Anomaly-Aware Gesture Recognition for Sign Language Translation Using Deep Learning

[1]Kancharla Harikrishna, [2]M. M. Rayudu
[1]M. Tech Student, Department of CSE, Prakasam Engineering College (Autonomous), Kandukur, India
[2]Associate Professor, Department of CSE, Prakasam Engineering College (Autonomous), Kandukur, India

ABSTRACT *- Deaf and hard-of-hearing individuals face communication barriers due to limited public sign language understanding and costly or impractical sensor-based tools. Traditional gesture systems also struggle with high computational complexity, large dataset requirements, and varying lighting, backgrounds, and hand orientations. This project proposes an anomaly-aware sign language translation system combining MediaPipe hand landmark extraction, deep learning models such as CNNs and RNNs, and language modeling to translate gestures into words and sentences in real time. The framework extracts 21-point hand landmarks from live video streams using OpenCV and MediaPipe, training deep neural models saved for fast gesture recognition. Detected signs are mapped to words with context-aware suggestions and dynamic sentence formation, including support for ambiguous or incomplete signs. By applying natural language processing alongside gesture classification, the system improves fluency and grammar in everyday communication. Built in Python on standard cameras, laptops, or smartphones, it offers a scalable, cost-effective assistive tool across diverse vocabularies and multiple sign languages. Overall, the project supports inclusive communication for the deaf and hard-of-hearing community across education, healthcare, workplaces, and daily life.*

KEYWORDS: *— Sign Language Translation, Gesture Recognition, MediaPipe, Deep Learning, Sentence Formation.*

INTRODUCTION

Millions of deaf and hard-of-hearing individuals face communication barriers due to limited public sign language awareness and costly or impractical sensor-, glove-, or depth-camera-based tools. To overcome these challenges, this project proposes an anomaly-aware sign language translation system using computer vision, deep learning, and natural language processing to convert gestures into text and

sentences in real time. Using OpenCV and MediaPipe to extract 21-point hand landmarks across varying lighting, backgrounds, and orientations, deep learning models (such as CNNs, RNNs, or transformers) classify signs and map them to words. A language model then provides context-aware word suggestions, handles incomplete or ambiguous signs, and forms coherent sentences on standard cameras, laptops, or smartphones. Ultimately, this scalable, cost-effective assistive system bridges communication gaps and supports inclusive participation across education, healthcare, workplaces, and daily life.

RELATED WORK

Sign language recognition has evolved from sensor-, glove-, and depth-camera systems toward vision-based machine learning and deep learning. Early studies used SVM, Decision Trees, KNN, Random Forest, ANN, Haar Cascade detection, and contour extraction for real-time gestures, while newer research leverages CNNs, RNNs, transformers, and attention-based GCNs to capture spatial keypoints and dynamic motion across lighting, backgrounds, and orientations.Combining gesture recognition with NLP and language models also aids word suggestions, ambiguity handling, and sentence formation despite complexity and dataset challenges, motivating our lightweight MediaPipe, deep learning, and language-driven approach.

LITERATURE SURVEY

The literature survey highlights how vision-based machine learning and deep learning models enhance sign language gesture recognition and translation compared to traditional sensor-heavy or classical rule/contour-based setups. While earlier approaches relied on wearable sensors, gloves, depth cameras, Haar Cascade detection, contour extraction, or classifiers such as SVM, Decision Trees, KNN, Random Forest, and ANN, they often faced hardware costs, background/lighting sensitivity, or high computational complexity. Recent studies incorporate deep neural architectures—including CNNs, RNNs, transformers, and attention-based GCNs alongside lightweight landmark tracking like MediaPipe to capture spatial keypoints and dynamic hand movements with higher accuracy and real-time speed. Furthermore, integrating gesture recognition with language modeling and NLP enables context-aware word suggestions, ambiguity handling, and coherent sentence formation, fostering more natural and inclusive communication for deaf and hard-of-hearing users.

EXISTING METHOD

Traditional sign language recognition systems often focus either on hand gestures

or body movements in isolation, whereas sign language involves the whole body, including facial expressions that convey emotion and grammar. More advanced approaches combine computer vision, machine learning, and deep learning across multiple gesture components—such as hand movements, facial expressions, and body posture—to interpret signs. However, integrating multiple modalities creates high computational complexity that strains real-time processing on resource-constrained devices. Performance also depends heavily on large, well-labeled datasets covering diverse signers, hand shapes, orientations, and environments, where poor or limited data hurts generalization. Furthermore, synchronizing hand gestures, facial expressions, and body posture in real time requires complex fusion algorithms and remains difficult under varying conditions. Overall, existing systems face computational overhead, dataset limitations, and multimodal synchronization challenges that limit practical real-time translation.

PROPOSED METHOD

The proposed system utilizes a real-time, anomaly-aware sign language translation framework combining OpenCV video streaming, MediaPipe hand landmark extraction, deep learning models such as Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), and Natural Language Processing (NLP) for word suggestion and sentence formation. Live video frames are captured from standard cameras or webcams, preprocessed via horizontal flipping, resizing, and color-space conversion, and analyzed with MediaPipe Hands to extract 21-point hand landmarks (x, y, z coordinates) stored as NumPy (.npy) arrays. A deep learning model—such as a CNN trained and saved as an .h5 model, or enhanced with RNNs/transformers—classifies static and dynamic gestures into corresponding words or characters across varying lighting, backgrounds, and hand orientations. A gesture detection and language modeling stage maps recognized signs to vocabulary, computes confidence scores, provides context-aware word suggestions, and handles incomplete or ambiguous signs to form grammatically coherent sentences. Built in Python using OpenCV, NumPy, and TensorFlow/Keras on accessible hardware like laptops or smartphones, the system supports diverse vocabularies and multiple sign languages. Ultimately, this lightweight approach reduces hardware cost, improves communication fluency, and enables practical real-time assistive communication for deaf and hard-of-hearing users.

SYSTEM ARCHITECTURE

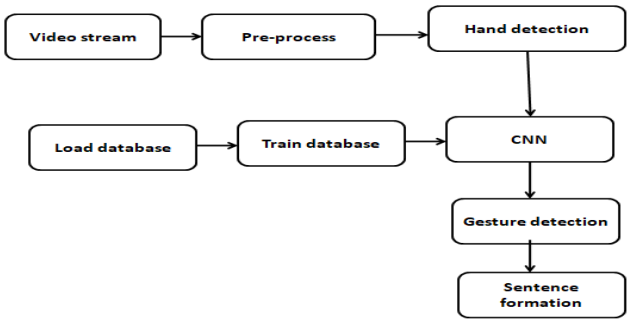


Fig. 1. Proposed architecture of the anomaly-aware gesture recognition system for sign language translation using a CNN.

METHODOLOGY DESCRIPTION

Video Streaming Module: The Video Streaming module initializes video capture via OpenCV and standard cameras or webcams, handling live video frames and real-time display streams in Python for continuous gesture input.

Preprocessing Module: The Preprocessing module prepares incoming video frames through horizontal flipping for selfie-view alignment, resizing to consistent input dimensions, and color-space conversions such as BGR to RGB and grayscale.

Hand Tracking and Landmark Extraction Module: The Hand Tracking module uses Google's MediaPipe Hands framework to locate hands, filter out background clutter, and extract 21 landmark points (x, y, and z coordinates covering fingertips, joints, and palm base) for static and dynamic gestures.

Database Creation and Storage Module: The Database Creation module collects extracted hand landmarks and gesture labels across diverse users, orientations, and conditions, storing structured representations as NumPy (.npy) arrays that serve as the system's knowledge base.

Deep Learning Training and Model Module: The Deep Learning module loads .npy gesture datasets into training and validation splits, applies augmentation (rotation, scaling, flipping), trains a Convolutional Neural Network (CNN)—optionally paired with Recurrent Neural Networks (RNNs) or transformers—via backpropagation and gradient descent, and saves the trained model as an .h5 file.

Gesture Detection Module: The Gesture Detection module loads the trained .h5 model with TensorFlow or Keras, classifies real-time hand landmarks into words, symbols, or characters mapped to a vocabulary database, and computes confidence scores to handle ambiguous or similar signs.

Sentence Formation Module: The Sentence Formation module applies Natural Language

Processing (NLP) and language modeling to sequence recognized words grammatically, generate context-aware word suggestions (such as predicting *"you"* after *"How"* and *"are"*), handle incomplete signs, and support text or speech output.

RESULTS AND DISCUSSION



Fig. 2: Home Page of the Anomaly-Aware Gesture Recognition System for Sign Language Translation Using Deep Learning

The home page introduces the system for real-time, AI-powered gesture recognition and sign language translation. The "Start Gesture Detection" button opens the live detection module, where hand gestures are captured and translated into text.

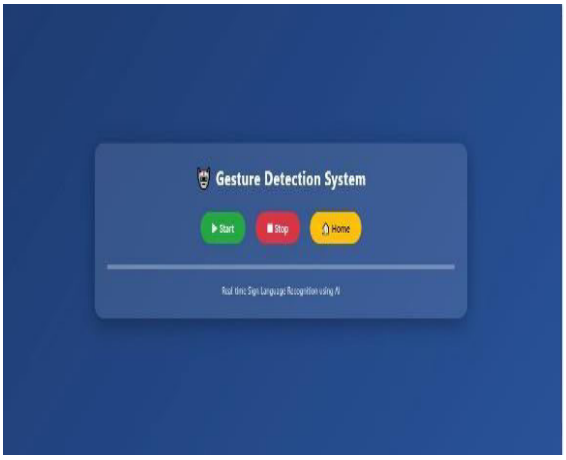


Fig. 3: Gesture Detection System Interface with Start, Stop and Home Controls

This page is the main control panel for real-time sign language recognition using AI. The Start and Stop buttons turn the camera-based gesture detection on and off, and the Home button returns the user to the main page.

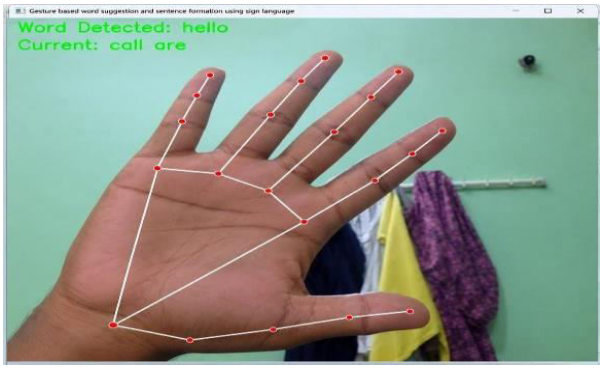


Fig.4: Real-Time Hand Gesture Detection with Landmark Tracking and Word Prediction

The system tracks the hand using key landmark points and connecting lines, and the detected gesture is shown on screen as the recognized word ("hello"). The current sequence of detected words is displayed as well, which supports sentence formation from sign language.

CONCLUSION

The project successfully integrates real-time computer vision, MediaPipe 21-point hand landmark tracking, deep learning models such as CNNs and RNNs, and language modeling to translate sign language gestures into sentences. By handling varying lighting, backgrounds, and hand orientations and mapping signs to vocabulary with confidence scores, the system delivers reliable recognition across static and dynamic gestures. With context-aware word suggestions, ambiguity handling for

incomplete signs, and Python-based execution on standard cameras, laptops, or smartphones without costly sensors or gloves, it offers a cost-effective, scalable assistive tool. Its adaptability to diverse vocabularies and multiple sign languages supports inclusive communication across education, healthcare, workplaces, and daily life.

FUTURE SCOPE

Future developments will expand the system to support larger vocabularies and multiple regional sign languages for broader global applicability. Incorporating advanced Natural Language Processing models such as transformers will further strengthen grammatical accuracy and context awareness during sentence formation. Adding real-time speech synthesis will enable instant spoken output from recognized gestures for smoother everyday interactions, while extending recognition to facial expressions and body posture will capture emotional and grammatical cues more completely. Furthermore, integrating cloud infrastructure will support continuous learning and personalization from user inputs, and deploying across mobile platforms will broaden accessibility without specialized hardware, evolving the solution into a comprehensive assistive communication platform for the deaf and hard-of-hearing community.

REFERENCES

1. T. Starner and A. Pentland, "Real-time American Sign Language recognition from video using hidden markov models," *Motion-Based Recognition*, pp. 227–243, 1997.
2. Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.
3. H. Li and W. Wang, "Gesture recognition based on CNN using depth information," *Sensors*, vol. 20, no. 1, p. 14, 2020.
4. C. H. Sung and Y. C. Chang, "A study on hand gesture recognition using CNN for human-computer interaction," *Applied Sciences*, vol. 10, no. 11, p. 3878, 2020.
5. M. M. Ameer and S. Dey, "Deep learning for hand gesture recognition: A comprehensive review," *Computer Science Review*, vol. 38, p. 100279, 2020.
6. S. Sivaramakrishnan and S. Jayaraman, "Hand gesture recognition using CNN and LSTM networks," *IEEE Access*, vol. 8, pp. 168953–168963, 2020.
7. P. Das, T. Ahmed, and M. F. Ali, "Static hand gesture recognition for American Sign Language using deep convolutional neural network," *IEEE TENSYMP*, pp. 1762–1765, 2020.
8. M. López and M. A. Fernández, "Real-time gesture recognition for sign language interpretation using deep learning techniques," *Sensors*, vol. 20, no. 18, p. 5238, 2020.

9. A. Khan and S. Fatima, "Hand gesture recognition using deep learning techniques: A review," *Int. J. Adv. Comput. Sci. Appl. (IJACSA)*, vol. 12, no. 3, pp. 1–9, 2021.
10. R. Kumar and P. Sharma, "Hand gesture recognition in video data using deep learning techniques," *Multimedia Tools and Applications*, vol. 80, no. 4, pp. 5655–5680, 2021.
11. Z. Zhou and H. Huang, "Hand gesture recognition based on deep learning: A survey," *J. Ambient Intell. Humaniz. Comput.*, vol. 12, no. 2, pp. 1771–1784, 2021.
12. A. Choudhary and S. Joshi, "A survey on gesture recognition techniques: A comprehensive review," *Artificial Intelligence Review*, vol. 54, no. 6, pp. 4419–4453, 2021.
13. R. Kandhasamy and R. Sundaram, "Real-time sign language recognition using deep learning techniques," *J. King Saud Univ. Comput. Inf. Sci.*, 2021.
14. M. M. Haque and S. Biswas, "Real-time gesture recognition based on machine learning techniques for human-machine interaction," *Sensors*, vol. 21, no. 1, p. 223, 2021.
15. C. M. Sandra, R. B. Reddy, P. M. Reddy, and L. Vamsi, "Sign language recognition using deep learning," *Zenodo*, 2026.