IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 14, April 2026
Impact Factor: 8.2

Real-Time Sign Language to Text Conversion Using Computer Vision

Prof. Somnath Mule[1], Sakshi Gentyal[2], Prathamesh Patil[3], Manmath Thigale[4] Govind Gore[5]
Head of Department, Computer Science and Engineering[1]
Students, Computer Science and Engineering[2,3,4,5]
MIT College of Railway Engineering & Research Barshi

Abstract: *Communication between deaf or hard-of-hearing individuals and the general population is often challenging due to the limited understanding of sign language. This communication gap creates barriers in education, employment, and social interaction. To address this issue, this research proposes an Artificial Intelligence (AI) based system that converts sign language gestures into readable text using computer vision and machine learning techniques.*

The proposed system captures hand gestures using a webcam and processes the visual data through image preprocessing and landmark detection techniques. The extracted hand landmarks are analyzed using machine learning and deep learning models to recognize sign language gestures. Once the gesture is recognized, the system converts it into the corresponding textual output in real time.The system is implemented as a web-based platform to ensure accessibility across multiple devices without requiring specialized hardware. The platform is developed using Python, OpenCV, MediaPipe, and Flask for the backend, while the frontend interface is built using HTML, CSS, and JavaScript to provide a responsive and user-friendly experience. Experimental results demonstrate that the proposed approach can effectively recognize sign language gestures and convert them into meaningful text. The system serves as an assistive technology that enhances accessibility, promotes inclusivity and helps bridge the communication gap between hearing and non-hearing individuals.

Keywords: *Sign Language Recognition, Computer Vision, Machine Learning, Gesture Recognition, Human Computer Interaction*

I. INTRODUCTION

Communication is a fundamental aspect of human interaction, enabling individuals to express thoughts, emotions, and intentions effectively. However, for people with hearing and speech impairments, communication with the wider society remains a persistent challenge. Sign language serves as a primary medium of communication for the deaf and mute community, but its understanding is often limited to those who have received specialized training. This creates a significant communication gap between sign language users and the general population, leading to social isolation and reduced accessibility in education, healthcare, and employment.With the rapid advancement of technology, particularly in the domains of computer vision and artificial intelligence, there is a growing opportunity to bridge this gap. Real-time sign language recognition systems aim to translate hand gestures and movements into readable text or audible speech, enabling seamless interaction between differently-abled individuals and others. Such systems not only enhance inclusivity but also promote independence and confidence among users.

Traditional approaches to sign language interpretation often rely on human interpreters or sensor-based gloves, which can be expensive, intrusive, and impractical for everyday use. In contrast, computer vision-based methods leverage cameras and image processing techniques to detect and interpret gestures in a non-invasive and scalable manner. By utilizing machine learning and deep learning models, these systems can learn complex patterns in hand shapes, orientations, and movements, allowing for accurate and efficient translation of sign language into text in real time.

DOI: 10.48175/IJARSCT-34070

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 6, Issue 14, April 2026 Impact Factor: 8.2

The proposed system, Real-Time Sign Language to Text Conversion Using Computer Vision, focuses on developing an intelligent, cost-effective, and user-friendly solution that captures live video input, processes hand gestures, and converts them into meaningful textual output. This approach not only reduces dependency on interpreters but also opens new possibilities for real-time communication in various environments such as classrooms, workplaces, and public services.

Furthermore, the integration of such technologies aligns with global efforts toward digital inclusion and accessibility. By combining computer vision techniques with robust classification algorithms, the system aims to achieve high accuracy, low latency, and adaptability to different lighting conditions and backgrounds. The ultimate goal is to create a reliable communication bridge that empowers individuals with hearing and speech disabilities to participate more actively in society.In summary, this research addresses a critical societal need by leveraging cutting-edge technological advancements to develop an efficient real-time sign language recognition system. It contributes to the broader vision of inclusive technology, where communication barriers are minimized, and equal opportunities are made accessible to all.

II. LITERATURE RIVIEW

Sign language recognition has gained significant attention in recent years as researchers aim to develop intelligent systems that can facilitate communication between deaf or hard-of-hearing individuals and the general population. Early approaches to sign language translation primarily relied on sensor-based devices, such as data gloves equipped with motion sensors and flex sensors. These systems were capable of capturing detailed finger movements and hand orientations, which allowed accurate gesture recognition. However, such solutions were expensive, required specialized hardware, and were not convenient for daily use. As a result, researchers began exploring vision-based approaches that use cameras and image processing techniques to detect and interpret hand gestures without requiring additional wearable devices.

With the advancement of computer vision and machine learning, vision-based sign language recognition systems have become more practical and accessible. Early vision-based systems used traditional image processing techniques such as skin color detection, edge detection, and contour analysis to identify hand gestures. These approaches attempted to segment the hand region from the background and analyze its shape and movement. Although these techniques provided a foundation for gesture recognition, their performance was often affected by environmental factors such as lighting conditions, background noise, and variations in hand position. Consequently, the accuracy and robustness of these systems were limited.

The emergence of deep learning techniques, particularly Convolutional Neural Networks (CNNs), significantly improved the performance of gesture recognition systems. CNNs are capable of automatically extracting complex visual features from images, making them highly effective for recognizing hand gestures in sign language datasets. Several studies have demonstrated that CNN-based models can achieve high accuracy in recognizing static hand gestures such as alphabet signs. Researchers have also explored advanced architectures, including 3D CNNs, Recurrent Neural Networks (RNNs), and Long Short-Term Memory (LSTM) networks, to recognize dynamic gestures and continuous sign language sentences. These deep learning approaches have enabled systems to process sequences of gestures and better understand temporal patterns in sign language communication.

In addition to gesture recognition models, researchers have also focused on developing user-friendly interfaces and accessible platforms for sign language translation systems. Many recent studies emphasize the importance of web-based and mobile-based applications that allow users to access sign language recognition tools using standard cameras available on smartphones or computers. Integrating machine learning models with modern web technologies has made it possible to create responsive platforms that can perform real-time gesture recognition and translation. Such systems provide greater accessibility compared to earlier hardware-dependent solutions.

Despite these advancements, several challenges remain in the field of sign language recognition. One of the major issues is the lack of large and diverse datasets for different sign languages, which limits the ability of models to generalize across various gestures and languages. Furthermore, many existing systems focus only on recognizing

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 14, April 2026
Impact Factor: 8.2

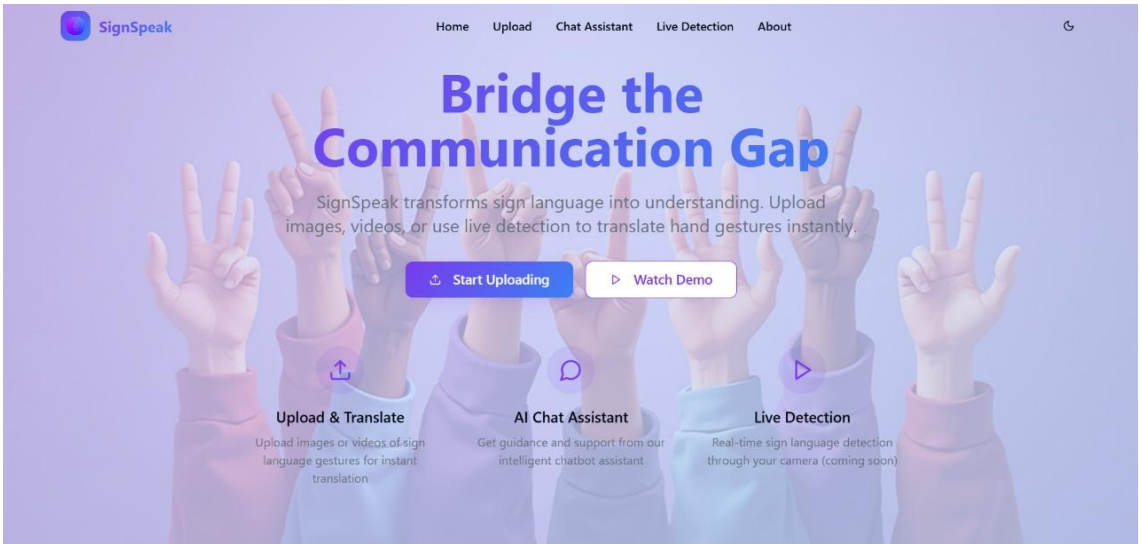
isolated gestures rather than continuous sentences, which restricts their practical use in real-life communication. Another limitation is the absence of multilingual support and educational features that could help users learn sign language effectively.

To address these challenges, recent research has focused on combining deep learning, computer vision, and web technologies to develop more robust and inclusive sign language translation systems. By leveraging these technologies, modern systems can recognize gestures more accurately, provide real-time translation, and support multiple languages. Such advancements have the potential to significantly reduce communication barriers and promote accessibility for the deaf community. The proposed system builds upon these developments by implementing an AI-based web platform capable of recognizing sign language gestures and converting them into text, thereby contributing to the advancement of assistive communication technologies.

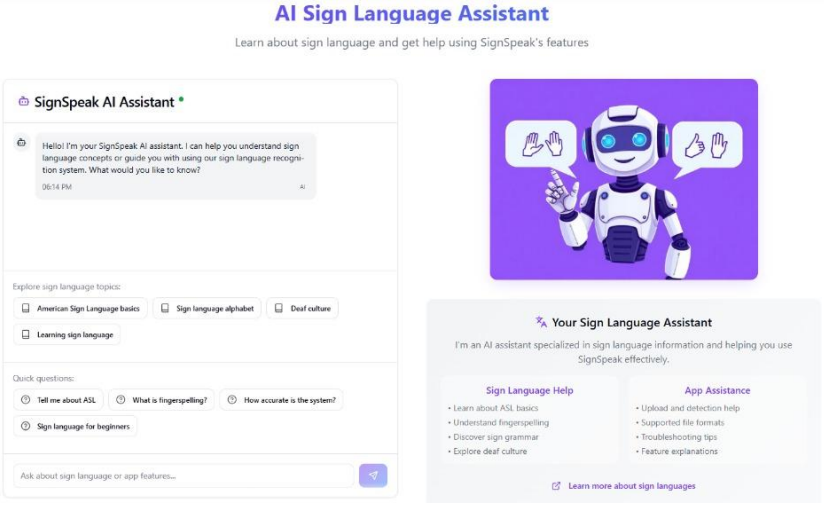
III. METHODOLOGY

The proposed system is designed to convert sign language gestures into readable text by utilizing computer vision and deep learning techniques. The methodology of the system involves several stages including data collection, preprocessing, feature extraction, model training, gesture recognition, and text generation. These stages work together to ensure accurate detection and interpretation of hand gestures in real time. The overall framework is developed as a web-based application so that users can access the system easily using a standard webcam without requiring any specialized hardware devices.

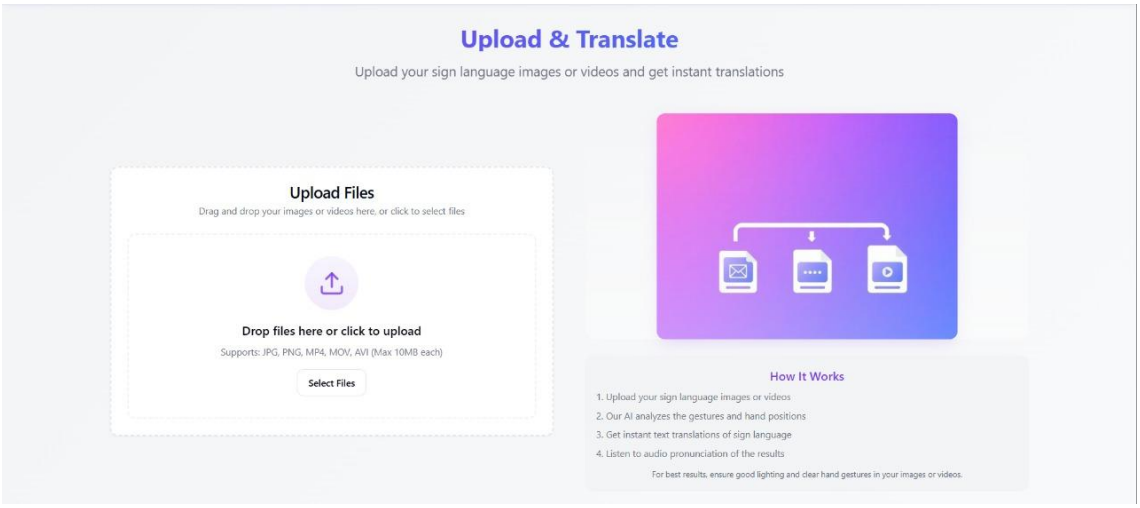
The first stage of the methodology involves the collection of a suitable dataset of sign language gestures. The dataset consists of images representing different hand gestures corresponding to alphabets, numbers, or commonly used words in sign language. These images can be obtained from publicly available sign language datasets or captured manually using a webcam to create a custom dataset. Collecting gesture samples from multiple individuals under different lighting conditions and hand orientations helps increase the diversity of the dataset and improves the robustness of the model.



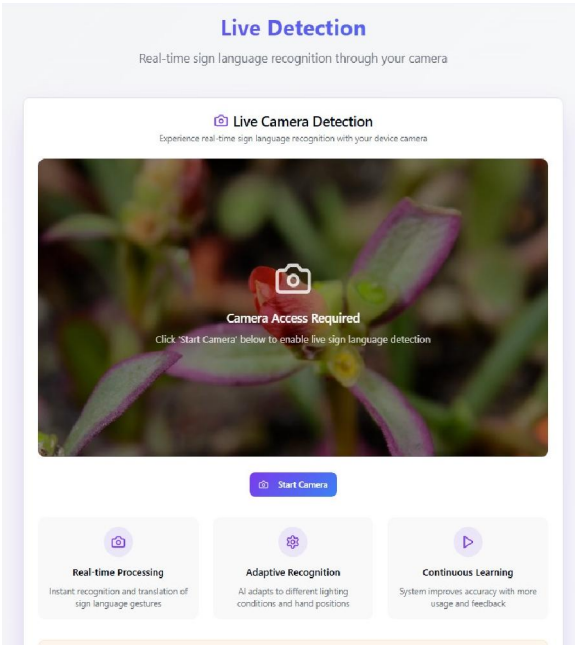
IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 14, April 2026
Impact Factor: 8.2

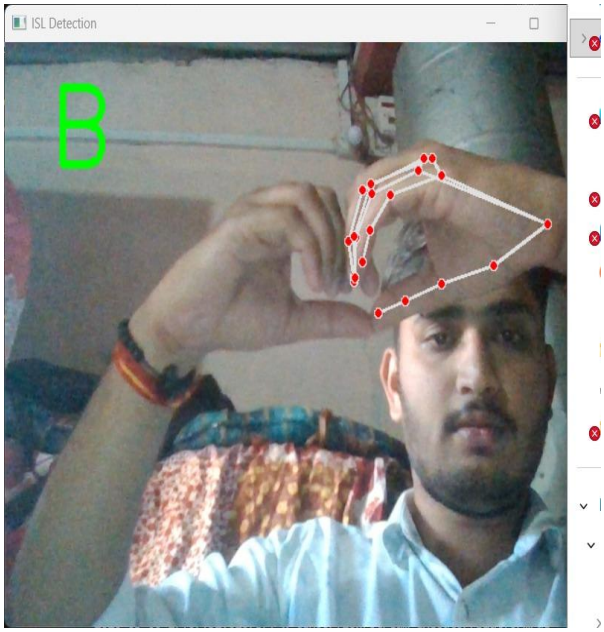


After the dataset is collected, the images undergo preprocessing to improve their quality and consistency before being used for model training. Preprocessing includes resizing the images to a fixed resolution, normalizing pixel values, and removing noise from the background. In addition, image segmentation techniques are applied to isolate the hand region from the surrounding background so that the model can focus on the relevant features of the gesture. These preprocessing steps help enhance the performance of the recognition model by reducing irrelevant information in the images.



IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 14, April 2026
Impact Factor: 8.2





Following preprocessing, feature extraction is performed using deep learning techniques. Convolutional Neural Networks (CNNs) are employed to automatically extract important visual features from the gesture images. CNN models are particularly effective for image-based tasks because they can identify patterns such as edges, shapes, and finger positions within the hand gesture. By learning hierarchical feature representations, the CNN model can effectively distinguish between different sign language gestures.

Once the features are extracted, the model is trained using labeled gesture images. During the training phase, the dataset is divided into training and testing subsets. The training dataset is used to teach the model how to recognize different gestures, while the testing dataset is used to evaluate its performance. The learning process involves adjusting the weights of the neural network using optimization techniques such as backpropagation and gradient descent. This training process allows the model to learn the relationship between input gesture images and their corresponding textual representations.

After the model has been successfully trained, it is deployed for real-time gesture recognition. The system captures live video frames from a webcam and processes each frame using image preprocessing techniques. The processed frames are then passed to the trained CNN model, which predicts the gesture class based on the learned patterns. Once a gesture is recognized, the system maps the predicted gesture to its corresponding textual output and displays the text on the screen. This allows users to perform sign language gestures and instantly obtain the equivalent text representation.

To ensure accessibility and ease of use, the entire recognition system is integrated into a web-based platform. The backend of the application is implemented using Python and Flask, while the frontend interface is developed using modern web technologies such as HTML, CSS, JavaScript, or ReactJS. This integration enables users to interact with the system through a simple graphical interface, activate the webcam, perform gestures, and view the translated text output in real time. Additionally, the system can incorporate multilingual support to display the recognized text in multiple languages, further enhancing its usability across different communities.

Through the combination of computer vision, deep learning, and web technologies, the proposed methodology provides an efficient and accessible solution for translating sign language gestures into text. The system aims to reduce communication barriers between hearing and non-hearing individuals and contribute to the development of inclusive assistive technologies.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 14, April 2026
Impact Factor: 8.2

IV. RESERCH GAP

In recent years, significant progress has been made in the field of sign language recognition using computer vision and deep learning techniques. Early research primarily focused on sensor-based systems, such as data gloves and wearable devices, which provided high accuracy in gesture capture but were costly, intrusive, and unsuitable for everyday use. This led to the emergence of vision-based approaches that utilize cameras and image processing techniques to detect hand gestures in a non-invasive manner. With the integration of advanced models such as Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), many systems have achieved impressive accuracy in recognizing predefined sets of static hand gestures under controlled environments.

However, despite these advancements, several critical gaps remain unaddressed. Most existing systems are limited to the recognition of isolated signs rather than continuous sequences, making them less effective for real-world communication where gestures occur in a fluid and context-dependent manner. Additionally, many models are trained and tested in controlled conditions with uniform backgrounds and lighting, which significantly limits their robustness when deployed in dynamic, real-world environments. Variations in illumination, camera angles, occlusions, and individual signing styles often lead to a noticeable drop in performance.

Another major limitation lies in the lack of generalization across users and sign languages. Existing approaches are typically dataset-specific and fail to adapt effectively to diverse users or regional variations in sign language. Furthermore, the scarcity of large-scale, well-annotated datasets continues to hinder the development of more accurate and scalable models. While some systems demonstrate high accuracy, they often do so at the cost of computational efficiency, making real-time implementation challenging due to latency and hardware constraints.

Moreover, current research largely focuses on gesture-level recognition without incorporating contextual or semantic understanding, which is essential for translating sign language into meaningful and grammatically correct text. As a result, the generated output may lack coherence and fail to capture the intended meaning of the user.

Therefore, there exists a clear need for a robust, real-time sign language recognition system that not only improves accuracy under diverse environmental conditions but also supports continuous gesture interpretation, enhances generalization across users, and operates efficiently with minimal latency. Addressing these challenges forms the core motivation of this research.

V. CONCLUSION

This research presents a real-time sign language recognition system that converts hand gestures into readable text using computer vision and machine learning techniques.

The system utilizes hand landmark detection and machine learning models to accurately recognize gestures captured through a webcam. By integrating the recognition model into a web-based platform, the system provides an accessible and user-friendly interface that does not require specialized hardware.

Another important aspect of the proposed system is its support for multilingual text output, including English, Hindi, and Marathi. This feature makes the system adaptable to different linguistic communities and increases its practical usability in real-world scenarios. The experimental implementation demonstrates that computer vision and deep learning techniques can be effectively combined to create assistive communication tools for people with hearing impairments.

Despite its promising performance, the system currently focuses mainly on recognizing individual gestures or alphabet-level signs. Future improvements may include the recognition of continuous sign language sentences, integration with speech synthesis to convert text into voice output, and the expansion of gesture datasets to improve accuracy and generalization. Additionally, advanced deep learning models and real-time mobile deployment could further enhance the system's performance and usability.

Overall, the proposed Sign Language to Text Conversion system demonstrates how modern AI technologies can contribute to inclusive communication and assistive technology development. By bridging the communication gap between hearing and non-hearing individuals, the system promotes social inclusion, accessibility, and equal

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 14, April 2026
Impact Factor: 8.2

participation in society. The research highlights the potential of intelligent human-computer interaction systems to create meaningful technological solutions that benefit diverse communities.

REFERENCES

[1]. K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in Proc. Int. Conf. Learning Representations (ICLR), 2015.
[2]. C. Szegedy, W. Liu, Y. Jia, and P. Sermanet, "Going deeper with convolutions," in Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR), 2015, pp. 1–9.
[3]. D. Tran, L. Bourdev, R. Fergus, L. Torresani, and M. Paluri, "Learning spatiotemporal features with 3D convolutional networks," in Proc. IEEE Int. Conf. Computer Vision (ICCV), 2015, pp. 4489–4497.
[4]. O. Koller, H. Ney, and R. Bowden, "Deep learning of mouth shapes for sign language," in Proc. IEEE Int. Conf. Computer Vision Workshops (ICCVW), 2015, pp. 85–91.
[5]. J. Huang, W. Zhou, H. Li, and W. Li, "Sign language recognition using 3D convolutional neural networks," in Proc. IEEE Int. Conf. Multimedia and Expo (ICME), 2015, pp. 1–6.
[6]. T. Starner, J. Weaver, and A. Pentland, "Real-time American sign language recognition using desk and wearable computer-based video," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 20, no. 12, pp. 1371–1375, Dec. 1998.
[7]. . P. Molchanov, S. Gupta, K. Kim, and K. Pulli, "Multi-sensor system for driver's hand-gesture recognition," in Proc. IEEE Int. Conf. Automatic Face & Gesture Recognition, 2015, pp. 1–8.
[8]. G. Pigou, S. Dieleman, P. Kindermans, and B. Schrauwen, "Sign language recognition using convolutional neural networks," in Proc. European Conf. Computer Vision Workshops (ECCVW), 2014, pp. 572–578