© 2026 IJRTI | Volume 11, Issue 5 May 2026 | ISSN: 2456-3315

Agentic AI-Based Sign Language Translation Using Multi-Agent Architecture

[1]Khushbu Mahale, [2]Vedant Sawant, [3]Shardul Mane, [4]Neha Nandurkar
[1] UG Scholar, [2] UG Scholar, [3] UG Scholar, [4] Assistant Professor
[1] Artificial Intelligence and Data Science,
[1]Rajiv Gandhi Institute of Technology, Mumbai, India
[1] Khushbumahale12@gmail.com , [2] vedantsawant0503@gmail.com , [3] shardulmane20@gmail.com ,
[4] neha.nandurkar@mctrgit.ac.in

***Abstract*—** Sign language is essential for facilitating communication among individuals with hearing and speech disabilities. Nevertheless, communication hurdles persist as many people lack knowledge of sign language, and automated translation tools have not reached their full potential. Recent progress in artificial intelligence, computer vision, and natural language processing has led to the creation of advanced systems that can convert spoken or written language into sign language. This paper proposes an Agentic AI–based text-to-sign language translation system that converts natural language sentences into animated sign language representations using a multi-agent architecture. The proposed framework consists of several intelligent agents responsible for text understanding, gloss translation, motion retrieval, and animation generation. The system first processes user text input and applies natural language processing techniques to detect sentence intent and normalize grammatical structures. A gloss translation agent converts the processed text into Indian Sign Language gloss representation, which is then used by a motion lookup agent to retrieve corresponding sign pose sequences. The motion data is processed and mapped to a 3D avatar skeleton using a Blender-based animation pipeline to generate realistic sign language animations. By combining Agen-tic AI decision-making with motion-based sign representation, the system enables scalable and modular sign language translation. The proposed approach aims to improve accessibility for hearing-impaired individuals while providing an interactive web-based communication platform.

***Index Terms*—** Agentic AI, Sign Language Translation, Natural Language Processing, Gesture Recognition, Computer Vision, 3D Avatar Animation Introduction.

I. INTRODUCTION

Communication is a vital part of human interaction; however, those with hearing or speech challenges frequently face issues when engaging with people who are unfamiliar with sign language. Sign language operates as a visual communication method that employs hand gestures, facial expressions, and body movements to express ideas. It is estimated that millions around the world depend on sign languages as their main mode of communication. [1] Despite its extensive usage, most of the hearing population is not knowledgeable about sign language, resulting in communication obstacles between hearing-impaired individuals and the larger community. Consequently, researchers have concentrated on creating automated systems for translating sign language that can help close this gap by transforming spoken or written language into sign language representations. [10]

Recent developments in artificial intelligence and deep learning have greatly enhanced technologies for recognizing and translating sign language. Methods in computer vision, particularly convolutional neural networks (CNNs), have demonstrated high precision in identifying hand gestures and converting them into written or spoken language. [6] Alongside gesture recognition, contemporary systems also utilize natural language processing and generative AI to facilitate immediate translation among text, speech, and sign language. Platforms powered by AI can comprehend linguistic context and create dynamic sign language animations using 3D avatars, thereby increasing accessibility and user-friendliness. [3]

Agentic AI architecture has become a potential method for creating intelligent systems that can coordinate several agents, organize tasks, and carry out intricate reasoning processes. These systems are made up of specialized agents that work together to accomplish high-level goals, making AI applications more adaptable and scalable. [2] Inspired by these developments, this study suggests an Agentic AI-based system for translating text to sign language. The system processes text input, transforms it into a sign language gloss representation, retrieves gesture motion sequences, and creates lifelike 3D avatar animations using a multi-agent architecture.

II. RELATE WORK

1. Application of AI in Sign Language Translation Domain

In recent years, Artificial Intelligence (AI) has significantly transformed accessibility technologies, particularly in sign language recognition and translation systems. This project aims to integrate AI into the communication ecosystem for Deaf and Hard-of-Hearing individuals by leveraging machine learning, natural language processing (NLP), computer vision, and 3D animation techniques. The authors reviewed over 60 research papers specifically focused on AI-based sign language recognition, multilingual translation, gesture detection, and avatar generation systems. Out of these, approximately 25 papers were selected based on relevance to Indian Sign Language (ISL), multilingual translation, and real-time accessibility solutions.

Traditional sign language systems primarily focus on either gesture-to-text or text-to-sign translation, but many lack multilingual capabilities, adaptive learning, or real-time 3D avatar support. Through extensive literature review and market analysis, it was observed that very few systems combine LLM-based natural language understanding, fallback rule-based gloss generation, named entity recognition, and user feedback loops into a unified framework. This research gap highlights the novelty of the proposed system, which aims to automate multilingual text-to-sign conversion while improving translation quality through continuous feedback and correction mechanisms.

2. Existing Research on AI-Based Sign Language Translation

Sensor-based systems that used wearable technology or sensors to record hand motions were among the first methods. These systems detected hand movements and converted them into vocal output employing hardware elements like accelerometers and gesture sensors. However, these systems were frequently constrained by their small language sets and need on hardware that is specialized [4]

Vision-based gesture recognition systems that use cameras and deep learning models are another extensively researched method. Strong performance in identifying hand gestures and converting them into text or speech outputs has been shown by CNN-based models. From pictures, these models pick up hierarchical visual characteristics and can get a high level of accuracy when trained on sizable gesture datasets. [6]

Another important development in this field is the use of generative AI and 3D animation technologies to produce dynamic sign language gestures. Generative AI systems combine natural language processing and computer graphics to generate animated sign language avatars, enabling real-time communication platforms that support web and mobile applications. [3]

Recent studies also highlight the potential of multi-agent AI systems in solving complex computational problems. Agentic AI architecture consists of multiple autonomous agents that collaborate through planning, reasoning, and tool usage to accomplish high-level tasks. These architectures offer improved adaptability and modularity compared to traditional AI systems. [2]

Natural Language Processing (NLP) plays a crucial role in sign language translation systems, particularly in translating words spoken or written into a sign language gloss. NLP techniques such as tokenization, stemming, and syntactic reordering are commonly used to convert sentences into a structure compatible with sign language grammar. Since sign languages often follow different grammatical rules compared to spoken languages, rule-based transformations and linguistic analysis are necessary for accurate translation. [13]

Several specialized agents make up agent-based artificial intelligence systems that collaborate to achieve complex goals through planning and reasoning. These systems enable distributed problem solving where each agent performs a specific task within a larger workflow. The modular nature of agent-based architecture allows improved scalability, flexibility, and maintainability compared to monolithic AI systems. [2]

Sign language identification problems have made extensive use of machine learning-driven methods. Najib (2025) proposed a multi-lingual sign language recognition system (MSLI) that utilizes machine learning models trained on multiple datasets across different languages. The system demonstrated high performance with training accuracy up to 95.87% and testing accuracy of 92.33%, highlighting the effectiveness of combining datasets for multilingual recognition. However, most existing systems focus primarily on recognition rather than complete translation pipelines. [14]

Recent research has investigated the application of artificial intelligence for improved communication accessibility for hearing-impaired individuals. Ramya et al. (2025) proposed an AI-enabled speech-to-sign language conversion system that leverages speech recognition and natural language processing techniques to translate spoken input into sign language animations. Their system utilizes deep learning models and a gesture database to generate accurate and real-time visual representations, demonstrating improved accessibility and usability across multiple devices. The study emphasizes how crucial it is to combine AI with assistive technology in order to close communication barriers. [15]

Recent advancements in agentic AI systems have introduced new approaches for handling complex workflows through multiple intelligent agents. Shukla (2025) proposed an adaptive monitoring framework for evaluating agentic AI systems using multi-dimensional metrics such as performance, robustness, safety, and human-centered interaction. The study introduces the Adaptive Multi-Dimensional Monitoring (AMDM) algorithm, which improves anomaly detection and system evaluation. The findings highlight the importance of balanced evaluation frameworks for ensuring reliability and scalability in multi-agent systems. [19]

3. Advanced AI Techniques in Sign Language Generation

Research into modern sign language generation models has evolved to include multiple AI techniques:

- Text-Based NLP Models: Transformer and LLM-based systems such as GPT, BERT, and Qwen improve sentence understanding and ISL gloss generation.
- Named Entity Recognition (NER): Tools like spaCy identify names, places, and uncommon words for fingerspelling.
- Rule-Based Fallback Systems: Grammar restructuring and gloss correction modules improve reliability when primary AI models fail.
- 3D Avatar Rendering: Three.js, Blender rigs, and motion keyframes create realistic visual sign animations.
- Feedback Learning Systems: SQLite or cloud-based memory systems continuously improve translation through user thumbs-up/down corrections.
- Multilingual Translation Layers: Translation engines support broader accessibility across multiple spoken languages before converting into sign

4. Comparative Analysis and Research Gap

While previous studies have explored various aspects of sign language technology such as gesture recognition, text-to-sign translation, avatar-based animation, and multilingual communication, most existing systems focus on individual components rather than delivering a complete end-to-end solution. Many research models emphasize either sign recognition through computer vision or basic translation using predefined gloss datasets, but they often lack contextual language understanding, multilingual adaptability, and real-time user interaction. Additionally, several systems depend solely on a single AI model, which can reduce reliability when handling unknown words, complex sentence structures, or model failures. Most existing solutions also provide limited support for proper noun recognition, fingerspelling, adaptive corrections, and continuous system improvement through user feedback.

In contrast, the proposed Multilingual Voice and Text to Sign Language Translator addresses these limitations by integrating advanced technologies such as LLM-based natural language processing, spaCy based named entity recognition, rule-based fallback mechanisms, missing token correction, SQLite memory learning, and real-time 3D avatar animation within a unified architecture. This comprehensive framework not only enhances translation accuracy and reliability but also introduces adaptive learning capabilities, making it more practical, scalable, and suitable for real-world accessibility applications. Thus, the proposed system fills a significant research gap by combining multiple AI-driven components into a robust, multilingual, and intelligent sign language translation platform.

III. FINDINGS AND DISCUSSION

1. Analysis

The findings indicate that the integration of Artificial Intelligence into multilingual sign language translation is a highly effective solution for bridging communication gaps between hearing individuals and the Deaf or Hard-of-Hearing community. The system demonstrates strong capability in converting text and voice inputs into meaningful sign language representations using advanced AI technologies such as NLP, LLMs, named entity recognition, and 3D avatar animation. Several key points emerged from the analysis:

1. Enhanced Accessibility: The AI-powered model significantly improves communication accessibility by translating user input into Indian Sign Language in real time. By incorporating multilingual support, the system broadens communication possibilities across diverse linguistic groups, making it highly practical for inclusive social, educational, and professional environments.
2. Contextual and Linguistic Accuracy: The integration of LLMs and NLP models allows the system to better understand sentence structure, intent, and context. Unlike static translation systems, the proposed framework generates more grammatically accurate ISL gloss while handling named entities and uncommon words through fingerspelling, thereby improving translation precision.
3. System Reliability Through Hybrid Architecture: A major strength of the model is its fallback mechanism. By combining LLM processing with rule-based gloss generation and missing token correction, the system remains functional even when AI model limitations arise. This hybrid architecture ensures robustness, continuity, and improved user trust.
4. Adaptive Learning and Personalization: The inclusion of SQLite-based memory and user feedback mechanisms allows the system to learn from corrections over time. This continuous improvement feature enhances translation quality and creates a more personalized and adaptive user experience.
5. Scalability and Future Applications: The modular design of the system architecture makes it highly scalable for future enhancements such as facial expression synthesis, emotion-aware translation, larger multilingual datasets, and wearable integrations. This creates strong potential for broader real-world adoption.2. System Performance and Practical Insights

2. Research Gaps

While the proposed system demonstrates significant advancements, several research gaps and challenges were identified:

1. Complexity of Sign Language Grammar: Sign languages possess unique grammatical structures that differ significantly from spoken languages. Although LLM integration improves translation, certain complex sentence formations, emotional contexts, and culturally nuanced expressions remain challenging.
2. Limited Standardized Datasets: One major limitation is the scarcity of comprehensive, high-quality datasets for Indian Sign Language and multilingual sign systems. Limited data diversity may affect system performance across regional signs and uncommon gestures.
3. Avatar Realism and Expression: While 3D avatar animation provides visual accessibility, current systems may lack full emotional depth, facial expressions, and subtle non-manual markers essential for natural sign communication.
4. User Feedback Integration Depth: Although feedback is incorporated, future systems require more sophisticated adaptive learning strategies to dynamically optimize translations based on broader user interactions.
5. Real-Time Processing Challenges: Maintaining low latency while processing multilingual translation, NLP, LLM reasoning, and animation rendering simultaneously remains computationally demanding.

IV. METHODOLOGY

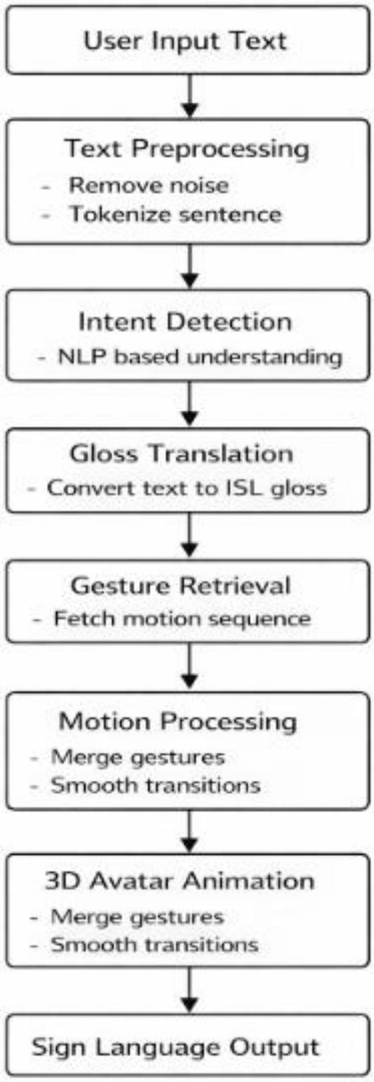

Fig. 1. The methodology workflow of the proposed text-to-sign language translation system.

A. User Input Text:

The process begins when the user enters a sentence or phrase through a web interface or application. This text serves as the initial input to the translation system.

B. Text Preprocessing:

The input sentence undergoes preprocessing to remove noise, punctuation, and unnecessary characters. Tokenization is performed to split the sentence into individual words or tokens for further processing.

C. Intent Detection:

The meaning as well as context of the input text are examined using Natural language processing, or NLP, methods. This step helps the system understand the structure of the sentence and identify the user's intent.

D. Gloss Translation:

The processed sentence is converted into Indian Sign Language (ISL) gloss representation. In this stage, the words are rearranged according to sign language grammar rules, which differ from conventional spoken language syntax.

E. Gesture Retrieval:

Each gloss word is mapped to a corresponding gesture sequence stored in the dataset. The system retrieves the required motion sequences for each word in the translated sentence.

F. Motion Processing:

The retrieved gesture sequences are combined into a continuous sequence. Transition smoothing is applied to ensure natural movement between consecutive gestures.

G. 3D Avatar Animation:

The processed motion data is applied to a rigged 3D avatar model. The avatar performs the corresponding gestures, producing a visual representation of the sign language translation.

H. Sign Language Output:

The final output is displayed as an animated sign language sequence that visually represents the input sentence.

V. SYSTEM ARCHITECTURE

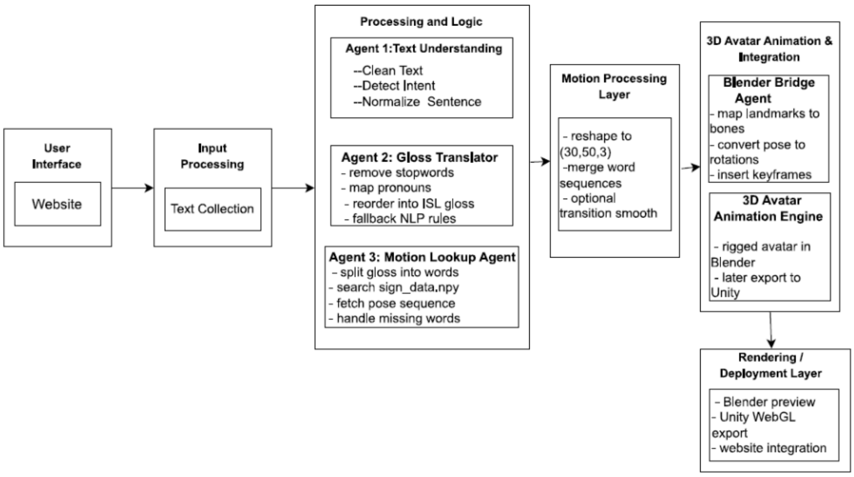

Fig. 2. System architecture of the proposed Agentic AI-based text-to-sign language translation

A. **User Interface (Website / App):** The entry point of the system. Users interact through a web or mobile interface where they type or input text they want converted into sign language. This layer handles all user-facing interaction and triggers the processing pipeline.

B. **Input Processing:** Text Collection Receives the raw text from the UI and prepares it for intelligent processing. This layer handles basic data capture, ensuring the input is ready to be passed downstream in a clean, structured format.

C. **Agent 1:** Text Understanding The first AI agent in the pipeline. It performs three core tasks — cleaning the text (removing noise, fixing typos), detecting the user's intent and normalizing the sentence into a standard grammatical form. This ensures consistent, high-quality input for the translation step.

D. **Agent 2:** Gloss Translator Translates normalized English into ISL gloss, which is the written representation of sign language. It removes filler stop words, maps pronouns to their ISL equivalents, reorders words according to ISL grammar and applies fallback NLP rules when standard mappings fail.

E. **Agent 3**: Motion Lookup Agent Takes the ISL gloss and resolves each word into actual motion data. It splits the gloss into individual signs, searches for a precompiled pose database that fetches the corresponding pose keyframe sequences, and gracefully handles any words not found in the database.

F. **Motion Processing Layer:** A data engineering layer that prepares raw pose sequences for animation. It reshapes pose arrays into a standardized (30 frames, 50 landmarks, 3 axes) tensor format, merges sequences from multiple words into a continuous motion stream and optionally smooths transitions between signs to make the animation look natural.

G. **Blender Bridge Agent:** Acts as the connector between pose data and 3D animation. It maps body landmark coordinates onto the avatar's rig bones, converts raw positional data into bone rotation values, and programmatically inserts keyframes into Blender's animation timeline.

H. **3D Avatar Animation Engine:** The core animation layer where the rigged 3D avatar lives inside Blender. It receives keyframe data and drives the avatar's movements. This layer is also the bridge to Unity for future export, enabling real-time rendering in game engines.

I. **Rendering / Deployment Layer:** The final output layer with three deployment targets — a Blender preview for development and testing, a Unity WebGL build for browser-based deployment, and direct website integration so end users can view the animated ISL avatar in real time

VI. RESULT

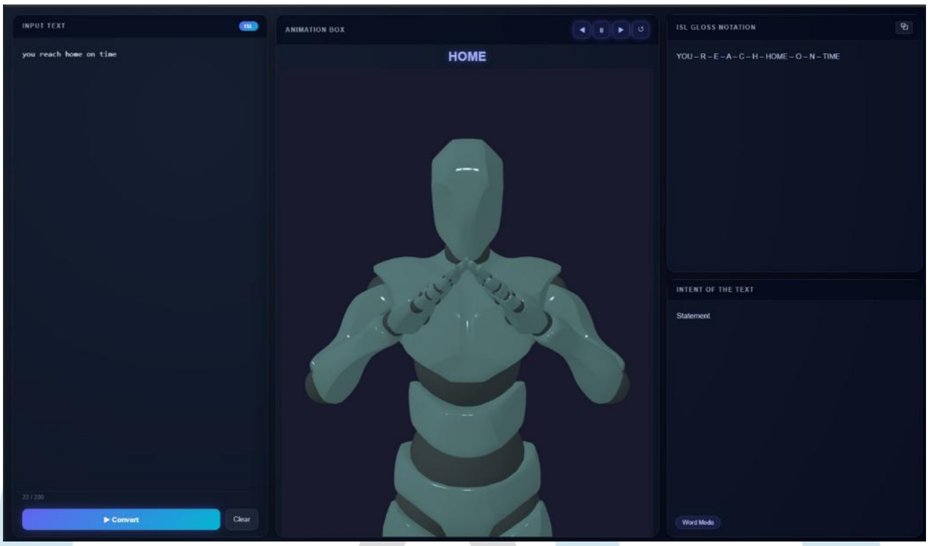

Fig. 3. Text-to-sign language translation output with 3D avatar animation.

The proposed method was assessed using real-time text input, as shown in Fig. 3. The input sentence "you reach home on time" is processed through the multi-agent pipeline. The system converts the input into ISL gloss notation, displayed as "YOU–R–E–A–C–H–HOME–O–N–TIME", and identifies the intent as a statement.

The corresponding gesture sequences are retrieved from the dataset and used to generate a 3D avatar animation, where the avatar performs the sign for "HOME". The results demonstrate that the system effectively translates text into structured gloss and generates meaningful sign language animations.

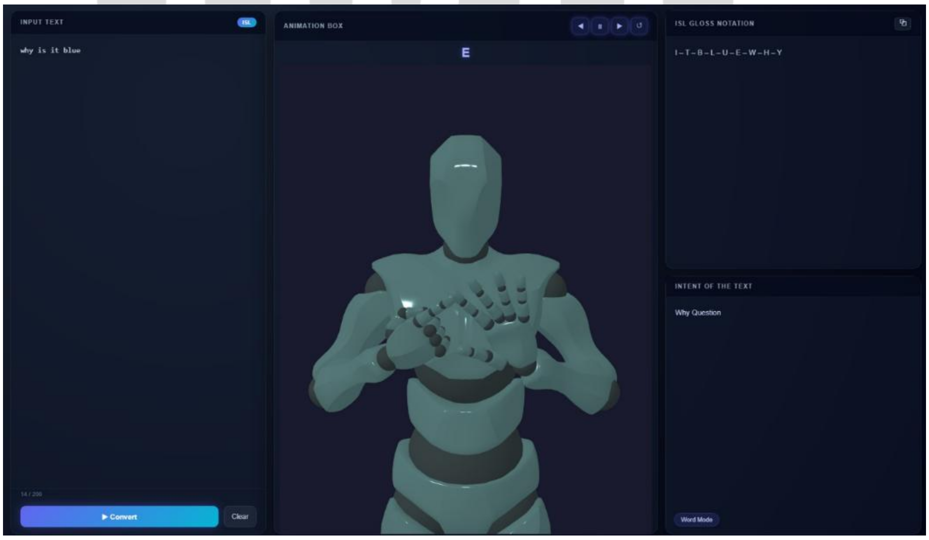

Fig. 4. Question-based text-to-sign language translation output with avatar animation.

The system was tested with the input sentence "why is it blue", as depicted in Figure 4. The multi-agent pipeline is used to process the input, where it is converted into ISL gloss notation displayed as "I–T–B–L–U–E–W–H–Y". The system correctly identifies the intent of the sentence as a "Why Question".

The corresponding gesture sequences are retrieved and used to generate the 3D avatar animation, where the avatar performs the required sign gestures. The results demonstrate that the system effectively handles interrogative sentences and generates appropriate sign language output.

VII. CONCLUSION

This paper presented an Agentic AI-based text-to-sign language translation framework that converts natural language input into animated sign language using a multi-agent architecture. The proposed system integrates natural language processing, motion sequence retrieval, and 3D animation technologies to generate realistic sign language gestures. The modular design of the system enables scalability and adaptability, allowing the integration of additional agents or improved machine learning models in the future. The proposed approach contributes to improving accessibility for hearing-impaired individuals and demonstrates the potential of Agentic AI systems in assistive communication technologies. Future work will focus on improving gesture datasets, enhancing avatar realism, and integrating real-time translation capabilities for multilingual communication. Additionally, incorporating advanced deep learning models and larger sign language datasets can further improve translation accuracy and system robustness. The integration of cloud-based deployment and real-time user interaction modules can also enable the system to be used in educational platforms, public communication systems, and assistive technologies for inclusive digital communication.

REFERENCES

[1] Alaghband M., Maghroor H. R., Garibay I. (2023), "A Survey on Sign Language Litera-ture", Machine Learning with Applications, Elsevier, Vol. 13, 100504. https://doi.org/10.1016/j.mlwa.2023.100504.

[2] Rouf M., Jadon J. S. (2025), "Generative AI for Text to Speech and Sign Language Translation", *2025 3rd International Conference on Disruptive Technologies (ICDT)*, IEEE.

[3] Shukla M. A. (2025), "Evaluating Agentic AI Systems: A Balanced Framework for Performance, Robustness, Safety and Beyond", Independent Research Study.

[4] Anusha L., Usha Devi Y. (2019), "Implementation of Gesture Based Voice and Language Translator for Dumb People", *International Conference Research Paper*.

[5] Sudhan M. B., Senthil Kumar K., Usha Bala V., Nirmala G., Abijith G. R., Karthikeyan G. (2025), "Real-Time Sign Language Translator with IoT and Flex Sensors for Multilingual Support", *7th International Conference on Inventive Material Science and Applications (ICIMA)*, IEEE.

[6] Ojha A., Pandey A., Maurya S., Thakur A., Dayananda P. (2020), "Sign Language to Text and Speech Translation in Real Time Using Convolutional Neural Network", *International Journal of Engineering Research & Technology (IJERT)*.

[7] Shukla M., Gupta H. (2025), "Advancing Sign Language Interpretation with Transfer Learning and Multimodal Features", Research Publication.

[8] Shukla M., Gupta H., Sharma A. (2025), "Sign Language Interpreter using Deep Learning", Research Article. https://doi.org/10.21203/rs.3.rs-7465375/v1.

[9] Sindhu K. S., Mehnaaz, Nikitha B., Varma P. L., Uddagiri C. (2024), "Sign Language Recognition and Translation Systems for Enhanced Communication for the Hearing Impaired", *International Conference on Cognitive, Green and Ubiquitous Computing*, IEEE.

[10] Antad S. M., Chakrabarty S., Bhat S., Bisen S., Jain S. (2024), "Sign Language Translation Across Multiple Languages", *International Conference on Emerging Systems and Intelligent Computing (ESIC)*, IEEE.

[11] Peguda J., Deepa R. N., Santosh V. S. S., Mounish V., Vijayalata Y. (2022), "Speech to Sign Language Translation for Indian Languages", *8th International Conference on Advanced Computing and Communication Systems (ICACCS)*, IEEE.

[12] Yadav A., Verma V. K., Saxena R., Srivastava V., Saini B. (2021), "Audio to Sign Language Translator Web Application", *International Conference on Computational Performance Evaluation (ComPE)*, IEEE.

[13] Najib F. M. (2025), "A Multi-lingual Sign Language Recognition System Using Machine Learning", *Multimedia Tools and Applications*, Springer, Vol. 84, pp. 27987–28011. https://doi.org/10.1007/s11042-024-20165-3

[14] Ramya, D., Keerthana, D., Madhav Priyan, S., Nanthene, R.E., Rajalingam, P.: AI enabled speech to sign language conversion. International Research Journal of Modernization in Engineering Technology and Science (IRJMETS), 7(4), 2025. DOI: https://doi.org/10.56726/IRJMETS72763

[15] Ulisses, J., Oliveira, T., Rocha, E., Escudeiro, P.M., Escudeiro, N., Barbosa, F.M.: Blind/Deaf communication API for assisted translated educational digital content. In: Proceedings of the IEEE Conference, 2020.

[16] Punekar, A.C., Nambiar, S.A., Laliya, R.S., Salian, S.R.: A translator for Indian sign language to text and speech. International Journal for Research in Applied Science and Engineering Technology (IJRASET), 8(6), 2020. DOI: https://doi.org/10.22214/ijraset.2020.6267

[17] Fox, N., Woll, B., Cormier, K.: Best practices for sign language technology research. Universal Access in the Information Society, 24, 69–77 (2025). https://doi.org/10.1007/s10209-023-01039-1

[18] Shukla, M.A.: Adaptive monitoring and real-world evaluation of agentic AI systems. arXiv preprint arXiv:2509.00115 (2025).

[19] Papadimitriou, K., Potamianos, G., Sapountzaki, G., Goulas, T., Efthimiou, E., Fotinea, S.-E., Maragos, P.: Greek sign language recognition for an education platform. Universal Access in the Information Society, 24, 51–68 (2025). https://doi.org/10.1007/s10209-023-01017-7

[20] Baskar, K., Azarudeen, S., Mythili, S., Reshma, K., Shalini, A. (2025). Voice to Sign-Language Translator Using NLP. *International Journal of Science, Engineering and Technology*, 13(1).

[21] Strobel, G., Schoormann, T., Banh, L., Möller, F. (2023). Artificial Intelligence for Sign Language Translation – A Design Science Research Study. *Communications of the Association for Information Systems*, 52.