International Journal of Scientific Research in Engineering & Technology
Volume6, Issue6 (November- December2026), PP: 56-61.
https://www.doi.org/10.59256/ijsreat.202606060009
www.ijsreat.com
ISSN No: 2583-1240

AI-Based Sign Language Recognition and Translation System

Dr Anitha G[1], Rakshita A P[2], Sanjana R C[3], Varshini K S[4], Vishakha P G[5]
[1]*Professor, Department of Information Science and Engineering, Bapuji Institute of Engineering and Technology, Davangere, Karnataka, India.*
[2,3,4,5]*Department of Information Science and Engineering, Bapuji Institute of Engineering and Technology, Davangere, Karnataka, India.*

To Cite this Article: *Dr Anitha G[1], Rakshita A P[2], Sanjana R C[3], Varshini K S[4], Vishakha P G[5], " AI-Based Sign Language Recognition and Translation System", International Journal of Scientific Research in Engineering & Technology, Volume 06, Issue 06, November-December2026, PP:56-61.*

Copyright: ©2026 This is an open access journal, and articles are distributed under the terms of the Creative Commons Attribution License; Which Permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Abstract: *The AI-Based Sign Language Recognition and Translation System is a computer vision-based application designed to facilitate communication between individuals with hearing and speech impairments and people who do not understand sign language. The system captures hand gestures through a webcam and uses OpenCV and MediaPipe to extract hand landmarks for gesture recognition. A machine learning model processes the extracted features to identify corresponding English alphabet signs and display the recognized letters as text through a Streamlit-based user interface. The system supports both single-hand and two-hand gesture recognition based on the trained datasets. Text-to-speech technology is used to convert recognized text into audible speech, while speech recognition enables spoken input to be converted into text for reverse communication. The system also incorporates multilingual translation and an Emergency Mode to support communication in different situations. Python serves as the primary programming language, while OpenCV, MediaPipe, NumPy, and Joblib support image processing, landmark extraction, and model prediction. The proposed system aims to reduce communication barriers and provide an accessible platform for sign language-based interaction. However, recognition performance depends on lighting conditions, camera quality, hand positioning, and variations in gestures. Further testing with diverse users and real-world conditions is required to evaluate the system's generalization and reliability.*

Keywords: *Sign Language Recognition, Gesture Recognition, Computer Vision, Machine Learning, MediaPipe, OpenCV, Text-to-Speech, Speech Recognition, Multilingual Translation.*

I. INTRODUCTION

Communication is an essential part of everyday life, enabling individuals to express their thoughts, needs, and emotions. However, deaf and mute individuals often face communication challenges when interacting with people who do not understand sign language. These barriers can affect daily activities, education, healthcare, and social interactions. Traditional communication methods may require the assistance of sign language interpreters, who may not always be available. Therefore, an intelligent system that can recognize sign language gestures and convert them into understandable text and speech can help improve communication accessibility.

The proposed AI-Based Sign Language Recognition and Translation System is a computer vision-based application developed using Python, OpenCV, and MediaPipe to recognize hand gestures in real time. The system captures webcam frames, detects hand landmarks, and uses a trained machine learning model to classify gestures into letters and words. Generative AI processes the recognized words into meaningful sentences, while multilingual translation, text-to-speech, and speech recognition support communication in both directions. Previous research has explored deep learning approaches to sign language processing [2], while transformer-based approaches integrate sign language recognition and translation [8].

The main objective of the system is to reduce communication barriers for individuals with hearing and speech impairments. It combines gesture recognition, sentence generation, translation, and speech processing through a Streamlit-based interface. An Emergency Mode supports urgent communication by detecting HELP or SOS gestures, obtaining the user's location, sending alerts to predefined contacts through WhatsApp, SMS, or Email, and activating a siren.

Furthermore, the system supports more inclusive communication by combining computer vision and artificial intelligence with speech technologies. The gesture recognition module identifies hand movements, while Generative AI processes recognized words to form meaningful sentences. Multilingual translation helps users communicate across selected languages, and reverse speech-to-text functionality allows hearing individuals to communicate with sign language users through text. However, the reliability of the system depends on factors such as lighting conditions, camera quality, hand positioning, variations in gestures, and the availability of external AI and communication services. Further testing with diverse users and real-world conditions is necessary to evaluate its performance and usability. This approach provides a foundation for developing accessible communication tools that promote independence, inclusivity, and safety for individuals with hearing and speech impairments.

II. MATERIALS AND METHODS

The proposed system follows a sequence of steps, beginning with webcam-based gesture capture and ending with text, speech, or translated output.

Gesture Capture:

The webcam captures the user's hand gestures in real time. The video frames serve as input for detecting and recognizing sign language gestures.

Image Processing and Landmark Extraction:

OpenCV processes the captured video frames, while MediaPipe detects the hands and extracts their landmark coordinates. These landmarks represent hand structure and provide features for gesture classification.

Gesture Recognition:

The extracted landmark features are passed to a trained machine learning model to identify the corresponding sign language letters or words. The recognition process supports single-hand and two-hand gestures represented in the training dataset.

Sentence Generation:

The recognized words are processed using Generative AI to generate meaningful and grammatically correct sentences. This helps convert individual recognized signs into more understandable text.

Translation and Speech Processing:

The generated text can be translated into selected languages. Text-to-speech technology converts the text into audible speech, while speech recognition converts spoken input into text for reverse communication.

Emergency Assistance:

The system includes an Emergency Mode to support urgent communication. When an emergency is triggered through the supported gesture or interface, the system can obtain the user's location, send alerts to predefined contacts through configured communication channels, activate a siren, and display an emergency notification.

Output Display:

The generated sentences, translated text, and available speech outputs are presented through the Streamlit interface, allowing users to access the system's communication features.

Materials and Technologies Used

The proposed system is developed using Python and Streamlit to provide an interactive interface for sign language recognition and translation. OpenCV and MediaPipe support video processing and hand landmark extraction, while a trained machine learning model classifies gestures. Generative AI supports sentence generation and multilingual translation. gTTS provides text-to-speech functionality, and SpeechRecognition supports reverse communication. Additional modules manage emergency contacts, location information, notifications, and siren activation.

Methodology

1. Input Collection: The webcam captures hand gestures, while a microphone supports spoken input for reverse communication.
2. Image Processing: OpenCV processes video frames, and MediaPipe detects hands and extracts landmark coordinates.
3. Gesture Classification: The extracted features are passed to the trained machine learning model to recognize the corresponding sign language gestures.
4. Sentence Generation: Recognized letters and words are processed using Generative AI to generate meaningful sentences.
5. Translation and Speech Output: The generated text is translated into a selected language when required and converted into speech using text-to-speech technology.
6. Reverse Communication: Speech recognition converts spoken input into text, supporting communication between sign language users and other individuals.
7. Emergency Management: The system supports emergency alerts, location sharing, and siren activation through its configured emergency features.
8. Output Display: The final text, translated content, speech output, and emergency notifications are presented through the Streamlit interface.

III. SYSTEM ARCHITECTURE

The proposed system consists of a gesture input module, image processing module, gesture recognition module, sentence generation module, translation and speech processing module, emergency assistance module, and output interface.

The input module captures hand gestures through a webcam, while OpenCV and MediaPipe process video frames and extract hand landmark coordinates. The gesture recognition module uses a trained machine learning model to classify the detected gestures into corresponding letters or words. The Generative AI module converts recognized words into meaningful

and grammatically correct sentences. The translation and speech processing module supports multilingual translation, text-to-speech conversion, and speech recognition for reverse communication. The emergency assistance module supports emergency alerts, location sharing, and siren activation through configured services. Finally, the Streamlit interface displays recognized text, generated sentences, translated content, and emergency notifications, providing an integrated platform for accessible communication.

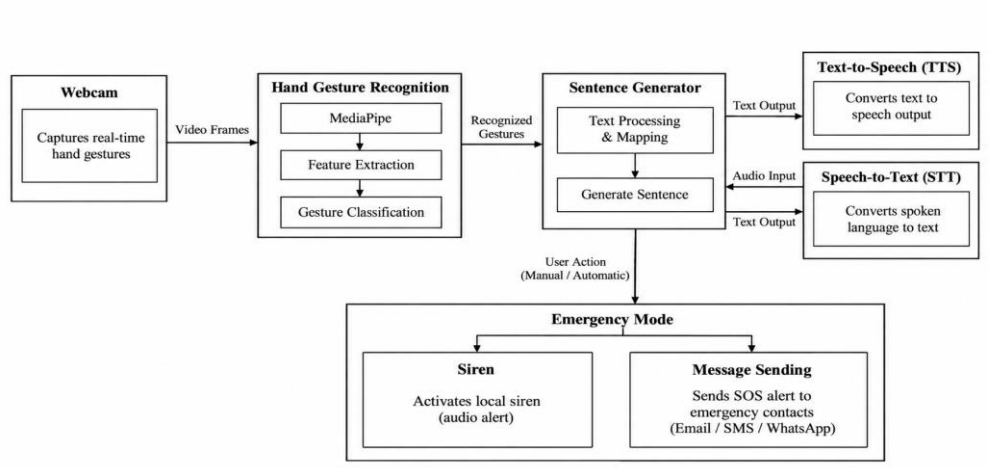


Fig.1 System Architecture of AI-Based Sign Language Recognition and Translation System

IV. RESULT AND DISCUSSION

The AI-Based Sign Language Recognition and Translation System is designed to improve communication between individuals with hearing and speech impairments and people who do not understand sign language. The system uses Artificial Intelligence (AI), computer vision, and machine learning to recognize hand gestures and convert them into meaningful text and speech.

The system captures hand gestures through a webcam and processes the video frames using OpenCV and MediaPipe to detect hands and extract landmark coordinates. A trained machine learning model classifies the gestures into corresponding letters or words. The recognized words are then processed using Generative AI to generate meaningful and grammatically correct sentences. The system also supports multilingual translation, text-to-speech conversion, and speech recognition for reverse communication.

The system integrates gesture recognition, sentence generation, translation, and speech processing into a single workflow through a Streamlit-based interface. In addition, the Emergency Mode supports urgent communication by detecting supported HELP or SOS gestures or interface activation, obtaining the user's location, sending alerts to predefined contacts through configured Wha tsApp, SMS, or Email services, and activating a siren. These features aim to improve accessibility, support inclusive communication, and enhance user safety.

However, variations in hand positioning, lighting conditions, camera quality, and gestures may affect recognition results. Sentence generation and emergency notifications may also depend on the availability of external AI and communication services. Therefore, further testing under diverse real-world conditions is necessary to evaluate the system's reliability, usability, and effectiveness.

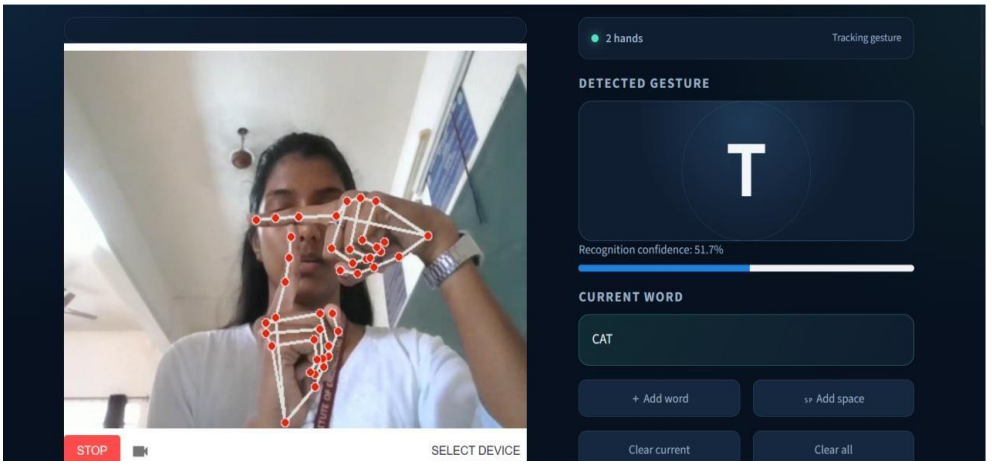


Fig. 2. Real-Time Sign Language Gesture Recognition

The figure 2 illustrates the real-time gesture recognition interface of the proposed system. The webcam captures the user's hand gestures, while MediaPipe detects and tracks hand landmarks. The trained machine learning model predicts the corresponding sign, displaying the detected letter and recognition confidence on the interface. The current word can be built by adding recognized letters and spaces, enabling users to convert sign gestures into text.

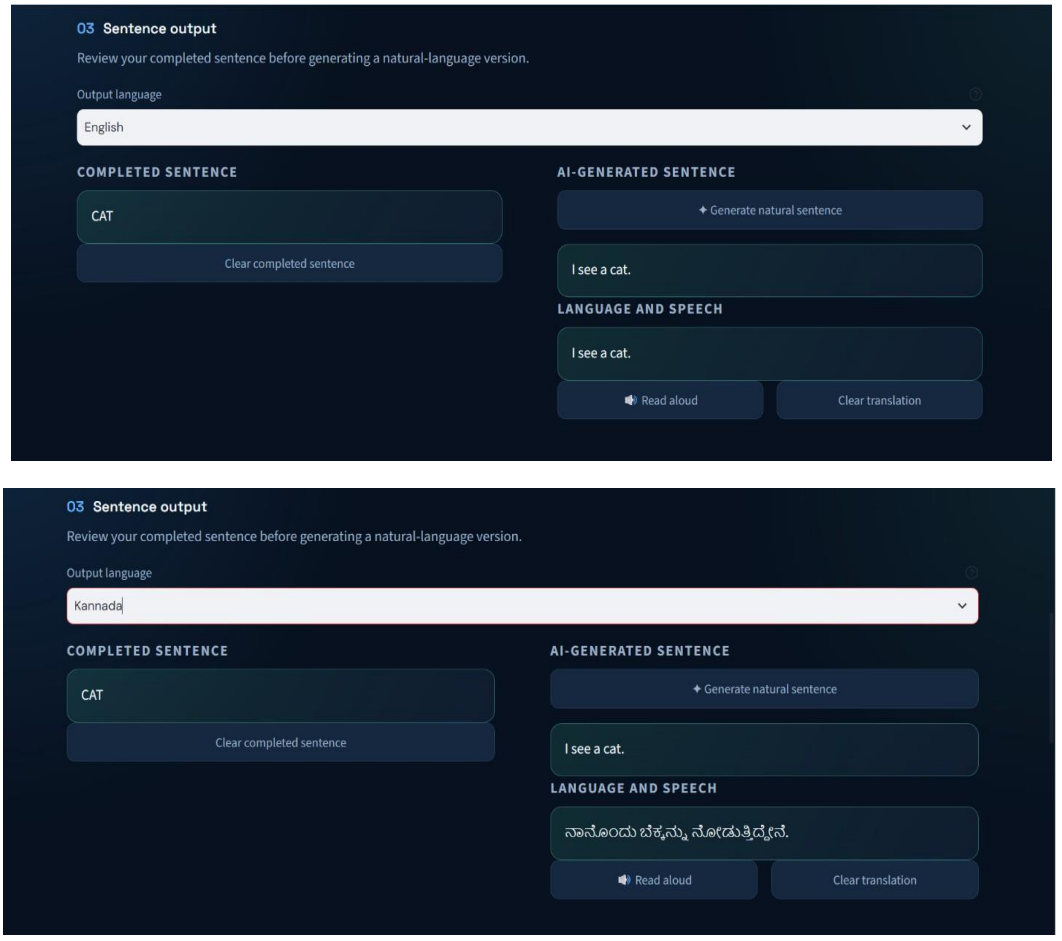


Fig. 3. AI-Based Sentence Generation and Multilingual Translation

The figure 3 illustrates the sentence generation and translation module of the proposed system. The recognized words are displayed as a completed sentence, which is processed using Generative AI to generate a meaningful and grammatically correct sentence. The system supports multiple output languages, such as English and Kannada, and displays the translated text. The text-to-speech feature enables users to listen to the generated or translated sentence, facilitating accessible and multilingual communication.

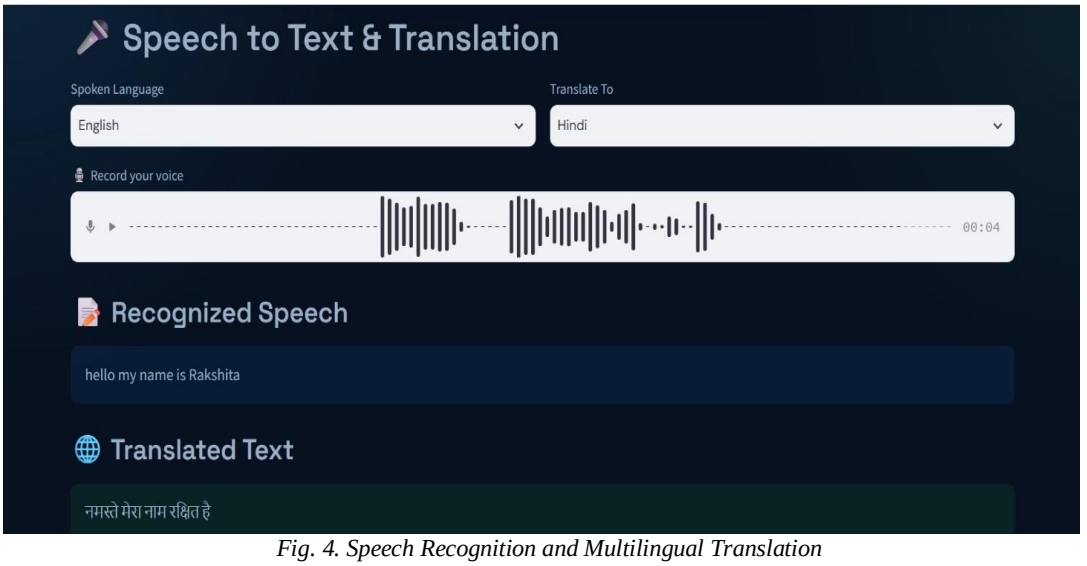


Fig. 4. Speech Recognition and Multilingual Translation

The figure 4 illustrates the speech-to-text and translation module of the proposed system. The user selects the spoken language and the target translation language, then records their voice. The system converts the recorded speech into text and displays the recognized sentence. The recognized text is then translated into the selected language, such as Hindi, supporting reverse communication and enabling users to communicate across different languages.



Fig. 5. Emergency Assistance Module

The figure 5 illustrates the Emergency Mode of the proposed system, which allows users to select an emergency condition such as Help, Danger, Robbery, Medical, Fire, Accident, Lost, Family, or Other. Based on the selected condition, the system uses Generative AI to prepare an SOS message and sends it to saved family members along with the user's location through configured communication services. This module aims to facilitate rapid emergency communication and improve user safety.

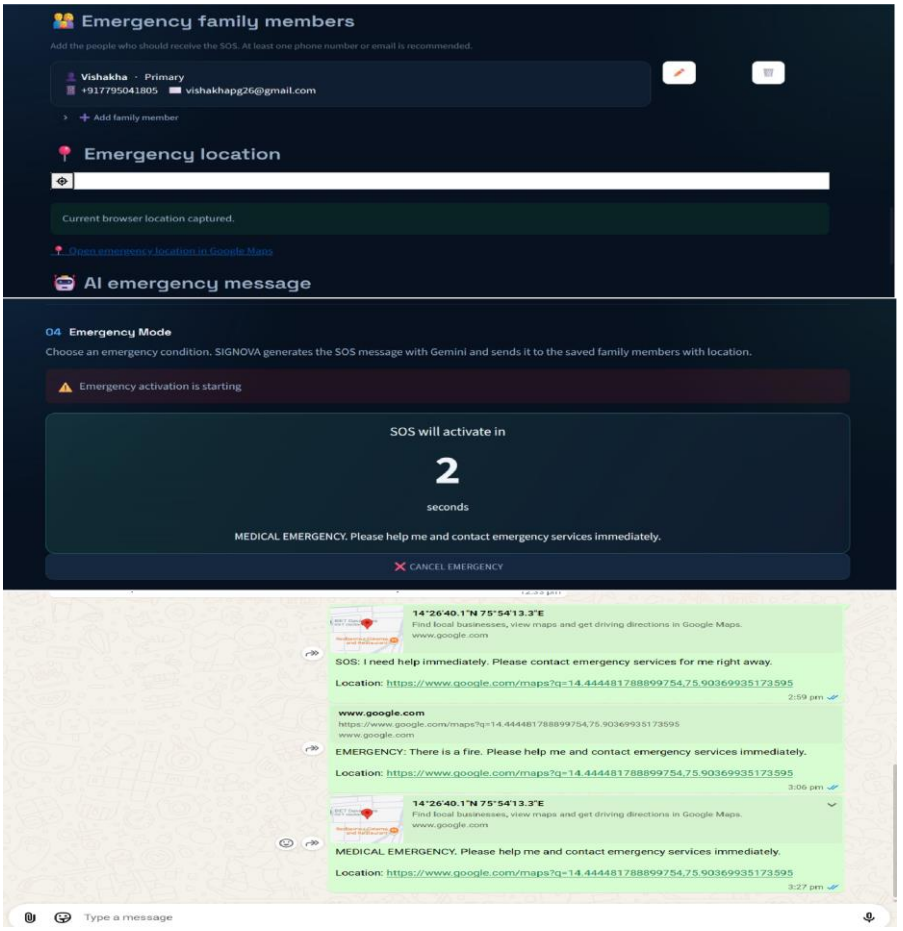


Fig. 6. Emergency Contact Management and SOS Alert System

The figure 6 illustrates the emergency assistance workflow of the proposed system. Users can manage emergency contacts and capture their current browser location, which can be shared through a Google Maps link. When an emergency is triggered, the system displays a countdown and generates an SOS message using Generative AI. After activation, the emergency message and location details are sent to the saved contacts through the configured communication service, such as WhatsApp. This feature enables faster sharing of emergency information and supports timely assistance during critical situations.

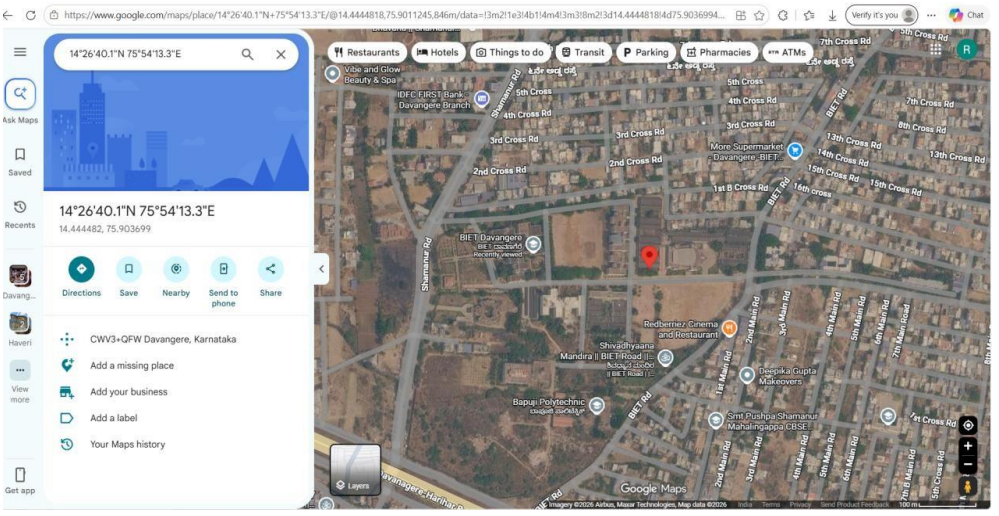


Fig. 7. Emergency Location Tracking Using Google Maps

The figure 7 shows the location-tracking feature of the proposed system using Google Maps. After the emergency mode is activated, the system captures the user's current browser location and generates a map link indicating the location coordinates. This location information can be shared with predefined emergency contacts along with the SOS message, helping them identify the user's approximate location and provide timely assistance.

V.CONCLUSION

The proposed AI-Based Sign Language Recognition and Translation System provides an integrated approach to recognizing hand gestures and converting them into meaningful text and speech using computer vision, machine learning, and Generative AI. By combining OpenCV, MediaPipe, gesture classification, sentence generation, and multilingual translation, the system supports communication between individuals with hearing and speech impairments and others. The integration of text-to-speech and speech recognition enables communication in both directions. Additionally, the Emergency Mode supports urgent communication by generating SOS messages, sharing location details with predefined contacts, and activating alerts through configured services. The Streamlit interface provides convenient access to these features within a single platform. However, the system's effectiveness depends on gesture recognition quality, environmental conditions, and the availability of external AI and communication services. Future improvements may focus on enhancing recognition reliability, expanding gesture vocabulary, improving sentence generation and translation, and conducting further testing under diverse real-world conditions.

REFERENCES

1. A. Khan, S. Jin, G.-H. Lee, G. E. Arzu, L. Minh Dang, T. N. Nguyen, W. Choi, and H. Moon, “Deep Learning Approaches for Continuous Sign Language Recognition: A Comprehensive Review,” IEEE Access, vol. 13, pp. 55524–55544, 2025. DOI: 10.1109/ACCESS.2025.3554046.
2. N. C. Camgoz, O. Koller, S. Hadfield, and R. Bowden, “Sign Language Transformers: Joint End-to-End Sign Language Recognition and Translation,” in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020, pp. 10023–10033. DOI: 10.1109/CVPR42600.2020.01004.
3. 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. 38, no. 23, pp. 1649–1667, 2024. DOI: 10.1080/01691864.2024.2442721. SQLAlchemy Authors, “SQLAlchemy Documentation.”
4. Z. Yu, S. Huang, Y. Cheng, and T. Birdal, “SignAvatars: A Large-scale 3D Sign Language Holistic Motion Dataset and Benchmark,” in Proceedings of the European Conference on Computer Vision (ECCV), 2024. OpenRouter, “OpenRouter Documentation.”
5. A. Núñez-Marcos, O. Perez-de-Viñaspre, and G. Labaka, “A Survey on Sign Language Machine Translation,” Expert Systems with Applications, vol. 213, article 118993, 2023. DOI: 10.1016/j.eswa.2022.118993. N. Reimers and I. Gurevych, “Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks,” in Proceedings of EMNLP-IJCNLP, 2019, pp. 3982–3992. DOI: 10.18653/v1/D19-1410.
6. D. D. Chakladar, P. Kumar, S. Mandal, P. P. Roy, M. Iwamura, and B.-G. Kim, “3D Avatar Approach for Continuous Sign Movement Using Speech/Text,” Applied Sciences, vol. 11, no. 8, article 3439, 2021. DOI: 10.3390/app11083439.
7. R. Rastgoo, K. Kiani, S. Escalera, and M. Sabokrou, “Sign Language Production: A Review,” in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, pp. 3451–3461, 2021. DOI: 10.1109/CVPRW53098.2021.00384.
8. [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, pp. 891–908, 2020. DOI: 10.1007/s11263-019-01281-2.