International Journal for Multidisciplinary Research (IJFMR)
E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

Ai Tool/App for Indian Sign Language(Isl) Generator from Audio-Visual Content in English/Hindi to Isl Content and Vice-Versa

D Dinesh Kumar[1], Revanth Sai[2], and Darwin[3], Nikhil [4], Rohit Pratap Singh[5]

[1,2,3,4]Student, Department of C&IT, Reva University [2]Student, Department of C&IT, Reva University
[5]Associate Professor, Department of C&IT, Reva University

Abstract
Indian Sign Language (ISL) is the primary visual communication medium for approximately 63 million deaf and hard-of-hearing individuals in India, yet computational resources for ISL processing remain severely limited compared to spoken languages. This paper presents a comprehensive framework for developing an AI-powered mobile application capable of bidirectional translation between English/Hindi audio-visual content and Indian Sign Language. The system integrates three critical components: large-scale dataset aggregation from sources including iSign benchmark (118,000 video-sentence pairs) and ISLTranslate (31,222 sentence pairs), advanced natural language processing pipelines for grammatical transformation, and computer vision techniques for sign generation and recognition. We discuss the architectural design of such systems, the processing pipeline from audio input through NLP tokenization, lemmatization, and part-of-speech tagging to final ISL video output, and present evaluation methodologies using Dynamic Time Warping and machine translation metrics. The paper incorporates findings from recent research on ISL datasets, linguistic properties unique to sign languages, and practical implementation considerations for real-world deployment on mobile platforms.

Keyword: Indian Sign Language, bidirectional translation, natural language processing, computer vision, mobile application, accessibility technology, sign language datasets, neural machine translation.

I. INTRODUCTION

Communication barriers faced by deaf and hard-of-hearing individuals remain a major accessibility challenge in India. According to World Health Organization estimates, India has over 63 million people with hearing impairment, yet only around 300 certified Indian Sign Language (ISL) interpreters. This imbalance severely limits access to education, healthcare, employment, and public services. Although ISL is one of the most widely used sign languages globally, it remains significantly under-resourced in computational and language technology support compared to spoken languages such as English and Hindi. Technology-driven solutions offer a promising approach to bridging this communication gap. A bidirectional translation system capable of converting English or Hindi speech and text into ISL video, and vice versa, can substantially enhance accessibility across multiple domains. Such systems can support inclusive education, enable effective communication in healthcare settings, improve access to public info-

rmation, and facilitate real-time communication through digital platforms.

Developing such a system involves several technical challenges. ISL datasets have historically been limited, with most earlier resources focused on isolated word recognition rather than continuous sentence-level translation. Recent datasets such as **iSign** and **ISLTranslate** have significantly improved data availability by providing large-scale sentence-aligned video resources. Additionally, ISL differs structurally from spoken languages, relying on spatial grammar, non-manual markers, and simultaneous articulation, which necessitates advanced linguistic transformation techniques rather than direct word-level translation.

This paper presents a comprehensive framework for a bidirectional ISL translation system by synthesizing research on ISL linguistics, dataset development, natural language processing, and computer vision. The subsequent sections discuss ISL linguistic characteristics, available datasets, system architecture for speech-to-ISL and ISL-to-text translation, evaluation methodologies, mobile deployment considerations, and future research directions.

II. LINGUISTIC PROPERTIES OF INDIAN SIGN LANGUAGE AND COMPUTATIONAL IMPLICATIONS

Indian Sign Language (ISL) is a natural visual–spatial language whose structure differs fundamentally from spoken languages such as English and Hindi. Meaning in ISL is conveyed through a combination of **manual components** (hand shapes, movements, orientations, and locations) and **non-manual components** (facial expressions, head movements, eye gaze, and body posture). These elements operate simultaneously, enabling parallel information transfer rather than linear sequencing.

A key linguistic feature of ISL is its use of **three-dimensional signing space**. Entities are assigned spatial locations around the signer, and subsequent references are made through directional movements or pointing. This spatial indexing provides discourse cohesion without explicit pronouns, a mechanism with no direct equivalent in spoken-language syntax. Computational models must therefore represent not only hand gestures but also precise spatial trajectories.

Non-manual markers are essential grammatical elements in ISL. Facial expressions and head movements indicate questions, negation, emphasis, and conditionals. These features must be accurately captured during video processing; systems relying solely on skeletal pose estimation risk losing critical grammatical information unless facial landmarks are incorporated.

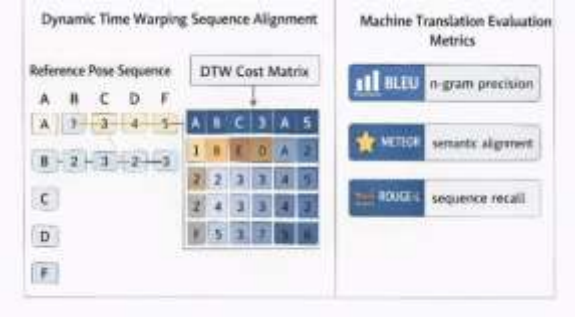
Fingerspelling is used to represent proper nouns and words lacking standardized signs. It consists of static hand shapes corresponding to alphabet letters and requires separate recognition mechanisms distinct from continuous sign movement detection.

ISL grammar follows a **Topic–Comment structure** rather than the Subject–Verb–Object order common in English. Many English function words, such as articles and auxiliary verbs, are omitted in ISL, with grammatical meaning conveyed through spatial and non-manual cues. As a result, direct word-by-word translation is linguistically incorrect and necessitates ISL-specific grammatical transformation.

Additionally, **mouth patterns** contribute semantic and prosodic information in ISL and are not equivalent to spoken-language lip movements. Capturing these patterns adds further complexity to computational processing.

These linguistic characteristics impose several requirements on ISL translation systems: datasets must include full upper-body and facial video capture; pose-based models should be augmented with facial

features; text-only intermediate representations may lead to information loss; and conventional text-based evaluation metrics are insufficient for assessing sign language translation quality.



III. MAJOR ISL DATASETS AND THEIR CHARACTERISTICS

The availability of large-scale, well-annotated datasets is a critical requirement for advancing machine learning-based Indian Sign Language (ISL) processing systems. Early ISL research primarily relied on small datasets focused on isolated word or alphabet recognition, while more recent efforts have shifted toward continuous sentence-level translation, which is more suitable for real-world communication tasks.

A. Early Isolated Word Datasets

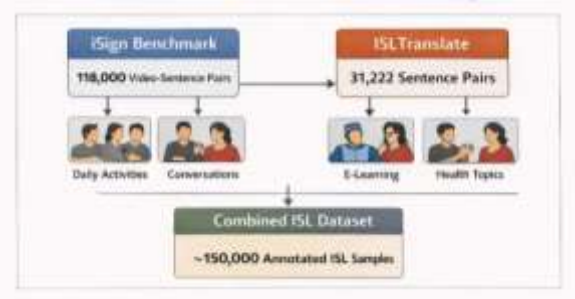
Initial ISL datasets developed in the early 2010s concentrated on static alphabets and numerals, typically containing a few hundred to a few thousand samples with limited signer diversity and controlled backgrounds. Subsequent datasets introduced dynamic gestures and expanded vocabularies, incorporating multiple signers and more realistic recording conditions. These datasets enabled high recognition accuracy for isolated signs using traditional machine learning and deep learning models. However, their focus on isolated gestures limited their usefulness for continuous sentence translation, where sign boundaries, co-articulation, and non-manual markers play a significant role.

B. ISL-CSLRT: Early Continuous Translation Dataset

The ISL-CSLRT dataset represented an important step toward continuous ISL translation by providing sentence-level video–text pairs. It consists of 700 videos covering approximately 100 sentences recorded by seven signers, with content largely drawn from educational and narrative contexts. While this dataset enabled early sequence-based translation research, its small size, limited signer diversity, and narrow domain coverage restricted its effectiveness for training large-scale neural translation models.

C. ISLTranslate Dataset

The ISLTranslate dataset significantly expanded available ISL translation resources by providing 31,222 ISL–English sentence and phrase pairs. The dataset was created using publicly available educational videos from the Indian Sign Language Research and Training Centre and the Deaf Enabled Foundation. Automated speech-to-text transcription, followed by manual verification, was used to obtain aligned English text. A notable design choice was the absence of intermediate gloss annotations, favoring end-to-end video-to-text translation for scalability. Validation by certified ISL experts demonstrated translation quality comparable to human benchmarks, despite some unavoidable noise from narration and signing mismatches.



D. iSign Benchmark

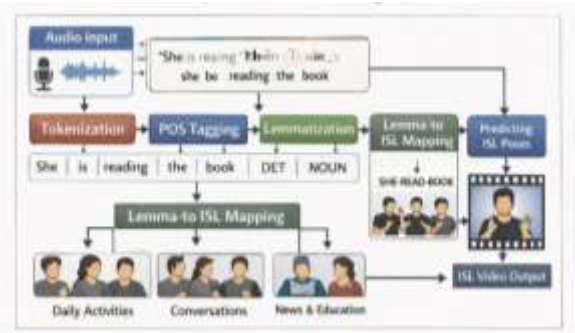
The iSign benchmark is the largest and most comprehensive ISL dataset to date, comprising 118,228 ISL–English video–text pairs aggregated from multiple sources, including educational content and ISL news broadcasts. Beyond dataset size, iSign introduces a standardized multi-task evaluation framework covering ISL-to-English translation, English-to-ISL pose generation, one-shot word recognition, word presence prediction, and semantic similarity assessment. Validation results indicate higher reliability compared to earlier datasets, reflecting improved data diversity and annotation quality. Although multiple reference translations would further strengthen evaluation, practical limitations in expert availability remain a challenge.

IV. SYSTEM ARCHITECTURE FOR SPEECH-TO-ISL TRANSLATION

An effective speech-to-Indian Sign Language (ISL) translation system requires the integration of multiple specialized components within a structured processing pipeline. The system converts English or Hindi speech and text into ISL video output through sequential stages, with optional feedback between stages to improve accuracy and fluency.

A. Audio Input and Speech Recognition

The translation process begins with audio input obtained from live speech, recorded files, or extracted audio from video content. Automatic Speech Recognition (ASR) converts this audio into textual form for further processing. While modern ASR systems achieve high accuracy for clear speech, transcription errors may occur due to accents, unclear audio, or specialized terminology. For real-time applications, such errors must be handled through confidence scoring or robust downstream processing rather than manual correction.



B. Natural Language Processing Pipeline

The transcribed text undergoes standard natural language processing steps including tokenization, part-of-speech tagging, lemmatization, and dependency parsing. These steps normalize word forms, identify grammatical roles, and capture sentence structure. This linguistic analysis prepares the input for

transformation into ISL-compatible representations and reduces vocabulary sparsity by mapping inflected words to their base forms.

C. ISL-Specific Grammar Transformation

Following general NLP processing, ISL-specific grammatical rules are applied. This includes removal of function words that are typically absent in ISL, reordering sentence structure to follow Topic–Comment patterns, resolving pronouns into explicit spatial references, and encoding negation or modality using non-manual markers. These transformations are essential, as direct word-by-word translation from English to ISL does not preserve grammatical or semantic correctness.

D. Lemma-to-Sign Mapping

The transformed text is mapped to ISL signs using a structured lexicon that links word lemmas to corresponding sign videos or pose sequences. When a direct match is unavailable, fallback strategies such as fingerspelling, semantic substitution, or morphological decomposition are used to ensure complete coverage of input content.

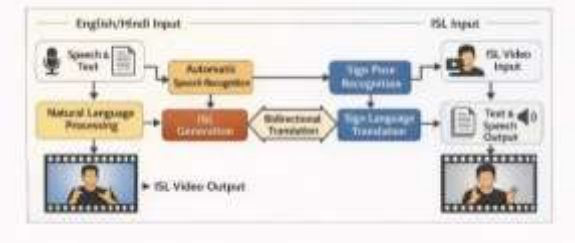
E. ISL Video Output Generation

The final stage generates visual ISL output. This can be achieved either by retrieving pre-recorded sign videos and assembling them into sequences or by generating pose-based sign representations that are rendered as animated signing avatars. While video retrieval is computationally efficient, pose-based generation offers greater flexibility and smoother transitions between signs, albeit at higher computational cost.

V. ISL-TO-ENGLISH TRANSLATION: THE REVERSE DIRECTION

A complete bidirectional translation system must also support translation from Indian Sign Language (ISL) video to English or Hindi text. This reverse translation task is more challenging than speech-to-ISL conversion due to the unstructured and high-dimensional nature of video input.

The translation process begins with feature extraction from ISL video. Two primary approaches are used: **RGB-based** and **pose-based** methods. RGB-based approaches process raw video frames and can capture detailed visual information, including non-manual cues, but are computationally expensive and sensitive to signer appearance. Pose-based approaches extract skeletal keypoints representing body and hand movements, offering improved efficiency and generalization across signers, though facial and non-manual information may be partially lost unless explicitly modeled.



Extracted features are processed using **temporal sequence models** such as recurrent neural networks or transformer architectures to learn the dynamics of continuous signing. The resulting representation is decoded into text using encoder–decoder frameworks similar to those employed in spoken-language machine translation.

ISL-to-text translation presents several challenges, including the absence of explicit sign boundaries,

variability in signing styles, the influence of non-manual markers across multiple signs, and the limited availability of annotated training data. These factors complicate accurate segmentation and translation.
System performance is commonly evaluated using standard machine translation metrics such as BLEU, METEOR, ROUGE-L, and Word Error Rate. However, these metrics have limitations when applied to visual-spatial languages, as they may not fully capture semantic accuracy, spatial meaning, or non-manual grammatical features present in ISL.

VI. MOBILE APPLICATION IMPLEMENTATION CONSIDERATIONS

Deploying Indian Sign Language (ISL) translation systems on mobile devices offers improved accessibility and portability but introduces constraints related to limited processing power, memory, battery capacity, and network variability. Mobile platforms must therefore balance performance, efficiency, and usability.
Automatic speech recognition can be implemented either on-device or using cloud-based services. Cloud-based approaches generally provide higher accuracy but depend on stable internet connectivity and introduce latency, whereas on-device models enable offline operation and preserve user privacy at the cost of reduced model complexity. Similar trade-offs apply to natural language processing components, where lightweight and optimized models are preferred for real-time mobile execution.
ISL video output on mobile platforms is typically achieved through retrieval of pre-recorded sign videos, which is computationally efficient and feasible even on low-end devices. In contrast, real-time generation of sign animations requires substantial resources and is better suited for server-side processing or simplified pose-based interpolation methods.
User interface design plays a critical role in mobile ISL applications. Interfaces should prioritize clear and sufficiently large sign video display, support flexible input methods such as text or voice, and offer controls for playback speed and repetition to enhance comprehension. Additionally, applications must account for varying network conditions by supporting offline functionality for commonly used signs and efficient data caching.
Energy efficiency is an important consideration due to the computational demands of video processing and model inference. Practical deployment requires optimized models, adaptive quality settings, and selective offloading of computation to cloud services when feasible.

VII. EVALUATION METRICS AND PERFORMANCE ASSESSMENT

Evaluating Indian Sign Language (ISL) translation systems presents unique challenges due to the visual–spatial nature of sign languages. Conventional text-based machine translation metrics are limited in their ability to assess whether translated output preserves semantic meaning, spatial structure, and non-manual grammatical features.
ISL-to-English translation is commonly evaluated using standard machine translation metrics such as BLEU, METEOR, ROUGE-L, and Word Error Rate. While these metrics provide a quantitative measure of textual similarity, low scores observed on ISL benchmarks highlight both the difficulty of the task and the inadequacy of text-only evaluation for sign languages.
For English-to-ISL translation, pose-based generation quality is often assessed using Dynamic Time Warping, which measures similarity between generated and reference pose sequences while accounting for variations in signing speed. However, interpretation of these scores remains challenging, as higher distances may reflect natural signing variation rather than poor translation quality.
Recent research emphasizes the need for sign-language-specific evaluation approaches that account for

spatial accuracy, non-manual markers, and overall perceptual similarity. Human evaluation by fluent ISL users remains the most reliable assessment method, although its high cost and limited scalability restrict its widespread use.

VIII. LIMITATIONS AND FUTURE RESEARCH DIRECTIONS

Despite recent progress, Indian Sign Language (ISL) translation systems continue to face several limitations that restrict their practical deployment. Although datasets such as iSign and ISLTranslate have expanded available resources, their size and domain diversity remain limited compared to spoken-language datasets. Most existing data focuses on formal and educational content, with insufficient coverage of conversational language, regional variation, and specialized domains.

Current systems also inadequately model **non-manual markers**, including facial expressions and head movements, which are essential to ISL grammar. Additionally, many datasets and models rely on a small number of signers, limiting generalization across different signing styles and regional variations of ISL. Fingerspelling, despite its frequent use in real-world communication, remains underrepresented in training data and system evaluation.

From a technical perspective, end-to-end video-to-text approaches face challenges in segmenting continuous signing and incorporating non-manual information. Achieving real-time performance for interactive applications further requires improvements in feature extraction efficiency and model optimization.

Future research should prioritize the development of larger and more diverse datasets, improved modeling of non-manual linguistic features, and evaluation methods grounded in sign-language-specific properties. User-centered studies involving deaf and hard-of-hearing individuals are essential to assess system usability, comprehensibility, and real-world effectiveness. Ethical considerations, including privacy protection and fairness across diverse user groups, should be integral to future system design and deployment.

IX. CONCLUSION

Bidirectional translation between English or Hindi and Indian Sign Language (ISL) has the potential to significantly enhance accessibility and inclusion for deaf and hard-of-hearing individuals in India. Recent advances in dataset development and system architectures integrating speech recognition, natural language processing, and computer vision provide a strong foundation for practical ISL translation systems.

Despite this progress, current systems still face limitations in accuracy, linguistic coverage, and real-time performance. Future work must focus on incorporating ISL-specific linguistic features, improving evaluation methodologies, expanding dataset diversity, and optimizing solutions for mobile deployment. Active involvement of the deaf community in system design and evaluation remains essential to ensure real-world relevance and usability.

Well-designed ISL translation systems cannot eliminate all communication barriers, but they can meaningfully improve access to education, healthcare, employment, and public services. Continued research and responsible deployment can help transform these systems from experimental prototypes into practical assistive technologies.

ACKNOWLEDGMENTS

The authors acknowledge contributions from researchers involved in ISL dataset development and sign

language processing, as well as organizations supporting Indian Sign Language research and community engagement. Special recognition is given to the deaf and hard-of-hearing community whose experiences motivate and guide this work.

REFERENCES

1. A. Joshi, R. Mohanty, M. Kanakanti, A. Mangla, S. Choudhary, M. Barbate, and A. Modi, "iSign: A benchmark for Indian Sign Language processing," *arXiv preprint arXiv:2407.05404*, 2024.
2. A. Joshi, S. Agrawal, and A. Modi, "ISLTranslate: Dataset for translating Indian Sign Language," in *Findings of the Association for Computational Linguistics (ACL)*, 2023.
3. P. Sharma, D. Tulsian, C. Verma, P. Sharma, and N. Nancy, "Translating speech to Indian Sign Language using natural language processing," *Future Internet*, vol. 14, no. 9, p. 253, 2022.
4. R. Damdoo and P. Kumar, "An integrative survey on Indian sign language recognition and translation," *IET Image Processing*, vol. 19, p. e70000, 2025.
5. N. C. Camgoz, O. Koller, S. Hadfield, and R. Bowden, "Sign language transformers: Joint end-to-end sign language recognition and translation," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020.
6. B. Saunders, N. C. Camgoz, and R. Bowden, "Progressive transformers for end-to-end sign language production," in *European Conference on Computer Vision (ECCV)*, 2020.
7. P. Selvaraj, G. Nc, P. Kumar, and M. Khapra, "OpenHands: Making sign language recognition accessible with pose-based pretrained models across languages," in *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (ACL)*, 2022.
8. N. C. Camgoz, S. Hadfield, O. Koller, H. Ney, and R. Bowden, "Neural sign language translation," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018.
9. D. Li, C. Rodriguez, Y. Yu, and D. Li, "Word-level deep sign language recognition from video: A new large-scale dataset and methods comparison," in *IEEE Winter Conference on Applications of Computer Vision (WACV)*, 2020.
10. A. Duarte, S. Palaskar, L. Ventura, D. Ghadiyaram, K. DeHaan, F. Metze, J. Torres, and X. Giro-i Nieto, "How2Sign: A large-scale multimodal dataset for continuous American Sign Language," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2021.
11. H. Zhou, W. Zhou, W. Qi, J. Pu, and L. Li, "Improving sign language translation with monolingual data by sign back-translation," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2021.
12. S. Ko, C. Kim, H. Jung, and C. Cho, "Neural sign language translation based on human keypoint estimation," *arXiv preprint arXiv:1811.11436*, 2018.
13. K. Yin, A. Moryossef, J. Hochgesang, Y. Goldberg, and M. Alikhani, "Including signed languages in natural language processing," in *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics (ACL)*, 2021.
14. S. Albanie, G. Varol, L. Momeni, H. Bull, T. Afouras, H. Chowdhury, N. Fox, B. Woll, R. Cooper, A. McParland, and A. Zisserman, "BOBSL: BBC-Oxford British Sign Language Dataset," 2021.
15. R. Elakkiya and B. Natarajan, "ISL-CSLTR: Indian sign language dataset for continuous sign language translation and recognition," *Mendeley Data*, 2021.
16. A. Joshi, A. Bhat, P. S, P. Gole, S. Gupta, S. Agarwal, and A. Modi, "CISLR: Corpus for Indian Sign Language recognition," in *Proceedings of the 2022 Conference on Empirical Methods in Natural*

Language Processing (EMNLP), 2022.
17. D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.
18. M. Müller, "Dynamic time warping," in *Information Retrieval for Music and Motion*, pp. 69–84, 2007.
19. K. Papineni, S. Roukos, T. Ward, and W. Zhu, "BLEU: A method for automatic evaluation of machine translation," in *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL)*, 2002.
20. C. Lin, "ROUGE: A package for automatic evaluation of summaries," in *Text Summarization Branches Out: Proceedings of the ACL-04 Workshop*, 2004.
21. World Health Organization, "Deafness and hearing loss," 2016.
22. Ethnologue, "The Indian Sign Language," 2022.