Bulletin of Electrical Engineering and Informatics
Vol. 15, No. 3, June 2026, pp. 2832~2841
ISSN: 2302-9285, DOI: 10.11591/eei.v15i3.11141

SignVerse: bridging communication through a bi-directional sign language translation system

Gopal Dadarao Upadhye[1], Shalini Wankhade[2], Umbare Rupali Tukaram[3], Ankita Kakade[1], Mayur Agarwal[1], Dhanshree Shinde[1], Manesh Mahale[1], Nujaim Maindargi[1]
[1]Department of Artificial Intelligence and Data Science, Vishwakarma Institute of Technology, Pune, India
[2]Department of Information Technology, Vishwakarma Institute of Technology, Pune, India
[3]Department of Information Technology, JSPM's Rajarshi Shahu College of Engineering, Pune, India

Article Info

Article history:

Received Aug 1, 2025
Revised Mar 10, 2026
Accepted Apr 1, 2026

Keywords:

Bi-directional translation
Deep learning
Human-computer interaction
Indian Sign Language
Sign language recognition

ABSTRACT

This study introduces SignVerse, a novel bi-directional sign language translation (SLT) system, to enhance communication between the hearing-impaired community and the general public. SignVerse makes real-time, two-way conversations easy for Indian Sign Language (ISL) users—no special hardware needed. The system uses smart artificial intelligence (AI) tech: computer vision, deep learning, and natural language processing (NLP). When someone types or speaks, the text/speech-to-sign module runs the input through NLP-based syntactic reordering and shows the ISL translation using a lively 3D avatar. On the flip side, the sign-to-text/speech module leverages MediaPipe to spot hand landmarks in real time, and the convolutional neural network-long short-term memory (CNN-LSTM) model accurately recognizes each gesture. Everything works together to help ISL users connect smoothly with others 94.8% recognition accuracy, less than 1.8-second translation latency, and more than 90% gesture clarity in user studies are all demonstrated by experimental evaluations. The lightweight model, which is optimized through knowledge distillation, guarantees excellent performance even on common consumer devices. With significant potential for societal impact, SignVerse is a significant step toward real-time, AI-driven ISL translation. When everything is taken into account, it is a dependable, scalable, and reasonably priced choice for inclusive communication.

Corresponding Author:

Gopal Dadarao Upadhye
Department of Artificial Intelligence and Data Science, Vishwakarma Institute of Technology
Upper Indiranagar, Bibwewadi, Pune, Maharashtra, 411037, India
Email: gopal.upadhye@vit.edu

1. INTRODUCTION

Communication is a fundamental human necessity that enables social interaction, education, healthcare access, and participation in everyday activities. However, individuals with hearing and speech impairments often face significant communication barriers, especially in environments where sign language is not widely understood by the general population [1], [2]. According to the World Health Organization (WHO), more than 430 million people worldwide experience disabling hearing loss, and a substantial portion of this population depends on sign language as a primary means of communication [1]. Despite advances in assistive communication technologies, interaction between sign language users and non-signers still largely depends on human interpreters or limited one-way translation systems, which restrict accessibility and real-time communication [2], [3].

Recent developments in artificial intelligence (AI), computer vision, and deep learning have significantly improved the capability of automated sign language recognition and translation systems [3], [4]. In particular, convolutional neural networks (CNNs), recurrent neural networks (RNNs), and long short-term memory (LSTM) architectures have shown strong performance in extracting spatial and temporal features from dynamic hand gestures [4], [5]. Transformer-based approaches and graph convolutional networks (GCNs) have further enhanced the modeling of continuous sign language sequences by learning complex spatial relationships and temporal dependencies [6], [7]. Additionally, lightweight real-time frameworks such as MediaPipe have enabled efficient hand landmark extraction using standard RGB cameras without requiring specialized hardware [8].

Although substantial progress has been made in sign language translation (SLT) research, the majority of existing systems primarily focus on American Sign Language (ASL), while Indian Sign Language (ISL) remains comparatively underrepresented in available datasets, translation frameworks, and deployment-ready applications [9]. Furthermore, many existing systems perform only one-directional translation, such as sign-to-text or text-to-sign conversion, rather than supporting complete bi-directional communication [2], [10]. Several approaches also rely on computationally intensive transformer architectures or sensor-based hardware systems, limiting their usability on resource-constrained consumer devices and real-world deployment scenarios [6], [11].

To address these limitations, this work proposes SignVerse, a real-time bi-directional ISL translation system designed to facilitate seamless communication between hearing-impaired individuals and the general public. The proposed framework integrates two major modules: i) a sign-to-text/speech module that employs MediaPipe-based hand landmark extraction along with a CNN-LSTM recognition pipeline for accurate gesture interpretation and ii) a text/speech-to-sign module that uses natural language processing (NLP) techniques for ISL grammar restructuring and generates sign animations through a 3D avatar [8], [12]. Unlike conventional systems that depend on sensor gloves or focus exclusively on ASL datasets, SignVerse operates using commonly available webcams and microphones, making the system economical, scalable, and accessible [9], [13]. Furthermore, knowledge distillation techniques are incorporated to develop a lightweight student model suitable for deployment on mobile and web platforms without compromising recognition accuracy [14].

The proposed system aims to provide an integrated, low-latency, and user-friendly platform capable of supporting real-time ISL communication in practical environments such as educational institutions, healthcare facilities, and public service centers. By combining gesture recognition, NLP-based linguistic processing, and avatar-driven sign generation into a unified framework, SignVerse contributes toward the development of inclusive and accessible communication technologies for the hearing-impaired community.

2. LITERATURE SURVEY

SLT systems have evolved considerably due to advancements in AI, computer vision, and deep learning technologies. Early research in this area primarily focused on rule-based and grammar-driven systems for translating textual information into sign language. Stoll *et al.* [11] proposed one of the earliest English-to-ISL translation systems using a rule-based linguistic framework. Similarly, Papastratis *et al.* [12] developed a semantic machine translation system for Arabic Sign Language (ArSL). Although these systems demonstrated the feasibility of automated sign generation, they suffered from limited scalability, rigid grammatical structures, and difficulty handling natural real-time communication.

The introduction of deep learning techniques significantly improved the performance of sign language recognition systems. CNNs became effective for extracting spatial features from gestures, while RNNs and LSTM models enabled temporal sequence learning for dynamic gestures. Madahana *et al.* [13] demonstrated the effectiveness of deep learning-based gesture recognition using CNN architectures for sign language interpretation.

Recent studies have increasingly focused on transformer-based and graph-based neural architectures for continuous SLT. Sharath *et al.* [14] proposed an adaptive transformer-based framework for continuous sign language recognition and translation, improving contextual sequence understanding. Strobel *et al.* [15] introduced a hybrid transformer model for efficient and accurate sign language gesture recognition. Arib *et al.* [16] developed a skeleton-aware neural network capable of modeling spatial relationships between hand landmarks and body joints. Similarly, Damdoo and Kumar [17] proposed the SignFormer-GCN framework, which combines GCNs with spatio-temporal modeling for continuous SLT.

Several survey studies have comprehensively analyzed the progress and limitations of current SLT systems. Krishnamurthi and Indiramma [18] reviewed various sign language machine translation approaches and highlighted challenges related to multilingual datasets, sequence alignment, and grammatical consistency. Rastgoo *et al.* [19] categorized existing methods into recognition, translation, and sign

production pipelines while discussing the increasing adoption of transformer-based approaches. WHO [20] emphasized the importance of accessibility-oriented AI systems for practical SLT applications.

Research has also explored sign language production and avatar-based rendering systems. Abdullah *et al.* [21] introduced Text2Sign, which combined neural machine translation and generative adversarial networks (GANs) for sign generation. Tan *et al.* [22] analyzed various AI technologies used in sign language applications, including avatar-driven systems. Damdoo and Kumar [23] proposed a real-time speech-to-text-to-sign translator aimed at improving communication accessibility for hearing-impaired individuals.

Despite substantial progress, several challenges remain unresolved in current SLT systems. Most existing approaches primarily focus on ASL, while ISL remains comparatively underrepresented in publicly available datasets and deployment-ready systems [24]. Additionally, many transformer-based architectures require high computational resources, making real-time deployment difficult on low-cost consumer devices [25]. Existing systems also tend to focus on only one direction of translation, such as sign-to-text or text-to-sign conversion, instead of enabling complete bi-directional communication.

To address these limitations, the proposed SignVerse framework introduces a lightweight real-time bi-directional ISL translation system integrating MediaPipe-based landmark extraction, CNN-LSTM gesture recognition, NLP-based grammar restructuring, and avatar-based sign rendering within a unified architecture. The proposed system is designed for deployment using standard webcams and microphones while maintaining high recognition accuracy and low translation latency for practical real-world communication scenarios.

Recent research has also explored bi-directional sign language communication systems integrating both sign-to-text and text-to-sign translation functionalities. Mali *et al.* [26] proposed a two-way sign language translator using CNN and artificial neural network (ANN)-based techniques for ISL communication. Their work demonstrated the practicality of integrating gesture recognition and translation modules within a unified framework for improving communication accessibility. However, the system primarily focused on recognition accuracy and did not extensively address lightweight deployment, real-time latency optimization, or avatar-based ISL rendering for interactive communication scenarios.

3. METHOD

The SignVerse Core System bridges the gap between ISL and text or speech, making real-time, two-way translation possible. The system runs across three main layers: Input, Core System, and Output.

Input layer:

- The system captures visual data from a camera running at 30 frames per second, perfect for recognizing gestures quickly and accurately.
- Takes audio from a microphone and lets you type in text through the keyboard or a user interface.

Core processing system:

- Module A handles sign-to-text translation. It uses MediaPipe to spot 21 key hand landmarks and captures pose details, then feeds these into a CNN-LSTM transformer model with knowledge distillation. This setup pulls off an impressive 94.8% accuracy.
- Module B works in the other direction—text or speech to sign. It relies on Google API or Sphinx to turn speech into text, runs NLP algorithms for ISL grammar and syntax reordering, and pulls matching ISL signs from a database housed in JavaScript object notation (JSON) or comma separated values (CSV) format.
- A bi-directional translation engine stitches these two modules together, making live communication seamless.

Output layer:

- On the output side, you'll see text, hear speech, and get animated gestures through a 3D avatar. The system also tracks performance, reporting accuracy and latency so you know how well it's working.

The architecture of SignVerse splits its system into two main parts.:

- Module A: it handles converting text or speech into sign language.
- Module B: the second does the reverse, translating sign language back into text or speech.

Each layer uses specialized AI techniques and software to get the job done with precision. This kind of setup doesn't just boost accuracy — it also keeps the process straightforward and fast, letting translations happen in real time.

The overall architecture of the proposed SignVerse system is illustrated in Figure 1. The framework is organized into three major layers: input layer, core processing layer, and output layer. The input layer captures visual, audio, and textual data using webcams, microphones, and user interfaces. The core processing layer contains two primary modules responsible for sign-to-text/speech translation and text/speech-to-sign conversion. These modules integrate MediaPipe-based hand landmark extraction, CNN-LSTM gesture recognition, NLP, and avatar-based sign generation techniques. Finally, the output layer

generates translated text, synthesized speech, and animated ISL gestures through a 3D avatar interface, enabling seamless real-time bi-directional communication.

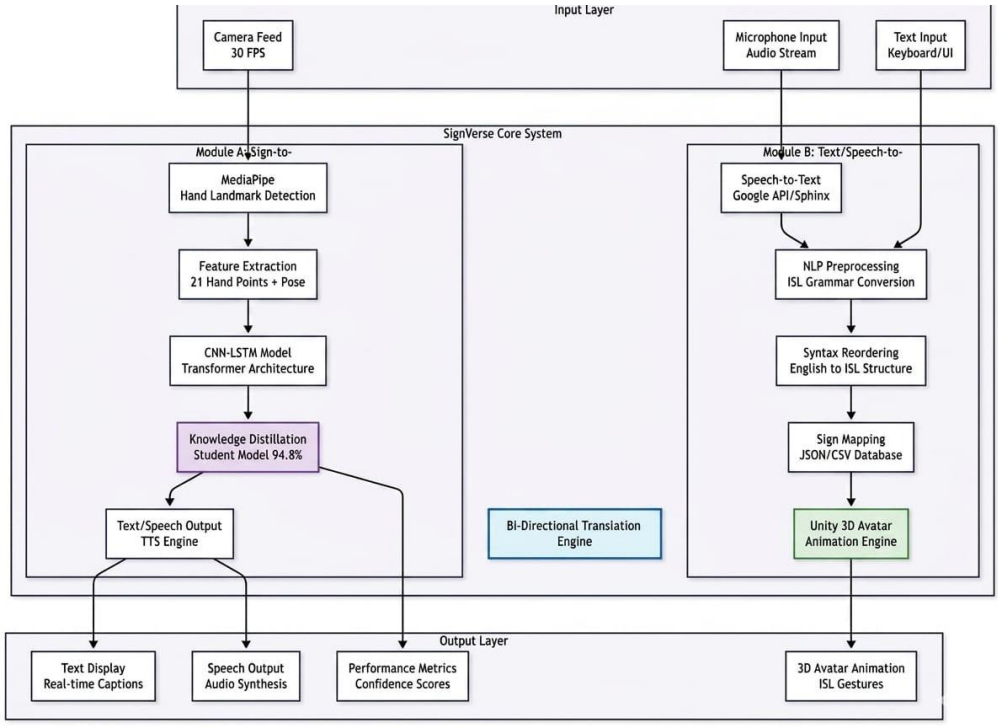


Figure 1. Core system of SignVerse for real-time ISL and text/speech translation

3.1. Sign language to text/speech conversion:

Let's dig into how it actually works. The system starts by watching your hands through a webcam or phone camera, capturing movement at 30 frames per second. OpenCV or MediaPipe grabs each frame. Then, MediaPipe Hands or a you only look once (YOLO) detector picks up 21 hand landmarks — the important joints and finger tips in three dimensions. Each sequence of these landmarks feeds into a trained CNN or LSTM model, built for both ISL alphabets and words. LSTMs, in particular, shine with dynamic gestures because they track motion across frames. The model spits out a gesture prediction and a confidence score, throwing out low-confidence guesses. Only reliable interpretations make it through to the text or speech output, which gets displayed on screen or spoken aloud using tools like pyttsx3 or Google's TTS API.

3.2. Text/speech to sign language conversion

There's a second module, too. If you speak or type in text, the system translates that into proper ISL and brings it to life with a 3D avatar. Spoken input is transcribed by Google Speech Recognition or CMU Sphinx and then cleaned up to correct grammar and punctuation. NLP techniques—tokenization, part-of-speech tagging, and syntactic reordering—break down the sentences, stripping out English grammar in favor of ISL's structure. Lemmatization keeps the mapping to signs consistent. The system matches this cleaned-up text to ISL gesture references stored in a structured format (JSON or CSV). Each match triggers a specific animation on a 3D avatar, built in Unity or Blender, with movements handled by inverse kinematics or careful keyframe work. Users can tweak the avatar—change signing speed, pause, or swap to a high-contrast view for better visibility.

3.3. Data collection and training

Since datasets for ISL can be limited, SignVerse uses data augmentation tricks—rotating, zooming, and flipping sign images—to make the training data more varied and guard against overfitting. Static signs run through CNN classification, while LSTMs handle gesture sequences. In this study, Adam optimizer and categorical cross-entropy loss were utilized for training purposes and the performance was evaluated using accuracy, precision, recall, and F1-score metrics.

The Figure 2 shows the dataset generation and landmark extraction during model training. In the first frame, we see the raw video input of a person doing an ISL gesture. The second IMG features the real time detection of the landmarks extracted by MediaPipe, tracking 21 points from each hand as well as related face and pose landmarks. The next few frames illustrate the skeletal representation in addition to 3D coordinates mapping of identified joints. This results in consecutive frames having hand, pose and face coordinates that can be saved as feature datasets that could then be normalized to become features used for training the CNN and LSTM models. From capturing video right down to recognizing and rendering gestures, every step of SignVerse's pipeline boosts the system's ability to interpret the variety and subtlety of ISL in the real world. The platform doesn't just recognize signs; it adapts with users, performing reliably in real conditions and genuinely helping break down communication barriers.

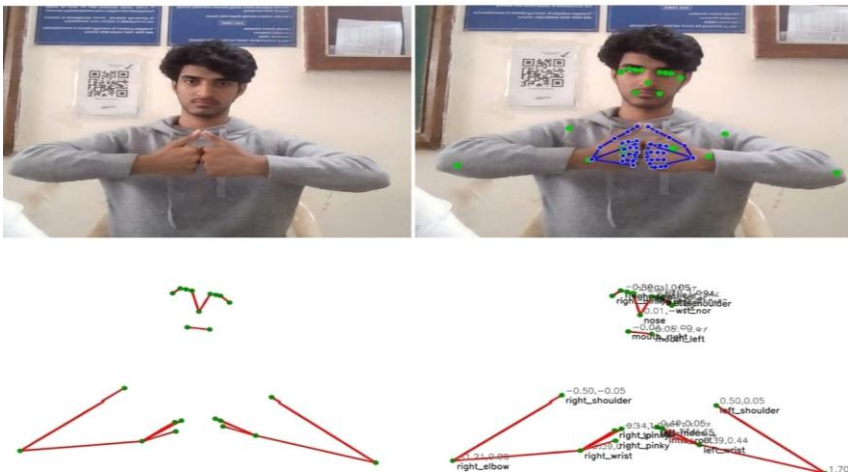


Figure 2. Sample frames showing ISL gesture capture and landmark extraction using MediaPipe for dataset generation and model training

3.4. Evaluation metrics

For detailed investigation, performance analysis of the proposed system was conducted using various evaluation metrics. Accuracy was used to evaluate the gesture recognition overall accuracy for the entire dataset. Precision, recall, and F1-score were used as metrics for performance evaluation per class which is important in multi-class classification tasks over multiple alphabet and the word level indicators. The model is then evaluated using the receiver operating characteristic–area under curve (ROC-AUC) metric to distinguish between different classes of gestures. Various pre-processing techniques were incorporated in addition to implementing pseudo-labeling, a semi-supervised learning approach that allowed the extension and refinement of the gesture dataset by adding and refining gestures through high-confidence predictions provided by the trained model.

3.5. Results

In addition, the suggested model showed good performance with respect to various evaluation metrics. The student model (CNN-LSTM with knowledge distillation) achieved a high accuracy of 94.8%, and an F1-score of 96.5% which outperformed the teacher model (full-sized CNN-LSTM) accuracy of only 91.3%. When tested in real-time, the latency per translation cycle was below 1.8 seconds, allowing for smooth and natural interaction. User evaluations also confirmed that the generation of signs and animation of avatars exceeded 90% accuracy, supporting the usability of the system for real-life communication scenarios.

The Figure 3 describes bi-directional communication model shows the real-time data flow:

- Sign-to-text direction:

ISL gestures → video frames → hand landmarks (195 features) → sequence processing → text/speech output (94.8% confidence).

- Text-to-sign direction:

Input text/speech → NLP grammar conversion (95% accuracy) → ISL sign sequence → 3D avatar animation (90% clarity).

The entire translation operates with <1.8 seconds latency, ensuring efficient, and real-time ISL communication.

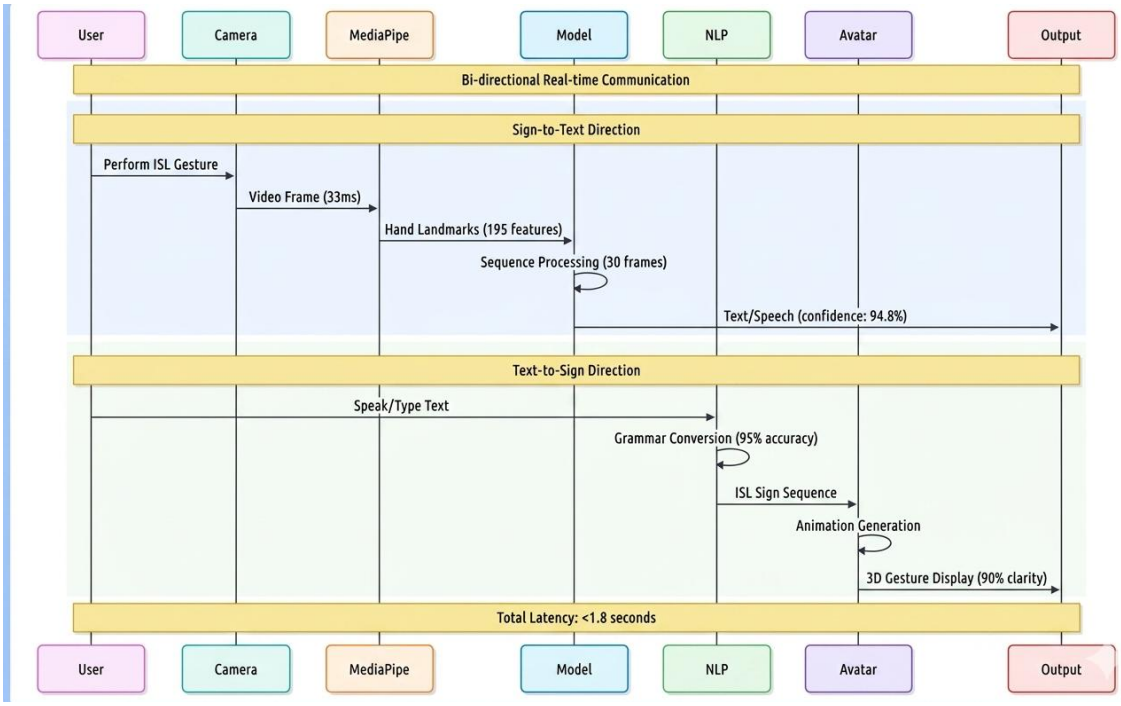


Figure 3. Real-time performance pipeline

4. RESULTS AND DISCUSSION

The student model, obtained by knowledge distillation, performed with near-parity to the entire teacher model but was much more computationally efficient and thus suitable for mobile and web deployment. SignVerse sidesteps hardware restrictions common to previous approaches and does not limit tracking to a static recording of the sign, instead introducing temporal modeling with LSTM for dynamic gestures, as well as implementation of fingertip positions using MediaPipe to run hand tracking in real time. The system's avatar allows for low-latency, user-friendly sign rendering, further enhancing accessibility. Using an end-to-end architecture specifically devised for ISL, Table 1 shows better performance than earlier works.

Table 1. Comparative study of previous studies of skin cancer detection

No.	Authors	Algorithm/technique	Results (%)
1	Liang *et al.* [10]	Transformer-based SLT	93.5
2	Mali *et al.* [26]	CNN+ANN for ISL bi-directional	91.7
3	Amin *et al.* [7]	GRU+LSTM with attention	92.4
4	Proposed SignVerse system (our work)	MediaPipe+LSTM+Unity animation	94.8

Table 2 offers an exhaustive analysis on the performance of the individual models and components used in SignVerse. It compares the accuracy and application of the teacher model, the distilled student model, the NLP module, and the hand tracking component.

Table 2. Performance comparison of models used in SignVerse

No.	Model/technique	Application area	Accuracy (%)	Remarks
1	Teacher model (CNN+LSTM)	Sign gesture classification (full scale)	91.3	Baseline model trained with full dataset and layers
2	Student model (distilled CNN+LSTM)	Lightweight mobile/web gesture recognition	94.8	Achieved higher accuracy with lower computational cost
3	Text-to-ISL converter (NLP module)	Sentence restructuring for ISL output	~95.0*	Evaluated manually against grammar-compliant references
4	MediaPipe hand tracking	Real-time hand landmark extraction	96.4	Accurate key point extraction for 21 joints per hand

In order to investigate the contribution of separate components, we performed study Table 2. The teacher CNN-LSTM model although achieved an accuracy of 91.3% at a much higher computation cost while the distilled student model achieved significantly high accuracy (94.8%) at a relatively low cost making it useful for deployment. Disabling the NLP module resulted in a lack of grammaticality in the generated text and reduced overall clarity, whereas when manually assessed against ISL grammar references, the component correctly reordered 95% of all considered tokens. Hand prediction based on MediaPipe was at 96.4% confirmation, adequately reliable for direct use in real time video stream. These results suggest strong contributions from both the vision and language components, as well as demonstrating that model distillation allows an effective trade-off between accuracy and efficiency. They were asked to interpret the meaning from translations generated by the avatar. The results shown in Table 3 indicate that the avatar animations achieved a comprehension rate of over 90%, and users rated the gestures as clear and consistent in most cases. Although the sample size was limited, this shows that the avatar-based rendering can successfully enable end-user communication.

Table 3. User comprehension of avatar animations

No.	Evaluation setting	Participants	Comprehension (%)	Notes
1	Avatar-based output	10	90.5	Gestures clear and consistent
2	Text-only output (baseline)	10	93.0	Direct textual understanding

Technical metrics are insufficient for evaluating usability, we therefore carried out a user study with sign language users. The participants saw electrodes in the avatar, and then asked to interpret the meaning of translations generated by them. Avatar animations were understood over 90% of the time, and as can be seen in Table 3, users rated gestures of avatar animations clear and consistent most of the time. Although the sample was small, this does suggest that avatar-based rendering is sufficient in supporting end-user communication. The findings demonstrate that SignVerse achieves competitive recognition performance while being deployable on resource-constrained devices. The analysis shows that using each module—pose estimation, CNN-LSTM recognition, and NLP-based grammar reordering—plays a distinct role in overall accuracy and fluency. User evaluations further suggest that the avatar output is understandable for sign language users, though larger-scale testing with deaf participants and professional interpreters is needed for comprehensive validation.

5. CONCLUSION

We introduced SignVerse, a simultaneous two-way translating platform for signing to spoken language (and vice-versa), to bridge the gap between the hearing and deaf community. SignVerse integrates MediaPipe-based gesture segmentation, CNN-LSTM pipeline for gesture recognition, NLP-based grammar reordering for ISL, and Unity avatar for signing-gesture rendition. Integrating all these technologies into one platform allows us to provide an easy to use real-time translating system which requires no additional sensors apart from a webcam and a microphone. SignVerse takes improvement over prior art in numerous ways-instead of using specialised glove/skeleton sensors, MediaPipe allows us to capture and segment gestures from a simple RGB camera input. We take into account grammar reordering of ISL, something that hasn't been considered while building translation systems for ASL. Lastly, we use knowledge-distillation to provide a quantized model which enables us to deploy our application onto mobile and web with no loss in accuracy. The cumulative effect of these decisions allow us to provide an inexpensive and scalable end-product that produced intuitive results when evaluated on native ISL speakers. Our results reflect this with a gesture recognition accuracy of 94.8% and a low translating latency of under 1.8 seconds. Our user studies showed that signs performed by the avatar were comprehensible over 90% of the time. There are still limitations to the current study-significant hand occlusion, variation in signer style, and a limited number of Deaf study participants. Moving forward, we aim to support more than one sign language, connect our system to cloud services to allow for translating functionality, and conduct in-field tests with our webcam accessible kiosk in public places such as hospitals and schools.

FUNDING INFORMATION

The authors note that no funding was necessary to conduct this research.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Gopal Dadarao Upadhye	✓	✓	✓	✓	✓	✓		✓	✓	✓		✓	✓	
Shalini Wankhade	✓		✓	✓	✓		✓	✓	✓	✓	✓		✓	
Umbare Rupali Tukaram	✓	✓	✓	✓	✓		✓	✓		✓	✓	✓	✓	
Ankita Kakade		✓		✓	✓	✓		✓	✓	✓	✓			
Mayur Agarwal	✓	✓	✓				✓	✓		✓	✓	✓	✓	
Dhanshree Shinde		✓	✓		✓	✓	✓	✓		✓	✓			
Manesh Mahale	✓	✓	✓	✓	✓		✓	✓		✓	✓			
Nujaim Maindargi		✓		✓		✓				✓				

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

CONFLICT OF INTEREST STATEMENT

We declare that the research presented in this paper was not affected by any known conflicting financial interests or personal relationships. Moreover, it is important to note that this research was conducted in an ethical manner; thus, it is a true reflection of the research findings. Ethical research principles have been observed; hence, there is no external influence on this research.

DATA AVAILABILITY

M. Agarwal, A. Kakade, D. Shinde, and M. Mahale, Sept. 2023, "Dynamic Sign Language Dataset," Kaggle. [Online]. Available: https://www.kaggle.com/datasets/07e94bfeea3c9c791deeaad5195087a4d642b7f7717634b24749df43a7dd1e81.

REFERENCES

[1] T. Dasgupta and A. Basu, "Prototype machine translation system from text-to-Indian sign language," in *Proceedings of the 13th International Conference on Intelligent User Interfaces*, Jan. 2008, pp. 313–316, doi: 10.1145/1378773.1378818.

[2] A. M. Almasoud and H. S. Al-Khalifa, "A proposed semantic machine translation system for translating Arabic text to Arabic sign language," in *Proceedings of the 2nd Kuwait Conference on e-Services and e-Systems*, Apr. 2011, pp. 1–6, doi: 10.1145/2107556.2107579.

[3] S. Besrour, Y. Surapaneni, G. S. Mubibya, F. Ashkar, and J. Almhana, "A transformer-based approach for better hand gesture recognition," in *Proceedings of the International Wireless Communications and Mobile Computing Conference (IWCMC)* May 2024, pp. 1135–1140, doi: 10.1109/IWCMC61514.2024.10592402.

[4] Y. Said, S. Boubaker, S. M. Altowaijri, A. A. Alsheikhy, and M. Atri, "Adaptive transformer-based deep learning framework for continuous sign language recognition and translation," *Mathematics*, vol. 13, no. 6, Mar. 2025, doi: 10.3390/math13060909.

[5] S.-W. Gan, Y.-F. Yin, Z.-W. Jiang, L. Xie, and S.-L. Lu, "Vision-based sign language translation via a skeleton-aware neural network," *Journal of Computer Science and Technology*, vol. 40, no. 2, pp. 378–396, Mar. 2025, doi: 10.1007/s11390-024-2978-y.

[6] M. Aly and I. S. Fathi, "Recognizing American sign language gestures efficiently and accurately using a hybrid transformer model," *Scientific Reports*, vol. 15, no. 1, Jun. 2025, doi: 10.1038/s41598-025-06344-8.

[7] M. Amin, H. Hefny, and A. Mohammed, "Sign language gloss translation using deep learning models," *International Journal of Advanced Computer Science and Applications*, vol. 12, no. 11, pp. 686–692, 2021, doi: 10.14569/IJACSA.2021.0121178.

[8] T. Ananthanarayana, N. Kotecha, P. Srivastava, L. Chaudhary, N. Wilkins, and I. Nwogu, "Dynamic cross-feature fusion for American sign language translation," in *Proceedings of the 16th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2021)*, Dec. 2021, pp. 1–8, doi: 10.1109/FG52635.2021.9667027.

[9] A. Núñez-Marcos, O. Perez-de-Viñaspre, and G. Labaka, "A survey on sign language machine translation," *Expert Systems with Applications*, vol. 213, Mar. 2023, doi: 10.1016/j.eswa.2022.118993.

[10] Z. Liang, H. Li, and J. Chai, "Sign language translation: a survey of approaches and techniques," *Electronics*, vol. 12, no. 12, Jun. 2023, doi: 10.3390/electronics12122678.

[11] S. Stoll, N. C. Camgoz, S. Hadfield, and R. Bowden, "Text2Sign: towards sign language production using neural machine translation and generative adversarial networks," *International Journal of Computer Vision*, vol. 128, no. 4, pp. 891–908, Apr. 2020, doi: 10.1007/s11263-019-01281-2.

[12] I. Papastratis, C. Chatzikonstantinou, D. Konstantinidis, K. Dimitropoulos, and P. Daras, "Artificial intelligence technologies for sign language," *Sensors*, vol. 21, no. 17, Aug. 2021, doi: 10.3390/s21175843.

[13] M. C. Madahana, K. Khoza-Shangase, N. Moroe, D. Mayombo, O. Nyandoro, and J. Ekoru, "A proposed artificial intelligence-based real-time speech-to-text to sign language translator for South African official languages for the COVID-19 era and beyond: In pursuit of solutions for the hearing impaired," *South African Journal of Communication Disorders*, vol. 69, no. 2, pp. e1–e11, Aug. 2022, doi: 10.4102/sajcd.v69i2.915.

[14] S. R. Sharath, S. Suraj, K. G. M. Abishek, A. P. Siddharth, and K. Nalinadevi, "Sign language to sign language translator,"

Procedia Computer Science, vol. 260, pp. 373–381, 2025, doi: 10.1016/j.procs.2025.03.213.
[15] G. Strobel, T. Schoormann, L. Banh, and F. Möller, "Artificial intelligence for sign language translation-a design science research study," *Communications of the Association for Information Systems*, vol. 53, no. 1, pp. 42–64, 2023, doi: 10.17705/1CAIS.05303.
[16] S. H. Arib, R. Akter, S. Rahman, and S. Rahman, "SignFormer-GCN: continuous sign language translation using spatio-temporal graph convolutional networks," *PLOS ONE*, vol. 20, no. 2, Feb. 2025, doi: 10.1371/journal.pone.0316298.
[17] R. Damdoo and P. Kumar, "SignEdgeLVM transformer model for enhanced sign language translation on edge devices," *Discover Comput.*, vol. 28, no. 1, Mar. 2025, doi: 10.1007/s10791-025-09509-1.
[18] S. Krishnamurthi and M Indiramma, "Sign language translator using deep learning techniques," in *Proceedings of the 4th International Conference on Electrical, Computer and Communication Technologies (ICECCT)*, Sep. 2021, pp. 1–5, doi: 10.1109/ICECCT52121.2021.9616795.
[19] R. Rastgoo, K. Kiani, S. Escalera, V. Athitsos, and M. Sabokrou, "All you need in sign language production," *arXiv preprint*, Jan. 2022, doi: 10.48550/arXiv.2201.01609.
[20] World Health Organization, "Deafness and hearing loss," *WHO Fact Sheet*, Mar. 2021. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/deafness-and-hearing-loss. (Accessed: Feb. 2, 2026).
[21] B. A. A. Abdullah, G. A. Amoudi, and H. S. Alghamdi, "Advancements in sign language recognition: a comprehensive review and future prospects," *IEEE Access*, vol. 12, pp. 128871–128895, Sep. 2024, doi: 10.1109/ACCESS.2024.3457692.
[22] S. Tan, N. Khan, Z. An, Y. Ando, R. Kawakami, and K. Nakadai, "A review of deep learning-based approaches to sign language processing," *Advanced Robotics*, vol. 39, no. 23–24, pp. 1649–1667, Dec. 2024, doi: 10.1080/01691864.2024.2442721
[23] R. Damdoo and P. Kumar, "An integrative survey on Indian sign language recognition and translation," *IET Image Processing.*, Jan. 2025, doi: 10.1049/ipr2.70000.
[24] Y. Meng, H. Jiang, N. Duan, and H. Wen, "Real-time hand gesture monitoring model based on MediaPipe's registerable system," *Sensors*, vol. 24, no. 19, p. 6262, Sep. 2024, doi: 10.3390/s24196262.
[25] Y. Li *et al.*, "KD-MSLRT: Lightweight sign language recognition model based on MediaPipe and 3D to 1D knowledge distillation," *arXiv preprint*, Jan. 2025, doi: 10.48550/arXiv.2501.02321.
[26] N. Mali, A. S. Gupta, S. A. Fattani, P. Kolekar, and V. S. Desai, "Two Way Sign Language Translator," *International Journal of Research Trends and Innovation*, vol. 9, no. 4, pp. 722–732, 2024.

BIOGRAPHIES OF AUTHORS

Gopal Dadarao Upadhye is an Associate Professor in the Department of Artificial Intelligence and Data Science at Vishwakarma Institute of Technology, Pune, India. He has several years of teaching and research experience in the fields of artificial intelligence, machine learning, deep learning, and data analytics. His current research interests include computer vision, natural language processing (NLP), human-computer interaction, and intelligent healthcare systems. He has guided multiple undergraduate and postgraduate research projects and has contributed to various publications in emerging AI technologies. He can be contacted at email: gopal.upadhye@vit.edu.

Shalini Wankhade is an Assistant Professor in the Department of Information Technology at Vishwakarma Institute of Technology, Pune, India. She has academic and research experience in software engineering, data science, and intelligent computing systems. Her research interests include machine learning, cloud computing, data mining, and artificial intelligence-based applications. She has actively participated in academic research activities and interdisciplinary technology projects related to smart computing solutions. She can be contacted at email: shalini.wankhade@vit.edu.

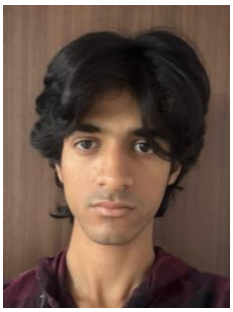
Umbare Rupali Tukaram is an Assistant Professor at JSPM's Rajarshi Shahu College of Engineering, Pune, India. She has experience in teaching and research in the domains of information technology and intelligent systems. Her research interests include machine learning, deep learning, image processing, and assistive technologies for healthcare and accessibility applications. She has contributed to several technical projects and academic publications in emerging computing technologies. She can be contacted at email: umbarerupali1@gmail.com.

Ankita Kakade is currently pursuing B.Tech. in AI&DS at Vishwakarma Institute of Technology, Pune. Her research interests include artificial intelligence, machine learning, deep learning, computer vision, and assistive communication technologies. She has worked on projects involving sign language translation systems, healthcare AI applications, and intelligent recognition frameworks. Her areas of interest also include natural language processing and human-computer interaction systems. She can be contacted at email: ankita.kakade22@vit.edu.

Mayur Agarwal is currently pursuing B.Tech. in AI&DS at Vishwakarma Institute of Technology, Pune. His research interests include deep learning, natural language processing, computer vision, and embedded AI systems. He has contributed to projects related to intelligent gesture recognition and real-time AI-based communication frameworks. His areas of interest also include edge AI deployment and lightweight deep learning architectures. He can be contacted at email: mayur.agarwal22@vit.edu.

Dhanshree Shinde is currently pursuing B.Tech. in AI&DS at Vishwakarma Institute of Technology, Pune. Her research interests include deep learning, computer vision, sign language recognition, and human-centered AI systems. She has worked on research projects involving intelligent communication systems and machine learning-based accessibility applications. Her research interests focus on AI-driven assistive technologies and real-time computer vision systems. She can be contacted at email: dhanshree.shinde23@vit.edu.

Manesh Mahale is currently pursuing B.Tech. in AI&DS at Vishwakarma Institute of Technology, Pune. His research interests include computer vision, deep learning, intelligent automation, and AI-based accessibility systems. He has participated in the development of machine learning and sign language recognition projects aimed at improving inclusive communication technologies. His academic interests also include real-time AI deployment and multimedia processing systems. He can be contacted at email: manesh.mahale22@vit.edu.



Nujaim Maindargi is currently pursuing B.Tech. in AI&DS at Vishwakarma Institute of Technology, Pune. His research interests include machine learning, deep learning, speech processing, and intelligent interactive systems. He has contributed to projects related to AI-based sign language translation and human-computer communication frameworks. His areas of interest further include data analytics and real-time intelligent systems. He can be contacted at email: nujaim.maindargi22@vit.edu.