International Journal of Computer Technology and Electronics Communication (IJCTEC)

| ISSN: 2320-0081 | www.ijctece.com ||A Peer-Reviewed, Refereed and Bimonthly Journal |

|| Volume 9, Issue 3, May- June 2026 ||

DOI: 10.15680/IJCTECE.2026.0903003

Speech and Text Conversion System for Sign Language using ML

Sangeetha D [1], Dharani Dharan K[2], Krishna A C[3], Karthikeyan C [4]

Assistant Professor, Department of Information Technology, Jaya Engineering College, Anna University, Chennai, Tamil Nadu, India[1]

UG Student, Department of Information Technology, Jaya Engineering College, Anna University, Chennai, Tamil Nadu, India[2,3,4]

Publication History: Received: 08.04.2026; Revised: 30.04.2026; Accepted: 03.05. 2026; Published: 08.05.2026.

ABSTRACT: This project presents a machine learning–based system that integrates computer vision and speech processing techniques to enable intelligent sign language recognition and translation. Deaf and mute individuals primarily rely on sign language for communication; however, the majority of people are not familiar with it, leading to a significant communication gap. Existing solutions depend heavily on human interpreters or manual interpretation methods, which are often costly, time-consuming, error-prone, and unsuitable for real-time communication due to their limited availability. To address these challenges, the proposed system introduces a two-way communication framework that converts sign language gestures into text and speech for non-sign users, while also transforming voice or text input from normal users into sign language for hearing-impaired individuals. By leveraging machine learning algorithms, the system ensures real-time performance, improved accuracy, and cost-effectiveness. This approach reduces dependency on human interpreters and enhances accessibility, enabling seamless and inclusive communication between hearing and speech-impaired individuals and the general population.

KEYWORDS: Machine Learning, Sign Language Recognition, Speech and Text Conversion, Two- Way Communication, Text-to-Speech (TTS).

I. INTRODUCTION

Artificial Intelligence (AI) has significantly transformed the way humans interact with technology, enabling intelligent systems to perform complex tasks such as image recognition, speech processing, and language translation. One of the major challenges in human communication exists between hearing-impaired individuals and normal users, as deaf and mute individuals primarily rely on sign language for interaction. However, most people are not familiar with sign language, communication which gap creates in a everyday environments such as education, healthcare, workplaces, and public services. This gap often leads to misunderstandings, delays, and social exclusion. Traditional communication methods depend on human interpreters, which are not always available and can be costly and inefficient for real-time communication. To address these challenges, this paper proposes a Speech and Text Conversion System for Sign Language using Machine Learning. The system aims to provide a two-way communication platform by converting sign language gestures into text and speech, and vice versa. The proposed solution leverages advancements in Machine Learning, Computer Vision, and Speech Processing to ensure real-time, accurate, and efficient communication.

II. LITERATURE SURVEY

Y. Han et al. (2025) proposed a deep learning-based sign language recognition system combining Spatial-Temporal Graph Convolutional Networks (ST-GCN) and Long Short-Term Memory (LSTM). The model captures both spatial and temporal features using a dual-stream architecture, where phonological features are extracted via Media Pipe key points and temporal patterns are modeled using SA-Conv LSTM. The fusion of both streams achieved high accuracy (up to 95.2%), outperforming traditional CNN and RNN approaches. Asma Khan et al. (2025) reviewed deep learning techniques for Continuous Sign Language Recognition (CSLR), highlighting models such as CNN, LSTM, Transformers, and Graph Convolutional Networks. The study emphasized challenges like real-time processing, signer variability, and limited datasets, concluding that integrated spatial-temporal frameworks are essential for improved performance.

| ISSN: 2320-0081 | www.ijctece.com ||A Peer-Reviewed, Refereed and Bimonthly Journal |

|| Volume 9, Issue 3, May- June 2026 ||

DOI: 10.15680/IJCTECE.2026.0903003

G. Batnasan et al. (2025) introduced a Head and Hands Tunneling (H&HT) method using Media Pipe and Vision Transformers to focus on critical regions such as hands and face. This region-based approach improved recognition accuracy on datasets like WLASL2000 by preserving fine-grained gesture details.

H. Ranjbar and A. Taheri (2025) developed the ISLR101 dataset for Iranian Sign Language, consisting of 4,614 video samples. Baseline models showed that visual-based approaches achieved higher accuracy (97.01%) compared to skeleton-based methods, highlighting the importance of rich visual features.

M. Geetha et al. (2025) proposed Sign Flow, a real-time Indian Sign Language recognition system using a combination of 3D CNN and Transformer models. The system effectively processes continuous gestures using Connectionist Temporal Classification (CTC), achieving a Word Error Rate (WER) of 19%.

III. PROBLEM STATEMENT

The communication gap between hearing-impaired individuals and the general population remains a significant challenge due to the lack of widespread understanding of sign language. Existing solutions are often limited by low accuracy, lack of real-time interaction, and minimal support for two-way communication between speech and sign language. Additionally, many systems fail to capture dynamic gestures, facial expressions, and contextual meaning, which are essential for effective communication. There is a need for an efficient, real-time, and user-friendly system that can accurately translate sign language into text or speech and convert spoken language into sign representations. Such a solution would improve accessibility, inclusivity, and independent communication for hearing-impaired individuals.

IV. RESERCH METHODOLOGY

The proposed Speech and Text Conversion System for Sign Language using Machine Learning effectively addresses the communication gap between hearing-impaired individuals and normal users by enabling real-time, two-way interaction. By integrating advanced technologies such as computer vision, deep learning models (CNN and LSTM), and speech processing techniques including Speech-to-Text and Text-to-Speech, the system achieves accurate recognition and seamless translation of sign language gestures into text and speech, and vice versa. The implementation significantly improves accessibility, reduces dependency on human interpreters, and enhances communication efficiency in real-world applications such as education, healthcare, and public services. Despite these advancements, challenges such as limited dataset availability, variability in gesture patterns, sensitivity to environmental conditions, and computational complexity still impact overall system performance. Future improvements can focus on enhancing model generalization, expanding multilingual sign language support, and optimizing the system for real-time performance on low-resource devices.

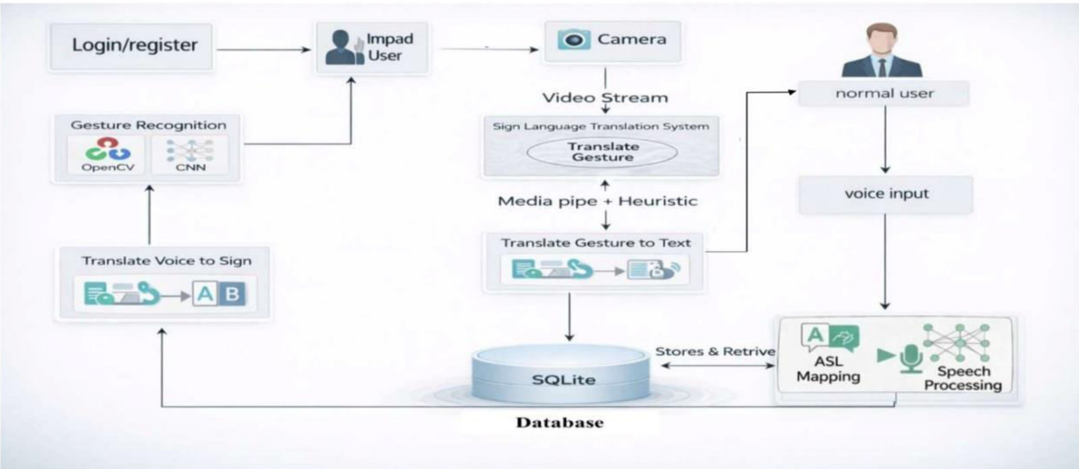

Fig 1 system architecture

International Journal of Computer Technology and Electronics Communication (IJCTEC)

| ISSN: 2320-0081 | www.ijctece.com ||A Peer-Reviewed, Refereed and Bimonthly Journal |

|| Volume 9, Issue 3, May- June 2026 ||

DOI: 10.15680/IJCTECE.2026.0903003

Authentication & Register Module

The Authentication & Registration Module enables users to securely create and access accounts within the system. It handles user registration by collecting and storing essential details, and performs authentication by verifying login credentials. The module ensures security through password encryption, input validation, and session management, while also providing features like error handling and password recovery to enhance usability and protect user data.

Visual Perception & Hand Tracking

The Visual Perception & Hand Tracking module captures real-time video input and processes it to detect and track hand movements. It identifies hand regions, extracts key landmarks such as fingers and joints, and continuously tracks their motion across frames. This processed visual data is then used for accurate gesture recognition

Gesture Interpretation & Translation

The Gesture Interpretation & Translation module analyses the tracked hand movements and extracts meaningful features to recognize specific gestures. Using trained models, it classifies these gestures into predefined signs and converts them into readable text and/or speech, enabling effective communication.

V. RESULT

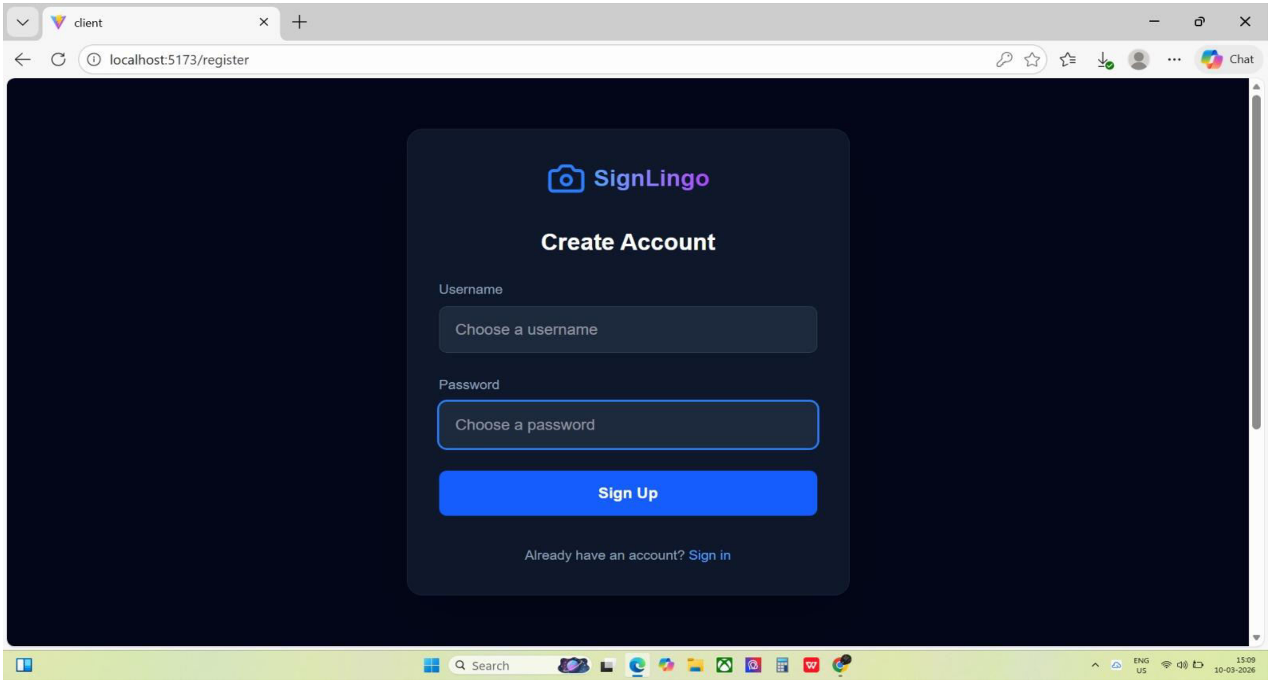

Fig 2 login page

This Fig 2 presents the user registration interface of the *Sign Lingo* system, designed to enable new users to create an account for accessing the platform's sign language translation services.

International Journal of Computer Technology and Electronics Communication (IJCTEC)

| ISSN: 2320-0081 | www.ijctece.com ||A Peer-Reviewed, Refereed and Bimonthly Journal |

|| Volume 9, Issue 3, May- June 2026 ||

DOI: 10.15680/IJCTECE.2026.0903003

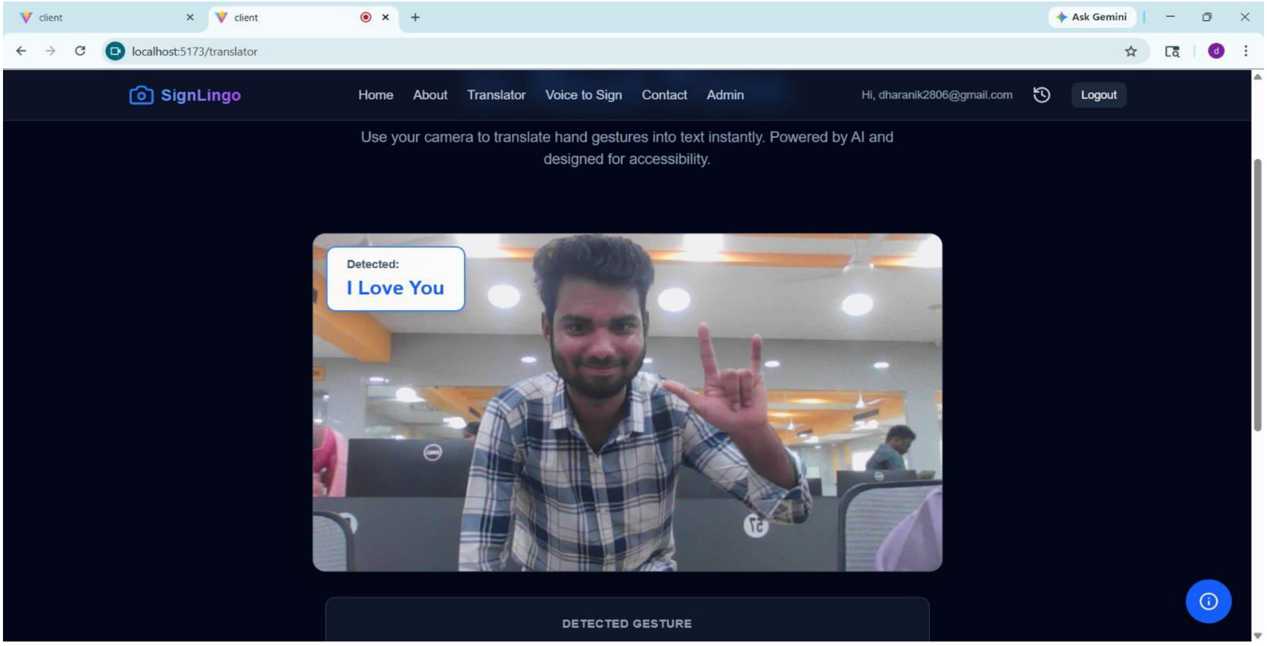

Fig 3 Translating gesture

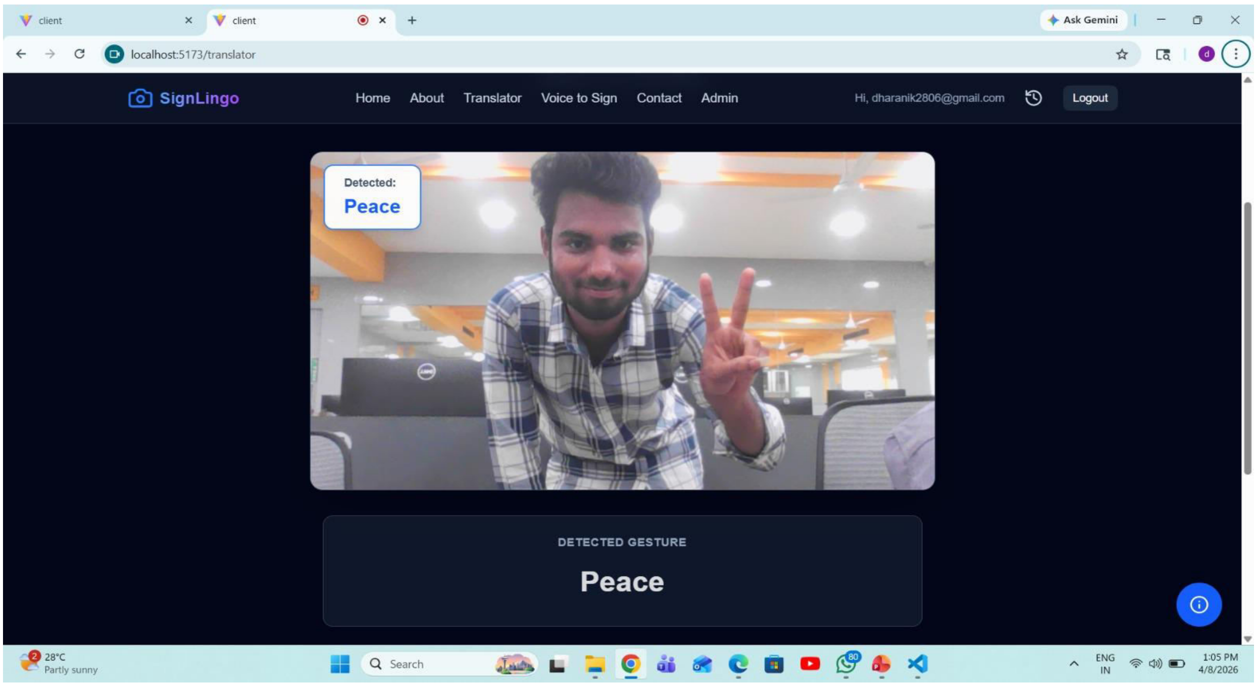

Fig 4 Translating gesture

This Fig 3&4 shows the Sign Lingo real-time gesture recognition interface. The system uses a webcam to capture hand gestures and displays the detected sign as “Peace” on the screen. The result is highlighted both on the video feed and below as “Detected Gesture: Peace,” demonstrating successful real-time sign language recognition.

| ISSN: 2320-0081 | www.ijctece.com ||A Peer-Reviewed, Refereed and Bimonthly Journal |

|| Volume 9, Issue 3, May- June 2026 ||

DOI: 10.15680/IJCTECE.2026.0903003

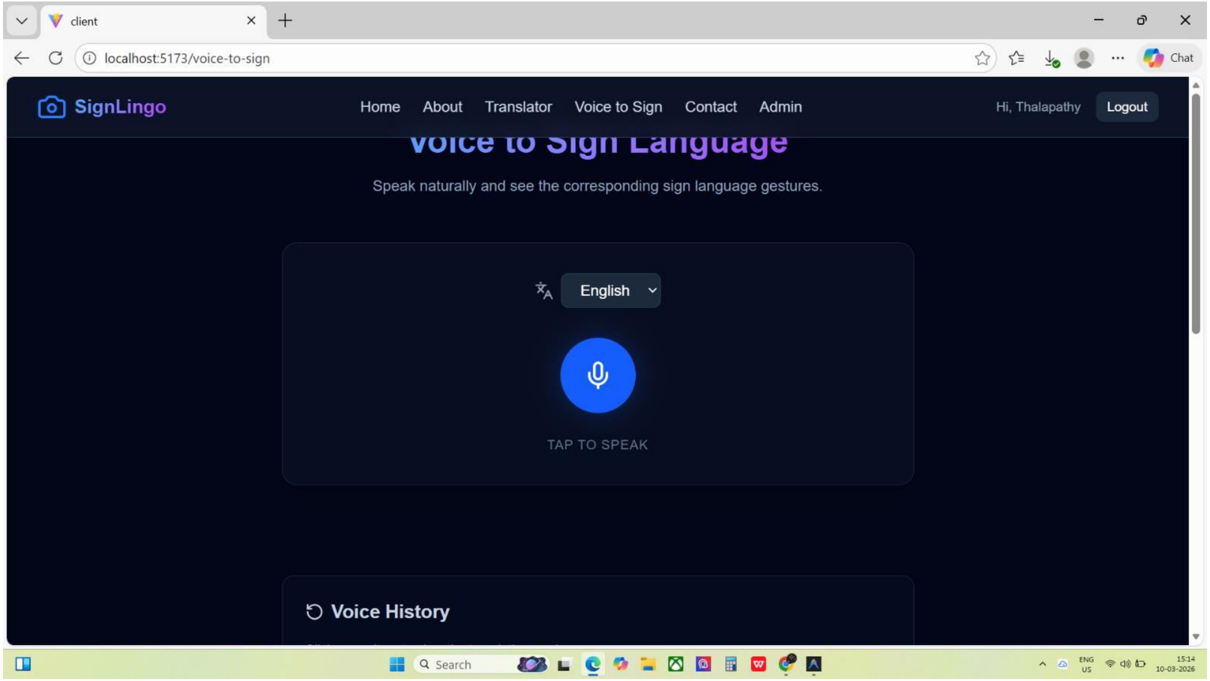

Fig 5 Voice to sign language

This Fig 5 illustrates the SignLingo voice-to-sign interface, which converts spoken language into corresponding sign language gestures. The interface includes a language selection option and a central microphone button that allows users to record speech. Upon activation, the system processes the input and generates the equivalent sign output. The design is simple and interactive, enabling users to communicate through voice-based input effectively.

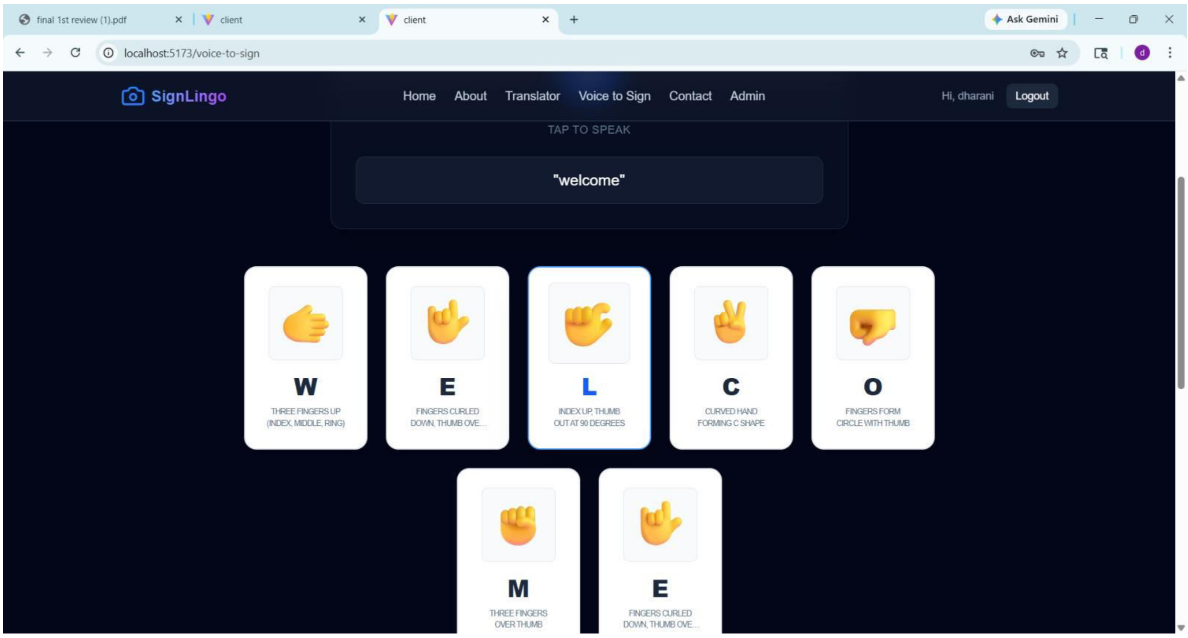

Fig 6 Translating voice to sign language

This Fig 6 shows the Sign Lingo voice-to-sign output interface. After the user speaks the word "welcome," the system converts it into sign language by breaking it into individual letters. Each letter (W, E, L, C, O, M, E) is displayed with its

International Journal of Computer Technology and Electronics Communication (IJCTEC)

| ISSN: 2320-0081 | www.ijctece.com ||A Peer-Reviewed, Refereed and Bimonthly Journal |

|| Volume 9, Issue 3, May- June 2026 ||

DOI: 10.15680/IJCTECE.2026.0903003

corresponding hand gesture, demonstrating the system's ability to translate speech into a sequence of sign representations.

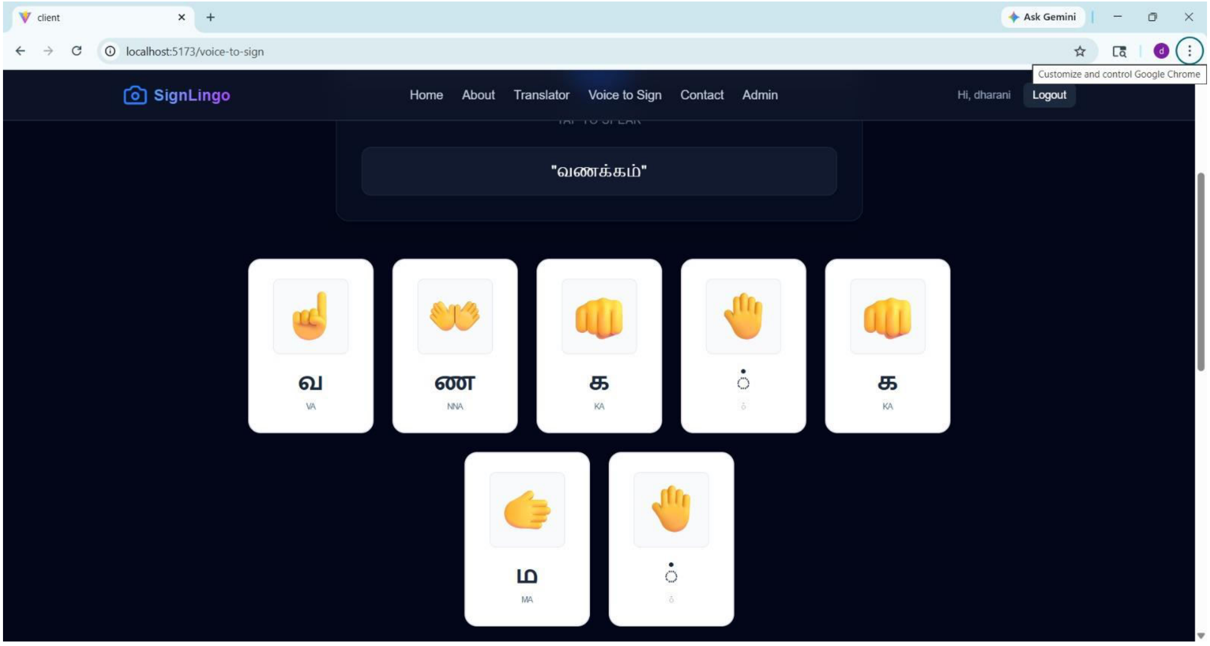

Fig 7 Translating voice to sign language

The Fig 7 shows the Voice-to-Sign translation interface of the proposed Sign Lingo system. The user provides a Tamil word as input, and the system converts it into corresponding sign representations. The word "வணக்கம்" is segmented into individual Tamil characters, and each character is displayed with its related hand sign and phonetic notation. This module helps hearing and speech-impaired individuals understand spoken or written Tamil through visual sign representations

VI. CONCLUSION & FUTURE ENHANCEMENT

The proposed Speech and Text Conversion System for Sign Language using Machine Learning effectively addresses the communication gap between hearing-impaired individuals and normal users by enabling real-time, two-way interaction. By integrating advanced technologies such as computer vision, deep learning models (CNN and LSTM), and speech processing techniques including Speech-to-Text and Text-to-Speech, the system achieves accurate recognition and seamless translation of sign language gestures into text and speech, and vice versa. The implementation significantly improves accessibility, reduces dependency on human interpreters, and enhances communication efficiency in real-world applications such as education, healthcare, and public services. Despite these advancements, challenges such as limited dataset availability, variability in gesture patterns, sensitivity to environmental conditions, and computational complexity still impact overall system performance. Future improvements can focus on enhancing model generalization, expanding multilingual sign language support, and optimizing the system for real-time performance on low-resource devices. Overall, the proposed system represents a meaningful step toward developing intelligent, scalable, and inclusive assistive communication solutions for the hearing-impaired community.

Future enhancements of the system can focus on expanding its capabilities and improving overall performance. One key improvement is the integration of multi-language sign support, including ISL, ASL, and other regional sign languages, to increase accessibility for diverse users. Incorporating advanced deep learning models such as Transformers and hybrid architectures can further enhance the accuracy of continuous gesture recognition and enable sentence-level translation. Additionally, expanding and refining the dataset with greater diversity and quality will improve model generalization across different environments. Optimizing the system for real-time performance with reduced latency will ensure smoother communication, even on low-end devices like smartphones. Furthermore, integrating facial expression and body gesture recognition can provide better contextual understanding and capture the emotional aspects of sign language communication.

| ISSN: 2320-0081 | www.ijctece.com ||A Peer-Reviewed, Refereed and Bimonthly Journal |

|| Volume 9, Issue 3, May- June 2026 ||

DOI: 10.15680/IJCTECE.2026.0903003

REFERENCES

1. Y. Han, Y. Han, and Q. Jiang, "Sign Language Recognition Using ST-GCN and LSTM with Dual-Stream Architecture," IEEE Access, vol. 13, pp. 12345–12360, Jan. 2025.
2. A. Khan, et al., "Deep Learning Approaches for Continuous Sign Language Recognition: A Comprehensive Review," IEEE Transactions on Neural Networks and Learning Systems, vol. 36, no. 2, pp. 890–905, Feb. 2025.
3. G. Batnasan, et al., "Head and Hands Tunnelling Pipeline for Improved Sign Language Recognition Using Vision Transformers," Pattern Recognition Letters, vol. 178, pp. 45–52, Mar. 2025.
4. H. Ranjbar and A. Taheri, "ISLR101: A Word-Level Iranian Sign Language Dataset for Scalable Recognition Systems," IEEE Sensors Journal, vol. 25, no. 4, pp. 3210–3219, Apr. 2025.
5. M. Geetha, et al., "Sign Flow: Real-Time Continuous Indian Sign Language Recognition Using Multi-Modal Deep Learning," Journal of Visual Communication and Image Representation, vol. 98, pp. 104012, Jan. 2025.
6. N. C. Campos, S. Hadfield, O. Koller, H. Ney, and R. Bowden, "Neural Sign Language Translation," Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, pp. 7771–7780, June 2018.
7. S. Hochreiter and J. Schmid Huber, "Long Short-Term Memory," Neural Computation, vol. 9, no. 8, pp. 1735–1780, Nov. 1997.
8. Y. LeCun, L. Botto, Y. Bengio, and P. Haffner, "Gradient-Based Learning Applied to Document Recognition," Proceedings of the IEEE, vol. 86, no. 11, pp. 2278–2324, Nov. 1998.
9. C. Lagares, et al., "Media Pipe: A Framework for Building Perception Pipelines," arrive preprint, arXiv:1906.08172, June 2019.
10. A. Graves and N. Jaitly, "Towards End-to-End Speech Recognition with Recurrent Neural Networks," Proceedings of the 31st International Conference on Machine Learning (ICML), Beijing, China, pp. 1764–1772, June 2014.
11. Y. Wang, et al., "Taco Tron: Towards End-to-End Speech Synthesis," Proceedings of Interspace, Stockholm, Sweden, pp. 4006–4010, Aug. 2017.
12. J. Devlin, M. Chang, K. Lee, and K. Toutanova, "BERT: Pre-Training of Deep Bidirectional Transformers for Language Understanding," *Proceedings of NAACL-HLT*, Minneapolis, MN, USA, pp. 4171–4186, June 2019.
13. G. Bradski, "The OpenCV Library," *Dr. Dobb's Journal of Software Tools*, vol. 25, pp. 120–125, Nov. 2000.
14. S. Albanie, et al., "BSL-1K: Scaling Up Co-Articulated Sign Language Recognition Using mouthing Cues," *Proceedings of the European Conference on Computer Vision (ECCV)*, Glasgow, UK, pp. 35–52, Aug. 2020.
15. O. Koller, H. Ney, and R. Bowden, "Deep Learning of Mouth Shapes for Sign Language," *Proceedings of the IEEE International Conference on Computer Vision Workshops (ICCVW)*, Santiago, Chile, pp. 85–91, Dec. 2015.
16. Seedha Devi, V., Selvi, D., Uma Maheshwari, K., & Yuvashree, G. (2026). Food linker: A smart system for global waste reduction. International Journal of Engineering & Extended Technologies Research (IJEETR), 8(3), 5012–5021. https://doi.org/10.15662/IJEETR.2026.0803002
17. Mathew, A., & Romasco, L. (2024). Forensic Investigation of Artificial Intelligence Systems. Research Updates in Mathematics and Computer Science Vol. 4, 154-164.
18. Sugumar, R. (2025). Designing Resilient and Scalable Cloud-Native Frameworks for Generative AI Content Production. International Journal of Research Publications in Engineering, Technology and Management (IJRPETM), 8(6), 13268-13279.
19. Rajasekar, M., Celine Kavida, A., & Anto Bennet, M. (2020). A pattern analysis based underwater video segmentation system for target object detection. Multidimensional Systems and Signal Processing, 31(4), 1579-1602.
20. Seedha Devi, V., Namitha, B., Divya Dharshini, J., & Livetha, K. (2026). A hybrid biometric and geo-fencing based smart attendance system. International Journal of Advanced Research in Computer Science and Technology (IJARCST), 9(3), 794–802. https://doi.org/10.15662/IJARCST.2026.0903002
21. Poornima, G., & Anand, L. (2025). Medical image fusion model using CT and MRI images based on dual scale weighted fusion based residual attention network with encoder-decoder architecture. Biomedical Signal Processing and Control, 108, 107932.
22. Vani, S., Malathi, P., Ramya, V. J., Sriman, B., Saravanan, M., & Srivel, R. (2024). An efficient black widow optimization-based faster R-CNN for classification of COVID-19 from CT images. Multimedia Systems, 30(2), 108.
23. Narayanan, L. K., Loganayagi, S., Hemavathi, R., Jayalakshmi, D., & Vimal, V. R. (2024, March). Machine learning-based predictive maintenance for industrial equipment optimization. In 2024 International Conference on Trends in Quantum Computing and Emerging Business Technologies (pp. 1-5). IEEE.

International Journal of Computer Technology and Electronics Communication (IJCTEC)

| ISSN: 2320-0081 | www.ijctece.com ||A Peer-Reviewed, Refereed and Bimonthly Journal |

|| Volume 9, Issue 3, May- June 2026 ||

DOI: 10.15680/IJCTECE.2026.0903003

24. Alangaram, S., Praveen, S., Rajesh, V., & Sanjai, A. (2026). Sales guard AI-driven decision intelligence platform for business optimization. International Journal of Engineering & Extended Technologies Research (IJEETR), 8(3), 5022–5031. https://doi.org/10.15662/IJEETR.2026.0803003
25. Sharma, K. P., Kumar, I., Singh, P. P., Anbazhagan, K., Albarakati, H. M., Bhatt, M. W., ... & Rana, A. (2024). Advancing spacecraft rendezvous and docking through safety reinforcement learning and ubiquitous learning principles. Computers in Human Behavior, 153, 108110.
26. Seedha Devi, V., Harshini, R., Dhana Lakshmi, E., Gayathri, N., & Nithesha, P. (2026). Low-code mobile application builder with AI-assisted features using Flutter & Firebase. International Journal of Research Publications in Engineering, Technology and Management (IJRPETM), 9(3), 1001–1010.
27. Socrates, S., Shanmugapriya, M., Murugeshwari, B., & Angalaeswari, S. (2024). Efficient Design for Implantable Device Constant Current Induction Doubly Fed Generating Incorporating Grid Connectivity. In Intelligent Solutions for Sustainable Power Grids (pp. 382-392). IGI Global Scientific Publishing.
28. Seedha Devi, V., Harshini, R., Dhana Lakshmi, E., Gayathri, N., & Nithesha, P. (2026). Low-code mobile application builder with AI-assisted features using Flutter & Firebase. International Journal of Research Publications in Engineering, Technology and Management (IJRPETM), 9(3), 1001–1010. https://doi.org/10.15662/IJRPETM.2026.0903001
29. Deivendran, P., Babu, P. S., Malathi, G., Anbazhagan, K., & Kumar, R. S. (2023). Emotion Recognition for Challenged People Facial Appearance in Social using Neural Network. arXiv preprint arXiv:2305.06842.
30. Seedha Devi, V., Divya Narasimman, S., Jayaprakash, S., & Mohamed Suhel, H. N. (2026). Smart IoT-based pedestrian power generator using DC motor. International Journal of Computer Technology and Electronics Communication (IJCTEC), 9(3), 990–999. https://doi.org/10.15680/IJCTECE.2026.0903002
31. Vimal, V. R., Anandan, P., & Induja, V. (2024). Estimating the perspicacious features of ECG recording based on template classification for detecting atrial fibrillation. International Journal of Advanced Intelligence Paradigms, 29(1), 17-27.
32. Archana, R., & Anand, L. (2025). Residual u-net with Self-Attention based deep convolutional adaptive capsule network for liver cancer segmentation and classification. Biomedical Signal Processing and Control, 105, 107665.
33. Kaliappan, S., Ragunthar, T., Ali, M., & Murugeshwari, B. (2024). Implementation of Virtual High Speed Data Transfer in Satellite Communication Systems Using PLC and Cloud Computing. In AI Approaches to Smart and Sustainable Power Systems (pp. 274-286). IGI Global Scientific Publishing.
34. Seedha Devi, V., Mahalakshimi, P. V., & Anitha, A. (2026). Automated skin disease analysis and detection using AI-powered mobile application. International Journal of Research and Applied Innovations (IJRAI), 9(3), 531–539. https://doi.org/10.15662/IJRAI.2026.0903004
35. Mathew, D. A. (2024). Time-triggered ethernet (ttethernet) and artificial Intelligence. International Journal of Development Research, 14.
36. Revathi, K. G., Ananth, B. J., Saravanan, M. L., & Kumar, A. R. (2021). Gps enabled vehicle location identification using gsm and fare collection using smart card. Turkish journal of computer and mathematics education, 12(10), 2657-2668.
37. Sugumar, R. (2025). Federated AI in Offline-First Mobile Health Architectures for Privacy-Preserving Clinical Intelligence. International Journal of Science, Research and Technology, 8(4), 14589-14600.
38. Rajasekar, M., Aruldoss, A. C., & Bennet, M. A. (2018). A novel method to detect corrosion in underwater infrastructure using an image processing. ARPN Journal of Engineering and Applied Science, 13(7), 2556-2561.
39. Alangaram, S., Udaykiran, M., Rajkumar, K., & Yogeeswaran, T. (2026). Enhancing customer churn prediction and retention for e-commerce. International Journal of Advanced Research in Computer Science & Technology (IJARCST), 9(3), 803–813. https://doi.org/10.15662/IJARCST.2026.0903003