# Sign Language Interpreter (SignaVision)

**AASHRAYA MAN SINGH, BHUVANESH S, TANIKANTI DINESH BABU**

*Department of Computer Science and Engineering,*
*Faculty of Engineering and Technology, JAIN (Deemed-to-be) University*
*Bangalore, India*
22btrcn004@jainuniversity.ac.in, 22btrcn048@jainuniversity.ac.in, 22btrcn292@jainuniversity.ac.in

***Abstract*** **- Communication is a fundamental human need, yet millions of individuals with hearing and speech impairments face barriers in expressing themselves to a wider audience. This project presents a real-time vision-based system that bridges this gap by translating American Sign Language (ASL) fingerspelling gestures into both text and audible speech. The system leverages a webcam to capture hand gestures, utilizes MediaPipe to extract hand landmarks, and employs a Convolutional Neural Network (CNN) to classify gestures with high accuracy, even in varied lighting and background conditions. By transforming recognized signs into spoken and written output, this solution offers an accessible and cost-effective method for inclusive communication. The proposed approach avoids reliance on specialized hardware, making it scalable and adaptable for real-world applications.**

**The system's design focuses on robustness and user-friendliness, addressing challenges such as background noise, inconsistent lighting, and gesture similarity by simplifying gesture inputs to skeletal landmark representations. Through the integration of machine learning and computer vision techniques, the model generalizes well across diverse users and environments. Additionally, a text-to-speech module enables spoken feedback, simulating natural communication and making it easier for non-sign language users to understand. This project not only demonstrates the feasibility of real-time sign language recognition but also highlights the potential of AI-driven solutions to enhance accessibility and foster inclusivity in digital communication platforms**.

.

**[*Keywords*— ASL Recognition, Computer Vision, CNN, MediaPipe, Hand Gesture Recognition, Deep Learning, Text-to-Speech, Real-Time Systems]**

## I. INTRODUCTION

### *A. Background and Motivation*

Sign language is one of the oldest and most expressive forms of communication used by individuals with hearing and speech impairments. Despite its richness and clarity, sign language often creates a communication barrier between those who use it and the majority of the population that does not. This disconnect can lead to social exclusion, difficulty in accessing services, and limitations in educational or professional environments. The motivation behind this project stems from a desire to bridge that gap using technology that is both accessible and intuitive. With advances in computer vision and deep learning, it's now possible to create systems that can interpret hand gestures in real-time and translate them into text or speech—bringing a level of inclusivity that was previously difficult to achieve without human interpreters.

### *B. Objective*

The primary objective of this project is to design and develop a real-time sign language interpreter system that translates American Sign Language (ASL) fingerspelling gestures into both textual and spoken output. The system uses a standard webcam to capture hand gestures and leverages a Convolutional Neural Network (CNN) to classify them accurately. Once a gesture is recognized, it is displayed as text on-screen and simultaneously converted into speech using a text-to-speech engine. The goal is to build a cost-effective, vision-based tool that enhances communication between sign language users and non-signers, without relying on any specialized hardware.

### *C. Delimitation of Research*

While the system shows promising results, it does operate within certain constraints. The current scope is limited to static ASL gestures corresponding to the English alphabet (A–Z), which means it does not yet support dynamic signs or full sentence-level interpretation. Gesture recognition relies heavily on the quality of the camera feed, and although MediaPipe handles various lighting and background conditions effectively, extremely poor environments can still impact accuracy. Furthermore, the system is trained on skeletal landmark representations of gestures, which means it may misclassify signs that look visually similar unless handled through additional logic. These limitations define the boundaries of what the current version of the system can achieve.

### *D. Benefits of Research*

Despite its limitations, the research offers practical benefits that can have a meaningful social impact. First and foremost, it provides an accessible communication tool for people who are deaf or mute, helping them interact more easily in situations where a human interpreter is not available. It also introduces non-sign language users to a new way of interacting with those who rely on gestures, encouraging inclusivity. From a technological perspective, the project explores a lightweight yet effective way to implement gesture recognition using only a webcam and common machine learning libraries, making it suitable for educational use, early-stage prototyping, or even deployment on consumer devices in the future.

## II. LITERATURE SURVEY

Recently, deep learning and computer vision techniques have shown to be effective for real-time sign language recognition. Rupesh Kumar, Ashutosh Bajpai and Ayush Sinha have developed a real time system for recognition of American Sign Language (ASL) using MediaPipe and CNN with 99.95% accuracy on ASL alphabets [1]. Similarly, an Indian Sign Language (ISL) recognition model based on CNN using transfer learning and a custom dataset was

proposed by Heramba Limaye et al. with high accuracy in real time [2]. The researchers at Florida Atlantic University combined MediaPipe and YOLOv8 for ASL gesture recognition and achieved an accuracy of 98% [3]. Other studies pointed out the effectiveness of CNNs in extracting spatial features for sign recognition tasks [4], and systems that combined MediaPipe and OpenCV showed efficient real-time gesture detection [5].

Various challenges in sign language recognition systems like environmental variations and dataset limitations were mentioned in their review by M. Madhiarasan and Partha Pratim Roy [6]. Besides, the continuous sign language recognition based on spatial and temporal learning was studied with a fully convolutional network (FCN) [7]. Hybrid methods using SIFT descriptors achieved 98.74% accuracy on datasets of Pakistani Sign Language [8]. Previous work on ANN-based systems showed promising recognition performance on curvature and convex hull feature extraction [9]. In addition, open-source implementations using MediaPipe and machine learning models have demonstrated practical applications of real-time sign language interpretation with audio output support [10].

## III. Objectives and Methodology

### *A. Introduction*

Sign language is a key mode of communication for the deaf and mute, but its comprehension is mainly confined to the hearing-impaired community, leading to communication difficulties in everyday situations. It's not always easy or cheap finding traditional interpreters. With the rise of deep learning and computer vision, intelligent systems for real-time gesture recognition have been proposed. The project is a vision-based American Sign Language (ASL) recognition system that uses a webcam and trained neural network to recognize the fingerspelling gestures of ASL and translate them into text and speech outputs [11]. A primary aim is to incorporate a strong Natural Language Processing (NLP) engine that facilitates conversational interactions, thereby eliminating the need for rigid command interfaces. By enabling intent recognition, keyword extraction, contextual comprehension, and adaptable query handling, the system aspires to support natural inquiries such as vehicle searches, comparisons, and administrative tasks. Tools like NLTK and spaCy are foundational to these NLP functionalities [2], [3].

### *B. Problem Definition*

This results in a huge communication gap between sign language users and non-users and creates difficulties for deaf or mute people in terms of accessibility and social inclusion. Current systems rely on expensive hardware like sensor gloves, or are sensitive to environmental factors like illumination and background noise. The challenge to be addressed in this project is to develop a low-cost, real-time, camera-based ASL recognition system capable of accurately translating hand gestures into text and audio outputs.

### *C. Model and System Architecture*

The proposed system has four stages which are data acquisition, preprocessing, gesture classification and output generation as shown in Fig. 3.1. A webcam captures live hand gestures and MediaPipe extracts hand landmarks from the frames. The landmarks are displayed on a white background, which reduces the noise and improves the consistency. The processed images are then classified using a Convolutional Neural Network (CNN) trained on ASL alphabet gestures . Finally, the recognised gesture is displayed as text and converted into speech by a text-to-speech engine.

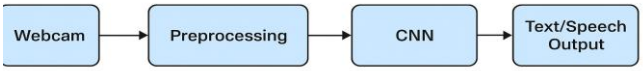


*Fig1: Flowchart depicting the working of the system architecture*

### *D. Proposed Algorithm*

The essential recognition algorithm is developed using a CNN to classify hand gestures according to their skeleton landmark patterns[12]. For training purposes, there will be about 180 images per letter (from A to Z) that were created using landmark information rendered on a plain canvas. In order to enhance the model's accuracy and avoid any misclassification of symbols, there was a need to categorize the 26 alphabets into eight logical classes (for example, [A, E, M, N, S, T] and [B, D, F, I, U, V, K, R, W]), and the second logical layer would differentiate within each group using their landmarks.

*Table 1: Tabular representation of the 8 segregated classes*

| **Class No.** | **Grouped Alphabets** | **Reason for Grouping** |
|---|---|---|
| Class 1 | Y, J | Both involve curved pinky/thumb movement; visually similar loops |
| Class 2 | C, O | Similar circular hand shapes |
| Class 3 | G, H | Hand in sideways position with extended fingers |
| Class 4 | B, D, F, I, U, V, K, R, W | All involve extended index/middle fingers in various orientations |
| Class 5 | P, Q, Z | Complex or angular finger shapes, often mistaken for one another |
| Class 6 | A, E, M, N, S, T | Closed fist or partially folded fingers; very compact signs |
| Class 7 | L | Unique hand shape (index + thumb extended), easy to isolate |
| Class 8 | X | Bent index finger gesture, less confused with others |

### E. Proposed Work

The proposed solution mainly concentrates on developing a sturdy, straightforward, and easily accessible sign language recognition system. As compared to the glove-based solutions, the proposed method only needs an ordinary webcam, thereby lowering the costs related to hardware and making the system easily accessible. MediaPipe serves as a tool for performing precise real-time hand tracking, while the CNN serves as a classifier for the gestures made. Moreover, the suggested work incorporates a text-to-speech component that not only provides text but also delivers the text in speech.

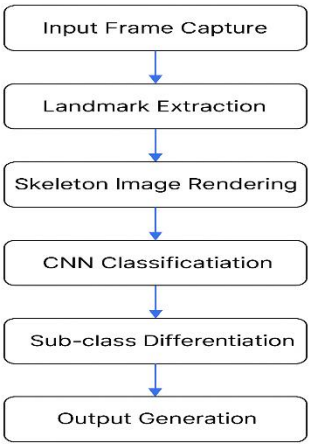


*Fig 2: Flowchart summarizing the flow of the work*

## IV. Implementation

The implementation of the proposed sign language interpreter system focuses on developing a real-time, camera-based solution for recognizing American Sign Language (ASL) gestures. The system integrates MediaPipe for hand landmark detection, OpenCV for image preprocessing, and a Convolutional Neural Network (CNN) for gesture classification. Recognized gestures are converted into both text and speech outputs using a text-to-speech engine, enabling effective and accessible communication in real time.

### A. Hardware and Software Design and Implementation

The implementation of the sign language translator system was achieved by the usage of low-cost components. First of all, the main hardware component used is an ordinary webcam that allows capturing video images of hand gestures in real time. The proposed approach does not use sensors but utilizes a webcam without the necessity to wear gloves or use any other specialized device.

The design process of the software was performed in Python 3.9 programming language using PyCharm Integrated Development Environment. For real-time detection of hand landmarks, a lightweight tool called MediaPipe was applied [14]. It extracts 21 hand landmarks in each image frame after which those images are processed by drawing skeleton hand structures on a white screen using the OpenCV library. This approach provides the reduction of background noise and lighting differences.

Processed images are resized and fed into CNN built in Keras with TensorFlow backend for alphabetic gesture identification. The predicted gesture is translated into text and further turned into speech via pyttsx3 text-to-speech tool [15].

### B. Software Algorithm

The working of the sign language interpreter is through a structured algorithm where the software captures the frames from a webcam, preprocesses the images using hand skeleton extraction algorithm, and classifies them based on gestures before generating an output. The first step involves continuous capture of video frames using a camera. Every frame captured is then passed through the MediaPipe process where 21 hand landmark coordinates that signify different hand joint locations are extracted. The MediaPipe result is then rendered into a hand skeleton image on a white background through the OpenCV library [16].

This process filters out any noise present in the environment that might affect the inputs. After the above steps, the image produced is grayscaled and resized depending on the required dimensions for CNN input. CNN uses the preprocessed data and predicts eight grouped gesture classes [17][18]. Different ASL alphabets are grouped into eight categories since they possess the same shape features.

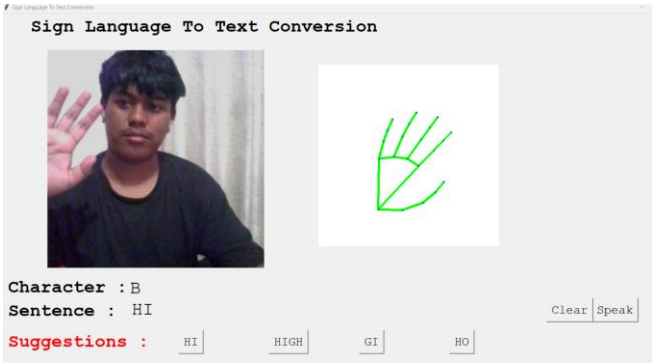


*Fig 3: Depiction of the real-time interpretation of the model*

If the CNN predicts a grouped gesture class, an additional rule-based classification step is applied to identify the exact alphabet within that group. This logic analyzes geometric relationships and relative positions between hand landmarks to distinguish visually similar gestures [19][20]. After the final letter is determined, it is displayed as text on the screen and simultaneously converted into speech using the pyttsx3 text-to-speech engine. This integrated pipeline enables the system to operate in real time while providing accurate and responsive visual and audio outputs.

## V. Experimental Results and Performance Analysis

The implementation of the system yielded encouraging results in terms of its accuracy, efficiency, and usability. The CNN yielded excellent performance in terms of ASL gestures recognition from skeletal landmark images using MediaPipe technology. The use of skeletal rendering rather than raw image classification provided higher robustness by mitigating background noise, different light conditions, and skin tone interference. Furthermore, clustering 26 ASL alphabets into 8 classes made gesture recognition easier by decreasing confusion between similar movements.

The system worked efficiently in terms of real-time operation with minimal time difference between gesture recognition and output display. The addition of text-to-speech capability contributed to improved usability by enabling two types of feedback (visual and auditory). The

model demonstrated high performance in terms of smooth processing of frames with single-hand gestures.

Nonetheless, there is a number of issues that may hinder effective system performance. There are some difficulties in recognition of gestures that were partly occluded or performed too fast/with an unusual angle. Gestures beyond the frame of webcam operation are not recognized by the algorithm. It seems that temporal models such as RNN and Transformer can help address these problems.

In summary, the findings have confirmed that the approach suggested to recognize ASL using cameras works well. The system managed to translate hand gestures into texts and speech outputs accurately while having a small and simple structure.

## VI. Conclusion and Future Scope

This project successfully developed a real-time American Sign Language (ASL) interpreter capable of converting hand gestures into both text and speech outputs. The system combines MediaPipe for hand landmark detection, OpenCV for preprocessing and skeletal rendering, and a Convolutional Neural Network (CNN) for gesture classification. By utilizing skeletal hand representations instead of raw images, the system achieved high recognition accuracy while remaining robust against background noise and lighting variations. The implementation demonstrated reliable real-time performance using only a standard webcam and commonly available Python libraries, making the solution cost-effective and accessible.

The integration of text-to-speech functionality further enhanced the usability of the system by enabling both visual and auditory communication. The achieved accuracy and responsiveness validate the effectiveness of combining computer vision and deep learning techniques for assistive communication technologies. The project successfully addressed the communication barrier between sign language users and non-signers while maintaining a lightweight and scalable architecture.

Although the current system focuses on static ASL fingerspelling gestures, several opportunities exist for future enhancement. The system can be extended to recognize dynamic gestures, complete words, and sentence-level sign language translation using sequence-based deep learning models such as Long Short-Term Memory (LSTM) networks or transformers. Future improvements may also include confidence-based feedback mechanisms, multi-hand gesture recognition, and mobile or embedded system deployment for wider accessibility. Integration with communication platforms, educational tools, and public service applications could further increase the practical impact of the system.

Overall, this project demonstrates the potential of artificial intelligence, computer vision, and deep learning in developing affordable and scalable solutions for inclusive communication and accessibility.

## References


1. R. Kumar, A. Bajpai, and A. Sinha, “Real-Time American Sign Language Recognition Using MediaPipe and CNN,” *arXiv preprint arXiv:2305.05296*, 2023. [Online]. Available: https://arxiv.org/abs/2305.05296
2. H. Limaye, S. Kumar, S. Yadav, and S. Choudhary, “American Sign Language Recognition using CNN,” *SSRN Electronic Journal*, 2022. [Online]. Available: https://papers.ssrn.com/sol3/Delivery.cfm/SSRN_ID4169172_code5298930.pdf
3. Florida Atlantic University, “AI Model Highly Accurate in Recognizing Sign Language Gestures,” *The Hearing Review*, 2024. [Online]. Available: https://hearingreview.com/inside-hearing/research/ai-model-highly-accurate-in-recognizing-sign-language-gestures
4. A. Chauhan, B. Bhushan, and S. Pundir, “American Sign Language Recognition Using Deep Convolutional Neural Networks,” *Materials Today: Proceedings*, vol. 46, pp. 6831–6835, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2666990021000471
5. M. Bali, “Sign Language Detection for Deaf Using Deep Learning, MediaPipe & OpenCV,” *Medium*, 2023. [Online]. Available: https://medium.com/@mayank.bali/sign-language-detection-for-deaf-using-deep-learning-mediapipe-u-opencv-4c5151e2374c
6. M. Madhiarasan and P. P. Roy, “A Comprehensive Review on Sign Language Recognition Systems,” *arXiv preprint arXiv:2204.03328*, 2022. [Online]. Available: https://arxiv.org/abs/2204.03328
7. K. L. Cheng, Y. Wang, and T. Wu, “Fully Convolutional Networks for Continuous Sign Language Recognition,” *arXiv preprint arXiv:2007.12402*, 2020. [Online]. Available: https://arxiv.org/abs/2007.12402
8. A. Qayyum, M. A. Jamil, and H. Sherazi, “Scale-Invariant Feature Based Deep CNN for Pakistani Sign Language Recognition,” *Journal of King Saud University - Computer and Information Sciences*, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1319157824000235
9. M. M. Islam, M. R. Hasan, and M. K. Hasan, “A Review on Vision-Based American Sign Language Recognition: Its Techniques and Outcomes,” *International Journal of Engineering & Technology*, vol. 7, no. 4.34, pp. 237–242, 2018. [Online]. Available: https://www.researchgate.net/publication/326643498
10. Laplaces42, “Real-Time Sign Language Interpreter Using MediaPipe,” *GitHub*, 2023. [Online]. Available: https://github.com/laplaces42/sign-language-interpreter
11. Neural Network based Indian Sign Language Recognition using hand crafted features,” 2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT), Kharagpur, India, 2020, pp. 1–6. Available: https://ieeexplore.ieee.org/document/9225294
12. R. H. Chile, R. V. Dharaskar and V. M. Wadhai, “Signet: A Deep Learning based Indian Sign Language Recognition System,” 2019 IEEE International Conference on Communication and Signal Processing (ICCSP), Chennai, India, 2019, pp. 0669–0673. Available: https://ieeexplore.ieee.org/document/8698006

13. A. Raut and P. N. Raut, “Real time Indian Sign Language Recognition System to aid deaf-dumb people,” 2011 IEEE 13th International Conference on Communication Technology, Jinan, China, 2011, pp. 699–702. Available: https://ieeexplore.ieee.org/document/6157974
14. N. S. Ameen and M. A. Khan, “ML Based Sign Language Recognition System,” 2021 International Conference on Innovative Trends in Information Technology (ICITIIT), Kottayam, India, 2021, pp. 1–6. Available: https://ieeexplore.ieee.org/document/9399594
15. V. Srivastava and A. Sharma, “Machine learning techniques for Indian sign language recognition,” 2017 International Conference on Current Trends in Computer, Electrical, Electronics and Communication (CTCEEC), Mysore, India, 2017, pp. 112–117. Available: https://ieeexplore.ieee.org/document/8454988
16. R. Kaur and V. Kaur, “Transfer Learning with L2 Norm Regularization for classifying static Two Hand Hindi Sign Language Gestures,” 2020 IEEE 9th International Conference on Communication Systems and Network Technologies (CSNT), Gwalior, India, 2020, pp. 239–243. Available: https://ieeexplore.ieee.org/document/9113661
17. D. Patel and D. Thakor, “Sign Language Recognition Based on Computer Vision,” 2021 IEEE International Conference on Artificial Intelligence and Computer Applications (ICAICA), Dalian, China, 2021, pp. 322–326. Available: https://ieeexplore.ieee.org/document/9498024
18. D. B. Patel and K. Bhatt, “A Vision-based System for Recognition of Words used in Indian Sign Language Using MediaPipe,” 2021 Sixth International Conference on Image Processing, Applications and Systems (IPAS), Genova, Italy, 2021, pp. 1–6. Available: https://ieeexplore.ieee.org/document/9742141
19. A. Mahanta, A. Singh and S. Ghosh, “Real-Time Indian Sign Language Recognition System using YOLOv3 Model,” 2021 Sixth International Conference on Image Processing, Applications and Systems (IPAS), Genova, Italy, 2021, pp. 1–6. Available: https://ieeexplore.ieee.org/document/9742198
20. A. Wadhwani and D. Patel, “Character and Word Level Gesture Recognition of Indian Sign Language,” 2023 IEEE 8th International Conference for Convergence in Technology (I2CT), Pune, India, 2023, pp. 1–6. Available: https://ieeexplore.ieee.org/document/10418341