

Artificial Intelligence-Based Indian Sign Language to Speech Conversion Using Computer Vision and Deep Learning: A Review

Meshwa Bhatt¹, Nirali Dave², and Hemangini Patel³

¹Research Scholar, Department of Computer Science, Vanita Vishram Women's University, Surat, Gujrat, India

^{2,3} Assistant Professor, Department of Computer Science, Vanita Vishram Women's University, Surat, Gujrat, India

Correspondence should be addressed to Meshwa Bhatt; bhattmeshwa324@gmail.com

Received: 19 Sep 2026; **1st Revision:** 3 Oct 2026; **2nd Revision:** 9 Oct 2026; **Accepted:** 15 Oct 2026;

Published Online: 26 Oct 2026

Copyright © 2026 Meshwa Bhatt et al. This is an open-access article distributed under the [Creative Commons Attribution 4.0 International License \(CC BY 4.0\)](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

ABSTRACT- Indian Sign language provides an important means of communication for deaf and hard of hearing individuals in India. However, communication might be difficult when people who do not understand ISL have to interact with Sign language users. Artificial Intelligence, Computer Vision and deep learning offer possible ways for recognizing signs from visual inputs and converting recognized information into text and speech. This paper reviews existing research on ISL recognition, translation, datasets and sign to speech conversion. It also discusses some techniques such as Convolutional Neural Network (CNN), Long Short-Term Memory (LSTM) Networks, Gated Recurrent Units (GRUs), and transformer-based models. It also examines datasets, such as INCLUDE and ISL Translate, alongside international sign language datasets used in related research work. The reviewed literature shows the distinction between recognizing individual signs and translating continuous sign sequences into full sentences. Some important challenges include limited and diverse ISL data, changes in the variation among signers, changes in lighting and background, and the need for effective real time processing. Based on the reviewed studies, the paper shows a conceptual framework that combines sign recognition, language processing, and text to speech conversion. The framework provides a direction for further future research; its complete implementation and performance evaluation remain outside the scope of this review.

KEYWORDS- Artificial Intelligence, Computer Vision, Deep Learning, Indian Sign Language, Sign Language Recognition, Text-to-Speech

I. INTRODUCTION

Communication becomes difficult for people who doesn't understand sign language. Because communication can be difficult for the person who doesn't understand sign language.

But with the help of Computer Vision and Artificial Intelligence methods for understanding information which is captured through camera. A camera captures the information while deep learning can analyze the frames. This pattern can then be converted into text or speech.

Early research was only done on individual sign language but now researchers have developed many deep learning techniques such as convolution and neural network CNN's long short-term memory LSTM networks and gated recurrent units GRU to improve the recognition and temporal information. [1], [4], [5], [7]

For the dataset which has supported this research is INCLUDE. Where large scale Indian Sign Language data was available for Recognition [2]. While, other dataset was ISL Translate which focused on translation of the Indian Sign Language and many were International Sign Language which includes WLASL. And How to sign also provided the contribution for this research. [11], [12]

Most recent studies have been shown by the researchers towards transformer base and gloss free translation base learning methods [10], [14], [15]. This approach was applied to understand longer sequence of translation of signs rather than considering every single individual signs.

Below Figure 1 illustrates the overall process of AI-based sign language recognition and text conversion. The process begins with capturing sign language through a camera. The captured video is processed to identify relevant hand and body information. Deep learning and computer vision techniques are then used to recognize the signs, which are converted into text for further processing.

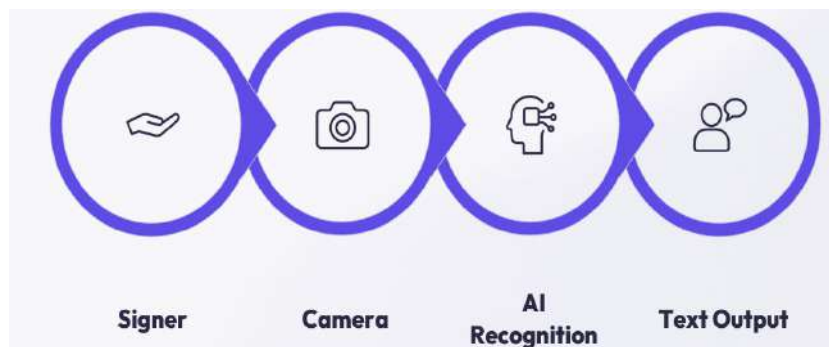


Figure 1: AI-Based Sign Language Recognition and Text Conversion Process

II. REVIEW METHODOLOGY

A. Research Question

The following questions guide the review: Indian Sign Language uses which Artificial Intelligence, Computer Vision, and deep learning techniques for detection and recognition? How are CNN, LSTM, GRU, and Transformer models used in sign language recognition and translation? Which datasets are available for Indian Sign Language and other sign languages? Which perspectives support continuous sign language translation and sign to speech conversion? What are the real-time ISL recognition and translation's current gaps and challenges?

B. Search Strategy

This review aims on the research that has been published and is related to Indian Sign Language recognition and translation, deep learning, computer vision, language models, datasets, sign to speech systems, and gloss-free translation. The studies which have been discussed in this manuscript covers the publication period from 2015 to 2024, based on cited literature. The search keywords include "Indian Sign Language", "ISL recognition", "sign language translation", "computer vision", "deep learning", "CNN", "LSTM", "GRU", "Transformer", "gloss-free translation", "language models", and "sign-to-speech".

C. Databases Used

The literature reviewed in this study was collected from IEEE Xplore, Springer, and Google Scholar. These sources were used to identify relevant research papers that are related to Indian Sign Language recognition, translation, datasets, deep learning, computer vision, continuous sign language translation, gloss-free translation, language models, and sign-to-speech systems.

D. Inclusion/Exclusion Criteria

The studies that were included addressed sign language recognition, translation, continuous sign language, datasets, deep learning, computer vision, transformer-based methods, gloss-free translation, language models, or related sign-to-speech approaches. The studies outside the scope of this review were not related to sign language recognition or translation.

E. Study Selection

This paper presents a narrative review of research on Indian Sign Language (ISL) recognition, translation and sign to speech conversion. The discussion has been done on 16 principal references published between 2015 and 2024,

selected for their relevance to recognition techniques, datasets, continuous sign language translation, multi model approaches, gloss-free translation, language models and speech generation. The selected literature was studied to identify the main approaches, available datasets, evaluation measures and research challenges. A complete numerical account of study identification and selection cannot be provided as the original search-result counts and screening records were not retained. For this reason, the manuscript is presented as a narrative review rather than a systematic review, and the 16 references should not be interpreted as an exhaustive collection of all relevant studies.

III. LITERATURE REVIEW

A. Early Indian Sign Language Recognition

Early research was found only on individual ISL Signs and gestures from videos or images. One study showed that the only focus is on continuous ISL sign Language [1]. This system showed that sign language should not only be considering individual signs but also taking sentences into consideration.

Another study showed signers independent recognition [3]. This was important because many signers do not only perform one signs; they move their hands from one motion to another so there can be difference in hand Movement provided by the signer. However, there are still many challenges which can be faced as background lighting and camera pose changes with the time.

B. Deep Learning for ISL Recognition

Deep learning has become an important method for sign language recognition CNN based methods can help features like hand move, finger position and hand orientation from images or videos.[5]

CNN and LSTM-based systems have also been used for ISL recognition [7]. In Such systems CNN takes useful information based on video frames, while the LSTM processes the movement of these features over time.

Other research has used MediaPipe with GRU models [4]. MediaPipe is helpful for hand landmarks, while the GRU can learn how these landmarks change during signing.

C. Datasets for Sign Language Recognition and Translation

Datasets are important for developing ISL sign recognition. INCLUDE is a large-scale dataset which is developed for Indian Sign Language Recognition [2]. While ISL Translate

focuses on translating Indian Sign Language into English continuous sentences.[8]

There are many other datasets such as WLASL [12] and how2sign [11] Are mainly based on American Sign Language Although they are not ISL datasets they provide powerful examples for developing Indian Sign Language recognition and translation system.

IV. DATASETS USED IN SIGN LANGUAGE RESEARCH

A. Include

INCLUDE is a large Datasets used for Indian Sign Language Recognition [2].

B. Isl Translate

ISLTranslate is designed for Indian Sign Language translation [8]. It is useful for research on translating Indian Sign Language into continuous English sentences.

C. Wlasl

WLASL is a large-scale American Sign Language dataset containing many word-level signs [12]

D. How2sign

How2Sign is a large multimodal dataset for continuous American Sign Language [11]. It is helpful for Studying longer sign sequence

E. Phoenix14t

PHOENIX14T is generally used in sign-language translation research [9], [10]. As it supported the development and evaluation of neural and Transformer-based translation methods.

V. CONTINUOUS SIGN LANGUAGE TRANSLATION AND TRANSFORMERS

Recognizing and translating is a two completely different things a recognition system may identify individual signs while a translation system needs to understand full sentence and a meaning of a complete sequence [9]

Camgoz et al. [9] introduced neural approaches for sign-language translation and later explored Transformer-based sign-language recognition and translation [10]. Study of different part of sequence is done with the help of transformer, making them useful for datasets are important for developing ISL sign recognition and longer sign language videos.

Sign 1 → Sign 2 → Sign 3 → Meaningful Sentence

This shows the development of sign-language research from simple gesture recognition toward continuous sign-language translation.

VI. MULTIMODAL AND GLOSS-FREE SIGN LANGUAGE TRANSLATION

The research has been started on more than just hand movements skeleton base and multi approaches can include information about hands bodies and movements [13].

Sign Video → Gloss → Sentence

Gloss-free systems aim to reduce this intermediate step:

Sign Video → Sentence

This approach still requires large and diverse training data; however, it can make the translation process more direct.

VII. AI AND LARGE LANGUAGE MODELS FOR SIGN LANGUAGE TRANSLATION

Research Recently have also showed the use of large language models for sign language translation Sign2GPT is an example of research using language models for gloss free sign language translation [16].

After the sign language has been recognized language models can help generate more natural sentences after the sign information. However, a model cannot directly understand the hand movements It first needs useful visual or sign representation from the recognition system

VIII. SIGN-TO-SPEECH CONVERSION

After the sign language has been recognized, the recognized sign information is then converted to a meaning sentence through language processing. Language models are helpful in generating more natural sentences from the recognized sign information. The final sentence is then converted to spoken output using a Text-To-Speech (TTS) system. This enables the system to provide the output which can be easily understood by the users who may not know sign language. The overall sign-to-speech process can thus be represented as following:

Sign Language → Sign Recognition → Language Processing → Sentence Generation → Text-to-Speech → Spoken Output

Because Text-to-Speech conversion provides the final spoken form of the translated sign language, it is an important component of sign-to-speech systems. However, the accuracy of sign recognition and the quality of the generated sentence is what determines the quality of the speech output.

IX. COMPARATIVE ANALYSIS OF EXISTING STUDIES

In below Table 1, it summarizes the principal studies reviewed in this paper, including their publication year, main topic, and main finding.

Table 1: Comparative Analysis of Existing Studies

Ref.	Year	Main Topic	Main Finding
[1]	2015	ISL Recognition	Identified continuous ISL gestures.
[2]	2020	ISL Dataset	Made a large ISL dataset.
[3]	2022	Signer Independent	Recognized signs from different users.
[4]	2022	MediaPipe + GRU	Landmarks recognized by ISL
[5]	2022	CNN + SVM	Indian Sign Language recognized
[6]	2022	Deep Learning	Detected and recognized signs.
[7]	2023	CNN + LSTM	Recognized signs from video formats.

[8]	2023	ISL Translation	Created a dataset for ISL translation.
[9]	2018	Neural Translation	Translated sign language using neural networks.
[10]	2020	Transformers	Recognition and Translation uses transformers
[11]	2021	Multimodal Dataset	A continuous sign language dataset created
[12]	2020	WLASL	Word-level sign recognition supported
[13]	2021	Skeleton Features	Recognition uses skeleton information
[14]	2023	Gloss-Free	Without gloss signs are translated
[15]	2023	Gloss-Free	Used end-to-end sign translation.
[16]	2024	Sign2GPT	Used LLMs for sign language translation.

X. PROPOSED AI-BASED ISL-TO-SPEECH ARCHITECTURE

Proposed architecture presented(see Figure 2) in this section is a conceptual framework developed from the findings of the literature reviewed in this paper. It has a combination of video capturing, preprocessing, hand and body detection, feature extraction, sequence modelling, sign recognition, language processing and text to speech conversion into a possible ISL to speech work flow. These components have a reflection of approaches which are discussed in the previous research on sign language recognition and translation. However, the complete architecture has not been implemented or experimentally validated in this study. It is shown to illustrate a possible system design and identify directions for future research rather than demonstrate measured system performance.



Figure 2: AI Based Indian Sign Language-to-Speech Architecture Proposed

Figure 2 illustrates the proposed architecture for converting Indian Sign Language into speech. It shows the complete processing pipeline, beginning with video capture and preprocessing, followed by hand and body detection, feature extraction, sequence processing, sign recognition, language processing, and finally text-to-speech conversion. In our proposed architecture, we followed these steps:

- **Step 1: Video Capture**

Webcam that records live video of the user's ISL signs

- **Step 2: Preprocessing**

The video is split up into frames, and basic preprocessing tasks like resizing and normalizing are carried out

- **Step 3 Recognition of Hands and Body**

A computer-vision method (e.g. MediaPipe) is used to detect hand and body landmark points. This helps to identify hand position, finger motion and body posture [4].

- **Step 4: Extraction of Features**

Important features like hand shape, finger position, movement and body position are extracted from the frames. Visual features can be extracted by CNN [7].

- **Step 5: Sequence Processing**

As sign language is a movement over time, models such as LSTM, GRU or Transformer can process the sequence of extracted features [4],[7] [10].

- **Step 6: Sign Recognition**

The trained model uses the processed video [2], [5], [7] to recognize the sign or sequence of signs.

- **Step 7: Language Processing**

The identified signs are then transformed into meaningful sentence using Natural Language Processing (NLP) [8],[9].

- **Step 8: Text To Speech Text-to-Speech Conversion**

The last sentence is converted to speech output by a Text-to-Speech (TTS) engine.

XI. PROPOSED SYSTEM FLOW

In below Figure 3 we represent the proposed System Flow and major stages involved in the Indian Sign Language recognition and speech conversion process. These steps demonstrate the sequential processing of video input to identify signs, produce meaningful text, and generate the final speech output. It explains the process of converting Indian Sign Language into Speech using AI, at first, a camera captures the sign with the help of a video, then it preprocesses using frame normalization, Next the system detects the Hand and Body Detection, the features are extracted using landmark and motion features, AI and deep learning techniques then recognize the signs and convert it into English text. Finally, Text-to-speech system converts the English text into audio, allowing people to understand the message.

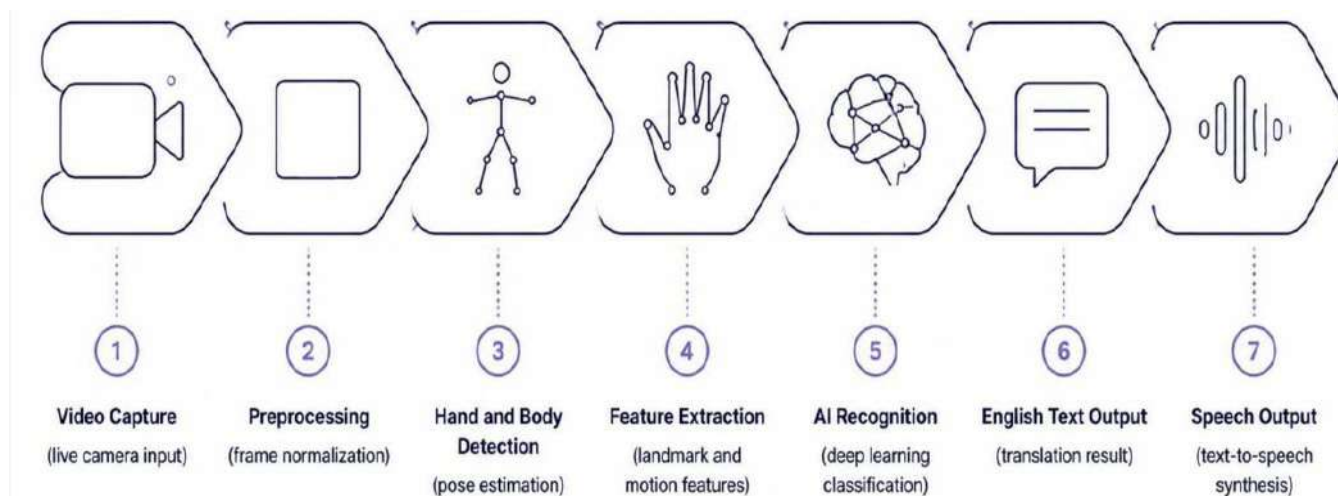


Figure 3: Stages of AI-Based Indian Sign Language Recognition and Speech Conversion.

XII. EVALUATION MEASURES USED IN PRIOR LITERATURE

The following evaluation metrics are included as measures used in prior literature to assess sign language recognition, translation, and processing performance. They are presented for review and comparison purposes only. No experimental results from the proposed AI-based ISL-to-speech architecture are reported or claimed in this review.

A. Accuracy

Accuracy measures the portion of correctly reorganized or classified Signs out of the total numbers of the predictions made by the system.

Accuracy = Number of Correct Predictions / Total numbers of predictions

B. Precision

Precision is a classification metrics that measures the proportion of positive predictions that are actually correct

Precision = TP / (TP + FP)

Where:

- TP = True Positive
- FP = False Positive

C. Recall

Recall measures how many of the actual signs belonging to a particular class are correctly identified by the system. It Shows the ability of the model to detect relevant signs.

$$\text{Recall} = \text{TP} / (\text{TP} + \text{FN})$$

Where:

- TP = True Positive
- FN = False Negative

D. F1-Score

F1-score is the performance metric use to measure precision and recall into a single number

$$\text{F1} = 2 \times (\text{Precision} \times \text{Recall}) / (\text{Precision} + \text{Recall})$$

A higher F1-Score indicates a better balance between Precision and Recall.

E. BLEU Score

BLEU (Bilingual Evaluation Understudy) can be used to evaluate the quality of translated sentences. It compares the sentence generated by a sign language translation system with a reference sentence. It is therefore useful for assessing the quality of the generated translation.

F. Word Error Rate

A Standard metric used to measure the accuracy of the Automated Speech recognition .WRE measures the minimum level of word level changes required to turn the machine transcript into the reference transcript

$$\text{WER} = (\text{S} + \text{D} + \text{I}) / \text{N}$$

Where:

- S = Substitutions
- D = Deletions
- I = Insertions
- N = Number of words in the correct sentence

A lower WER indicates fewer errors between the generated output and the reference sentence.

G. Processing Speed

Processing speed is important for real-time sign language recognition and sign-to-speech systems. Prior studies can evaluate processing performance using measures such as:

- **Frames Per Second (FPS):** Shows how many video frames it can capture per seconds
- **Processing Time:** Measures the time required by the system to process the input.
- **Response Time:** Measures the time taken by the system to produce an output after receiving an input.
- **Latency:** Represents the delay between the input sign and the corresponding system output.

These measures are useful for comparing the suitability of different approaches for real-time sign language recognition and translation applications.

In the below [Table 2](#), we present a comparative analysis of 16 Existing Studies on Sign Language and translation published between 2015 and 2024.It summarizes the methods used and the evaluation metric reported in the literature ,including Study methods, Accuracy, Precision, Recall F1-score,BLEU – 4,WER FPS and Latency. This Studies uses different techniques and dataset, and some do not report all evaluation metric .Therefore this table summarizes findings from existing literature does not present experimental results from the proposed architecture.

Table 2: Comparison Analysis of existing Sign language Recognition and translation Studies Based on Evaluation Metric

Ref.	Year	Study / Method	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)	BLEU-4	WER (%)	FPS	Latency (ms)
[1]	2015	Continuous ISL Gesture Recognition	94% accuracy	Not reported	Not reported	Not reported	Not reported	Not reported	Not reported	Not reported
[2]	2020	INCLUDE Dataset	94.5% (INCLUDE-50); 85.6% (INCLUDE); 92.1% (ASLLVD, 48 classes)	Not reported	Not reported	Not reported	Not reported	Not reported	Not reported	Not reported
[3]	2022	Signer-Independent ISL Recognition	90.1% (static); 89% (dynamic)	Not reported	Not reported	Not reported	Not reported	Not reported	Not reported	Not reported
[4]	2022	MediaPipe + GRU	95% (average)	Not reported	Not reported	Not reported	Not reported	Not reported	Not reported	Not reported
[5]	2022	SURF + SVM + CNN	99.17% (SVM); 99.64% (CNN)	99.09% (SVM); 99.57% (CNN)	99.02% (SVM); 99.57% (CNN)	99.096% (SVM); 99.57% (CNN)	Not reported	Not reported	Not reported	Not reported
[6]	2022	DeepSign	97%	97%	97%	97%	Not reported	Not reported	Not reported	Not reported
[7]	2023	CNN + LSTM for ISL Recognition	Not available	Not available	Not available	Not available	Not available	Not available	Not available	Not available

[8]	2023	ISLTranslate Dataset	Not reported	Not reported	Not reported	Not reported	6.09 (Pose-SLT baseline)	61.88 (281-pair human-validation sample)	Not reported	Not reported
[9]	2018	Neural Sign Language Translation	Not available	Not available	Not available	Not available	Not available	Not available	Not available	Not available
[10]	2020	Sign Language Transformers	Not reported	Not reported	Not reported	Not reported	20.17 (Sign-to-Text); 21.32 (best Sign2Gloss2Text translation model)	24.59% (Sign2Gloss); 26.16% (best translation model)	Not reported	Not reported
[11]	2021	How2Sign Dataset	Not available	Not available	Not available	Not available	Not available	Not available	Not available	Not available
[12]	2020	WLASL Dataset / Word-Level Recognition	62.63% top accuracy (2,000 classes)	Not reported	Not reported	Not reported	Not reported	Not reported	Not reported	Not reported
[13]	2021	Skeleton-Aware Multimodal Recognition	Not available	Not available	Not available	Not available	Not available	Not available	Not available	Not available
[14]	2023	Gloss-Free Sign Language Translation	Not available	Not available	Not available	Not available	Not available	Not available	Not available	Not available
[15]	2023	Gloss-Free End-to-End Translation	Not reported (translation study)	Not reported	Not reported	Not reported	7.06 (OpenASL test); 2.24 (How2Sign test)	Not reported	Not reported	Not reported

XIII. RESEARCH GAPS AND CHALLENGES

Problem of limited ISL dataset. More large and diverse Indian Sign Language dataset are required.

Different people may perform Similar signs differently which can affect the system.

Problem of background and lightings due to different type of lightings in background camera positions which can reduce the recognition performance.

Recognizing individual sign is easier than understanding of complete sequence of sign

Hand movements alone cannot only provide all the information which signer wants to convey so the body movement and facial expressions can also be important part of the communication.

The system needs to recognize sign quickly as the communication becomes easy and faster while maintaining good performance.

Recognized signs may be converted into meaningful sentences and understandable speech.

XIV. FUTURE SCOPE

- To develop a ISL datasets
- Including many users and many signing styles.
- To support different Types of lighting and backgrounds.
- Coordinating hand information boarding information and facial information.
- Improving real time processing.
- Using continuous translations based on transformer.
- Improving the gloss free translation.

- Better sentences formation using NLP and language models
- Developing direct ISL to speed systems

XV. CONCLUSION

The studies discussed in this paper cover different perspectives, including CNN based recognition, temporal modelling using LSTM and GRU Networks, Transformer-based translation, and Gloss-free methods. The review also shows datasets such as INCLUDE and ISL Translate which supports ISL recognition and translation. Appropriate models and evaluation methods are required for different tasks such as recognizing individual signs and translating continuous sign sequences as indicated by the literature. Challenges remain in dataset diversity environmental conditions, signer variations real time processing and continuous sentence translation. This paper outlines a conceptual ISL to speech framework that connects visual signs recognition with Language processing and speech generation in order to address needs. The effectiveness of complete framework cannot yet be established as it has not been implemented or experimentally evaluated in this study. Future research should be focusing more on the development and evaluation of integrated systems using diverse ISL data and suitable recognition, translation, and processing speed measures.

CONFLICTS OF INTEREST

The authors declare that they have no Conflicts of Interest.

REFERENCES

- [1] K. Tripathi and N. B. G. C. Nandi, "Continuous Indian sign language gesture recognition and sentence formation," *Procedia Computer Science*, vol. 54, pp. 523–531, 2015. Available: <https://doi.org/10.1016/j.procs.2015.06.060>
- [2] Sridhar, R. G. Ganesan, P. Kumar, and M. Khapra, "Include: A large-scale dataset for Indian sign language recognition," in *Proc. 28th ACM Int. Conf. Multimedia*, 2020, pp. 1366–1375. Available: <https://doi.org/10.1145/3394171.3413528>
- [3] P. K. Athira, C. J. Sruthi, and A. Lijiya, "A signer independent sign language recognition with co-articulation elimination from live videos: An Indian scenario," *Journal of King Saud University—Computer and Information Sciences*, vol. 34, no. 3, pp. 771–781, 2022. Available: <https://doi.org/10.1016/j.jksuci.2019.05.002>
- [4] B. Subramanian, B. Olimov, S. M. Naik, S. Kim, K. H. Park, and J. Kim, "An integrated MediaPipe-optimized GRU model for Indian sign language recognition," *Scientific Reports*, vol. 12, Art. no. 11964, 2022. Available: <https://doi.org/10.1038/s41598-022-15998-7>
- [5] S. Katoch, V. Singh, and U. S. Tiwary, "Indian sign language recognition system using SURF with SVM and CNN," *Array*, vol. 14, Art. no. 100141, 2022. Available: <https://doi.org/10.1016/j.array.2022.100141>
- [6] D. Kothadiya, C. Bhatt, K. Sapariya, K. Patel, A. B. Gil-González, and J. M. Corchado, "DeepSign: Sign language detection and recognition using deep learning," *Electronics*, vol. 11, no. 11, Art. no. 1780, 2022. Available: <https://doi.org/10.3390/electronics11111780>
- [7] S. Bhamare and S. Bhamare, "Translating the unspoken: Deep learning approaches to Indian sign language recognition using CNN and LSTM networks," in *Proc. Annu. Int. Conf. Emerging Research Areas: Int. Conf. Intelligent Systems (AICERA/ICIS)*, 2023, pp. 1–6. Available: <https://doi.org/10.1109/AICERA/ICIS59538.2023.10420379>
- [8] Joshi, S. Agrawal, and A. Modi, "ISLTranslate: Dataset for translating Indian sign language," in *Findings of the Association for Computational Linguistics: ACL 2023*, 2023, pp. 10466–10475. Available: <https://doi.org/10.18653/v1/2023.findings-acl.665>
- [9] N. C. Camgoz, S. Hadfield, O. Koller, H. Ney, and R. Bowden, "Neural sign language translation," in *Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition (CVPR)*, 2018, pp. 7784–7793. Available: <https://doi.org/10.1109/CVPR.2018.00812>
- [10] N. C. Camgoz, O. Koller, S. Hadfield, and R. Bowden, "Sign language transformers: Joint end-to-end sign language recognition and translation," in *Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition (CVPR)*, 2020, pp. 10023–10033. Available: <https://doi.org/10.1109/CVPR42600.2020.01004>
- [11] A. Duarte, S. Palaskar, L. Ventura, D. Ghadiyaram, K. DeHaan, F. Metze, et al., "How2Sign: A large-scale multimodal dataset for continuous American sign language," in *Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition (CVPR)*, 2021, pp. 2734–2743. Available: <https://doi.org/10.1109/CVPR46437.2021.00276>
- [12] D. Li, C. Rodriguez Opazo, X. Yu, and H. Li, "Word-level deep sign language recognition from video: A new large-scale dataset and methods comparison," in *Proc. IEEE/CVF Winter Conf. Applications of Computer Vision (WACV)*, 2020, pp. 1459–1469. Available: <https://doi.org/10.1109/WACV45572.2020.9093512>
- [13] S. Jiang, B. Sun, L. Wang, Y. Bai, K. Li, and Y. Fu, "Skeleton aware multi-modal sign language recognition," in *Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2021, pp. 3413–3423. Available: <https://doi.org/10.1109/CVPRW53098.2021.00380>
- [14] B. Zhou, Z. Chen, A. Clapés, J. Wan, Y. Liang, S. Escalera, et al., "Gloss-free sign language translation: Improving from visual-language pretraining," in *Proc. IEEE/CVF Int. Conf. Computer Vision (ICCV)*, 2023, pp. 20871–20881. Available: <https://doi.org/10.1109/ICCV51070.2023.01908>
- [15] K. Lin, X. Wang, L. Zhu, K. Sun, B. Zhang, and Y. Yang, "Gloss-free end-to-end sign language translation," in *Proc. 61st Annu. Meeting Association for Computational Linguistics (ACL)*, vol. 1: Long Papers, 2023, pp. 12904–12916. Available: <https://doi.org/10.18653/v1/2023.acl-long.722>
- [16] R. Wong, N. C. Camgoz, and R. Bowden, "Sign2GPT: Leveraging large language models for gloss-free sign language translation," in *Proc. Int. Conf. Learning Representations (ICLR)*, 2024. Available: <https://tinyurl.com/n8ynaz7s>

ABOUT THE AUTHORS



Meshwa Bhatt Meshwa Bhatt is a Research Scholar in Computer Science and Information Technology at Vanita Vishram Women's University, Surat, Gujarat, India. She is currently pursuing a Bachelor of Science in Information Technology (B.Sc. IT) Honours with Research at the same university.



Dr. Nirali Dave specializes in Image Processing and Artificial Intelligence (AI). She received her Ph.D. in Computer Science from Hemchandracharya North Gujarat University (HNGU), Patan, Gujarat, in 2018, her Master of Computer Applications (MCA) from Maharaja Krishnakumarsinhji Bhavnagar University (MKBU), Bhavnagar, in 2005, and her Bachelor of Computer Applications (BCA) from Gujarat University, Ahmedabad, in 2002. Her teaching subjects include Operating Systems, Data Structures, Computer Graphics, and Software Engineering. Her research interests include Information Security and Cyber Forensics. She has 18 years of academic experience and received a Best Paper Presentation Award for her work titled "A Novel Approach for Superpixel Watermarking Scheme for Tamper Detection and Image Authentication."



Dr. Hemangini Patel specializes in Computer Science and Information Technology. She received her Ph.D. in Computer Science from Veer Narmad South Gujarat University in 2018, her Master of Science in Information Technology (M.Sc. IT) from the same university in 2005, and her Bachelor of Science in Information Technology (B.Sc. IT) in 2003. Her research interests include Information Retrieval, Web Mining, Machine Learning, and Deep Learning. She has 18 years of academic experience and is recognized as a Ph.D./M.Phil. guide and a postgraduate (PG) teacher.