International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 06 | June - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

Real-Time American Sign Language Recognition Using Machine Learning

Khan Abdullah Riyaz Ahmed, Mr. Ali Karim Sayed

Department of B.Voc (Artificial Intelligence and Data Science)

Anjuman-I-Islam's Abdul Razzaq Kalsekar Polytechnic, New Panvel

Abstract

This paper presents the design and implementation of a real-time sign language recognition system focused on detecting static hand gestures representing the English alphabets (A–Z). The project leverages computer vision and machine learning to provide an accessible, low-cost tool for communication between hearing-impaired individuals and non-signers. The system captures webcam input, extracts hand landmarks using Google's Mediapipe framework, and classifies gestures through a trained Random Forest model. A Streamlit-based user interface displays the detected letters and enables real-time sentence construction with editing controls. The model, trained on a custom dataset, achieved an accuracy of 99.71% in testing. This paper discusses the system's methodology, architecture, dataset preparation, model training, and evaluation. The results highlight the potential of combining machine learning and vision-based solutions to support inclusive communication technologies.

Key Words: Sign Language, Gesture Recognition, Mediapipe, Random Forest, Streamlit, Computer Vision, Accessibility, Human-Computer Interaction

1. INTRODUCTION

Sign language is a powerful form of communication for individuals with hearing or speech impairments. However, the language barrier between signers and non-signers creates communication challenges in day-to-day scenarios. With the increasing availability of machine learning and vision-based technologies, there is an opportunity to bridge this gap using real-time gesture recognition systems.

This research focuses on building a real-time sign language detection system capable of recognizing static ASL (American Sign Language) gestures corresponding to alphabets A to Z. The project avoids expensive hardware by relying on standard webcams and open-source libraries, making it suitable for both educational and practical environments.

2. METHODOLOGY

2.1 Data Collection and Preparation

The dataset for training was collected using a custom script (collect_imgs.py) which uses the Mediapipe library to detect hand landmarks from webcam input. For each letter (A to Z), 200 samples were recorded, saved in .npy format, and then compiled into a single dataset using create_dataset.py. Each sample consists of 42 normalized (x, y) coordinate values derived from 21 landmark points.

2.2 Model Selection and Training

A Random Forest Classifier from the scikit-learn library was chosen for its speed, robustness, and simplicity. The model was trained on the compiled dataset (data.pickle) and achieved a classification accuracy of 99.71% on the test set. The model was saved as model.p for real-time inference.

2.3 Real-Time Inference Pipeline

During inference, live webcam frames are processed using Mediapipe to extract landmarks. The extracted features are normalized and passed to the trained Random Forest model to predict the corresponding alphabet. The prediction is displayed in real-time using OpenCV and Streamlit.

2.4 Streamlit App Interface

The Streamlit-based web app provides:

- Start/stop detection button
- Backspace, space, and clear controls
- Editable sentence window
- Real-time prediction visualization

This app is designed for both desktop and mobile use.

3. RESULTS

- Model Accuracy: 99.71% (on unseen test data)
- Confusion Matrix: Strong diagonal dominance with minimal misclassifications between similar signs (A-T, C-O).
- Interface Testing: Real-time frame rate remains above 15 FPS with no major lags.
- Usability: The app is user-friendly and allows real-time feedback with editing controls for sentence correction.

4. CONCLUSIONS

This research successfully demonstrates a low-cost, real-time hand gesture recognition system for the English alphabet using machine learning and computer vision. The use of simple tools like Mediapipe, scikit-learn, and Streamlit makes the system accessible to students and developers alike. While the current system handles static gestures, it serves as a strong foundation for dynamic gesture recognition in the future.

ACKNOWLEDGEMENT

I would like to thank my faculty, mentors, and institution for their constant support throughout the project. Special thanks to open-source contributors whose tools and libraries made this work possible.

REFERENCES

1. Lugaresi, C., et al. “MediaPipe: A Framework for Building Perception Pipelines.” arXiv preprint arXiv:1906.08172, 2019.
2. scikit-learn Documentation. https://scikit-learn.org
3. OpenCV Documentation. https://docs.opencv.org
4. Streamlit Documentation. https://docs.streamlit.io

Author Bio

Abdullah Khan is a final-year student of B.Voc in Artificial Intelligence and Data Science at Anjuman-I-Islam’s Abdul Razzaq Kalsekar Polytechnic, New Panvel. He is passionate about accessible AI technologies and applying machine learning to solve real-world problems.