from text. By tokenizing and processing the input text, gTTS ensures that the syn- thesized speech is fluid and easy to understand. The low operational latency of the gTTS module ensures real-time spoken translation of ISL gestures into Hindi. This capability makes the system effective for practical applications, enabling seamless and immediate communication in dynamic scenarios. By combining visual, textual, and auditory outputs, the system exemplifies multimodal communication, enhancing its usability and impact across different user groups.

V. EXPERIMENTAL SETUP

For efficient training of the Indian Sign Language (ISL) recognition model and fine-tuning the Helsinki-NLP English- to-Hindi translation model, a high-performance hardware setup is required. The system is powered by an NVIDIA A100 (80GB HBM2e) GPU, providing high memory bandwidth and optimized tensor operations for deep learning tasks. It is supported by an AMD EPYC 7742 (64C/128T, 2.25 GHz) CPU, ensuring fast data processing and model execution. To handle large datasets and prevent memory bottlenecks, 128GB DDR4 RAM is utilized, along with a 2TB NVMe SSD (Samsung 990 PRO) for high-speed storage and checkpoint saving.

The software environment is built on Python 3.9, with deep learning models implemented using PyTorch 2.1 and TensorFlow 2.15, both optimized with CUDA 12.1 and cuDNN for GPU acceleration. The Hugging Face Transformers (v4.35) library is used for loading and fine-tuning the Helsinki-NLP translation model, while OpenCV 4.7 is employed for image preprocessing in ISL recognition. NLTK and SacreBLEU are used for evaluating translation performance, and GTTS (Google Text-to-Speech) is integrated for generating spoken Hindi output. This setup ensures efficient execution of ISL recognition, translation, and speech synthesis in a unified pipeline.

VI. EVALUATION METRICS

This section covers the metrics used to evaluate ISL recognition accuracy, translation quality, and speech synthesis performance.

A. Gesture Recognition Metrics

For multi-class classification using LSTM, we evaluate the following metrics:

a) Accuracy:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

Where:

- TP = True Positives
- TN = True Negatives
- FP = False Positives
- FN = False Negatives

b) Precision (Positive Predictive Value, PPV):

$$\text{Precision} = \frac{TP}{TP + FP} \quad (2)$$

Measures how many predicted gestures were actually correct.

c) Recall (Sensitivity, True Positive Rate, TPR):

$$\text{Recall} = \frac{TP}{TP + FN} \quad (3)$$

Measures how many actual gestures were correctly classified.

d) F1-score (Harmonic Mean of Precision and Recall):

$$F1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (4)$$

e) Confusion Matrix: A table showing the values of TP , FP , FN , and TN for different classes is used to analyze the performance of the classification.

B. Machine Translation Metrics (English-to-Hindi LLM)

To evaluate the quality of English-to-Hindi translation, we use the following metrics:

a) BLEU (Bilingual Evaluation Understudy):

$$\text{BLEU} = \exp \left(\sum_{n=1}^{N} w_n \log p_n \right) \times \min \left(1, e^{1-\frac{r}{c}} \right) \quad (5)$$

Where:

- p_n is the precision for n-grams
- r/c is the brevity penalty

b) ROUGE (Recall-Oriented Understudy for Gisting Evaluation):

- ROUGE-N (n-gram overlap):

$$\text{ROUGE-N} = \frac{\sum \text{matched n-grams}}{\sum \text{total n-grams in reference}} \quad (6)$$

- ROUGE-L (Longest Common Subsequence - LCS):

$$\text{ROUGE-L} = \frac{\text{LCS length}}{\text{reference length}} \quad (7)$$

VII. CONCLUSION

This study presents an end-to-end framework for recognizing and translating Indian Sign Language (ISL), combining deep learning driven sign classification with transformer-based language translation. The system utilizes an LSTM model for gesture recognition and a fine-tuned Helsinki-NLP translation model. By integrating text-to-speech synthesis through gTTS, the system offers spoken Hindi output, greatly improving real-time interaction for ISL users. This method improves the accessibility of ISL recognition systems, while also closing essential communication gaps for the hearing-impaired community in India.

Expanding on previous studies using CNNs, RNNs, and SVM with MediaPipe for gesture recognition, this research is progressing to implement an LSTM model alongside LLM for translation. Employing transformer-based models ensures greater precision in understanding the linguistic subtleties between ISL and Hindi.

Through the use of sophisticated machine learning methods and utilizing datasets that reflect regional diversity, this re- search lays the foundation for more inclusive systems designed to address India's linguistic and cultural diversity. Moreover, the system's multilingual features, producing outputs in Hindi and the possibility of adding other regional languages, high- light its scalability and social significance.

VIII. FUTURE WORK

Most of the existing methods of Indian sign language (ISL) have been developed under restricted conditions, such as limited movement ranges, uniform lighting, controlled background, and a limited number of signatories. These limitations reduce the generalization of systems to real world scenarios. In addition, most of the research is focused on static gestures, especially including digits (0–9) and alphabets, which neglects dynamic gestures at the level of word and sentence.

Future progress in recognition of ISL requires significant research on isolated, continuous, and more natural events.

This includes the use of time and spatial information in various lighting and environmental conditions. To increase the accuracy and reliability of gesture recognition systems, it is necessary to recognize and classify time data. Advanced methods, for example, supervised, unattended, semi insufficient, and transmission learning methods, as well as the structure of neural networks, need to be explored so that machines may detect a broader variety of gestures.

Another challenge lies in the regional specificity of ISL. Signs and gestures often differ between states and even districts, making it difficult to create a model that satisfies all potential users at the same time. This variability requires the development of adaptable systems capable of processing a regional dialect in ISL. In addition, the unavailability of extensive, standardized, and regional datasets is a significant narrow profile for research progress. To resolve this, there is a need for a comprehensive dataset that combines regional variations into a unified resource for wider applicability.

In order to increase more inclusive ISL recognition systems, future research should also focus on multilingual translation capabilities. For example, outputs could be generated in different regional languages apart from Hindi. Future research can build on this base by exploring larger datasets, optimizing the time of inference for real-time deployment, and incorporation of AI marginal solutions for low source environments. Ethical issues, including the safeguarding of individual data and community participation in the evaluation of the system, will be crucial to secure widespread acceptance and trust among users.

A solution to the aforementioned difficulties, future efforts can pave the way for strong and inclusive ISL recognition systems capable of bridging communication gaps across various linguistic and cultural environments.

REFERENCES

[1] T. B. Brown et al., ”Language Models are Few-Shot Learners,” Advances in Neural Information Processing Systems (NeurIPS), vol. 33, pp. 1877–1901, 2020.

[2] C. Raffel et al., ”Exploring the Limits of Transfer Learning with a Uni- fied Text-to-Text Transformer,” Journal of Machine Learning Research, vol. 21, no. 140, pp. 1–67, 2020.

[3] S. Koller et al., ”Sign Language Recognition, Generation, and Transla- tion: An Interdisciplinary Perspective,” arXiv preprint arXiv:1908.08597, 2019

[4] signer.ai, ”The 5 parameters of Indian Sign Language (ISL),” signer.ai, 2025. [Online]. Available: https://signer.ai/the-5-parameters-of-indian- sign-language-isl/.

[5] M. Kumar, S. S. Visagan, T. S. Mahajan, and A. Natarajan, ”En- hanced Sign Language Translation between American Sign Language (ASL) and Indian Sign Language (ISL) Using LLMs,” arXiv preprint arXiv:2411.12685, 2024.

[6] S. Srivastava, S. Singh, Pooja, and S. Prakash, ”Continuous Sign Language Recognition System using Deep Learning with MediaPipe Holistic,” arXiv preprint arXiv:2411.04517, 2024.

[7] S. Fang, L. Wang, C. Zheng, Y. Tian, and C. Chen, ”SignLLM: Sign Language Production Large Language Models,” arXiv preprint arXiv:2405.10718, 2024.

[8] J. Gong, L. G. Foo, Y. He, H. Rahmani, and J. Liu, ”LLMs are Good Sign Language Translators,” arXiv preprint arXiv:2404.00925, 2024.

[9] E. J. Hwang, S. Cho, J. Lee, and J. C. Park, ”An Efficient Sign Language Translation Using Spatial Configuration and Motion Dynamics with LLMs,” arXiv preprint arXiv:2408.10593, 2024.

[10] P. Zhang, H. Yin, Z. Wang, W. Chen, S. Li, D. Wang, H. Lu, and X. Jia, ”EvSign: Sign Language Recognition and Translation with Streaming Events,” arXiv preprint arXiv:2407.12593, 2024.

[11] Mishra, A., Verma, R., Singh, P., Sharma, K., ”SignSpeak: An Indian Sign Language Recognition System Using Machine Learning,” Proc. IEEE Int. Conf. on Artificial Intelligence and Machine Learning (AIML), 2024, pp. 1-6.

[12] FDMSE ISL Dataset, Papers with Code, 2023. [Online]. Available: https://paperswithcode.com/dataset/fdmse-isl.

[13] S. Kushwaha, ”Indian Sign Language Dataset,” Kaggle, 2023. [Online]. Available: https://www.kaggle.com/datasets/soumyakushwaha/indian- sign-language-dataset.

[14] J. Johnson, J. Joseph, M. Reji, and M. George, ”Indian Sign Language Skeletal-point NumPy

array using MediaPipe," IEEE DataPort, Apr. 2023. [Online]. Available: https://ieee-dataport.org/documents/indian- sign-language-skeletal-point-numpy-array-using-mediapipe.

[15] DrBlack00, "ISL CSLTR: Indian Sign Language Dataset," Kaggle, 2023. [Online]. Available: https://www.kaggle.com/datasets/drblack00/isl- csltr-indian-sign-language-dataset.

[16] S. Joshi, S. S. Bedi, and A. R. Joshi, "Indian Sign Language (ISL) Dataset," Zenodo, 2020. [Online]. Available: https://zenodo.org/records/4010759.

[17] H. M. Wajid and M. Z. Islam, "Indian Sign Language (ISL) Dataset," Mendeley Data, vol. 1, 2021. [Online]. Available: https://data.mendeley.com/datasets/yx7kdssfjp/1.

[18] H. K. Vashisth, T. Tarafder, R. Aziz, M. Arora, and Alpana, "Hand Gesture Recognition in Indian Sign Language Using Deep Learning," Engineering Proceedings, vol. 59, no. 1, p. 96, 2023. [Online]. Available: https://doi.org/10.3390/engproc2023059096.

[19] D. Kothadiya, C. Bhatt, K. Sapariya, K. Patel, A.-B. Gil-Gonza´lez, and J. M. Corchado, "Deepsign: Sign Language Detection and Recognition Using Deep Learning," Electronics, vol. 11, no. 11, p. 1780, 2022. [Online]. Available: https://doi.org/10.3390/electronics11111780.

[20] R. Dhiman, G. Joshi, and C. Rama Krishna, "A deep learning approach for Indian sign language gestures clas-sification with different back- grounds," J. Phys. Conf. Ser., vol. 1950, no. 1, 2021.

[21] P. Likhar, N. K. Bhagat and R. G N, "Deep Learning Methods for Indian Sign Language Recognition," 2020 IEEE 10th International Conference on Consumer Electronics (ICCE-Berlin), Berlin, Germany, 2020, pp. 1- 6, doi: 10.1109/ICCE-Berlin50680.2020.9352194.

[22] Gomathi, V., and Mariappan H. Muthu. "Real-time recognition of Indian sign language." Second International Conference on Computational Intelligence in Data Science. 2019.

[23] J. Tiedemann and S. Thottingal, "OPUS-MT—Building open trans- lation services for the world," in Proceedings of the 22nd Annual Conference of the European Association for Machine Translation (EAMT), Lisbon, Portugal, 2020, pp. 479–480. [Online]. Available: https://aclanthology.org/2020.eamt-1.61

[24] Sarvam AI,"OpenHathi-7B-Hi-v0.1-Base," Hugging Face, 2024. [Online]. Available: https://huggingface.co/sarvamai/OpenHathi-7B-Hi- v0.1-Base

[25] J. Gala, T. Jayakumar, J. Aktar Husain, A. Kumar M, M. S. U. R. Khan, Kanojia, R. Puduppully, M. M. Khapra, R. Dabre, R. Murthy, and Kunchukuttan, "Airavata: Introducing Hindi Instruction-tuned LLM", AI4Bharat, 2024. [Online].

[26] AI4Bharat, IndicTrans2: Models and code for high-quality NMT in Indic languages, GitHub repository, 2023. [Online]. Available: https://github.com/AI4Bharat/IndicTrans2.

[27] P. Jain, R. Chandra, A. Sharma, and A. Sharma, "ISL-CSLTR: Indian Sign Language Dataset for Continuous Sign Language Recognition and Translation," arXiv preprint, arXiv:2105.05394, 2021.

[28] Y. Zhang, Z. Wu, and X. Yang, "MediaPipe Hands: Real-time Hand Tracking and Gesture Recognition," Google Research, 2020.

[29] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need," Advances in Neural Information Processing Systems (NeurIPS), vol. 30, 2017. [Online].