

SILENT VOICES AMPLIFIED: DECIPHERING SIGN LANGUAGE WITH DEEP LEARNING FOR DUMB AND DEAF

Project Reference No.: 48S_BE_3150

College : Vivekananda College of Engineering and Technology, Puttur
Branch : Artificial Intelligence and Machine Learning
Guide(s) : Prof. Akshaya D Shetty
Student(s): Ms. Sinchana B
Ms. Dhruthi S Bhat
Mr. Sagar K R

Keywords:

American Sign Language, MobileNetV2, MediaPipe, Gesture Recognition, Deep Learning, Real-Time Detection.

Introduction:

In a world where communication is key, individuals with hearing and speech impairments face significant challenges in expressing themselves effectively. American Sign Language (ASL) serves as a bridge, but it is often misunderstood or unknown by the general population, leading to a communication gap. With advancements in artificial intelligence and computer vision, there is a growing potential to create systems that can interpret sign language in real-time.

This project proposes the development of a deep learning-based system that can detect and recognize ASL gestures from live video input. Using lightweight models like MobileNetV2 and accurate hand tracking through MediaPipe, the system aims to classify 14 ASL gestures (5 vowels and 9 numbers) and display the corresponding text output. This will serve as an assistive tool for the deaf and hard-of-hearing community and promote inclusive communication in various social and professional environments.

Objectives:

1. To collect and prepare a diverse dataset of 14 ASL gestures (A, E, I, O, U, 1–9) for training and evaluation.
2. To design a robust deep learning model based on MobileNetV2 for gesture recognition.
3. To implement MediaPipe-based hand tracking for accurate input capture.
4. To develop a real-time system that can predict and display gesture outputs from live video.
5. To ensure the model's robustness under varied lighting and background conditions.

Methodology:

The project will begin with the creation of a dataset comprising images of 14 ASL gestures. Each gesture will be recorded through video and converted into image frames. These images will undergo preprocessing techniques such as resizing, grayscale conversion, normalization, and data augmentation.

A MobileNetV2 model pre-trained on ImageNet will be adapted for the task by fine-tuning its final layers. Dense layers with dropout and batch normalization will be added to improve accuracy and prevent overfitting. The training process will involve techniques like early stopping and learning rate adjustment for optimal performance.

MediaPipe will be used to detect and isolate hand landmarks from the webcam input. The detected region of interest will be processed and passed into the trained model to generate predictions. OpenCV will be used to display real-time predictions and confidence scores on the video feed.

The final system will be tested under different lighting conditions and hand orientations to ensure generalization. The model will be saved in a deployable format, and the complete application will be built using Python and TensorFlow.

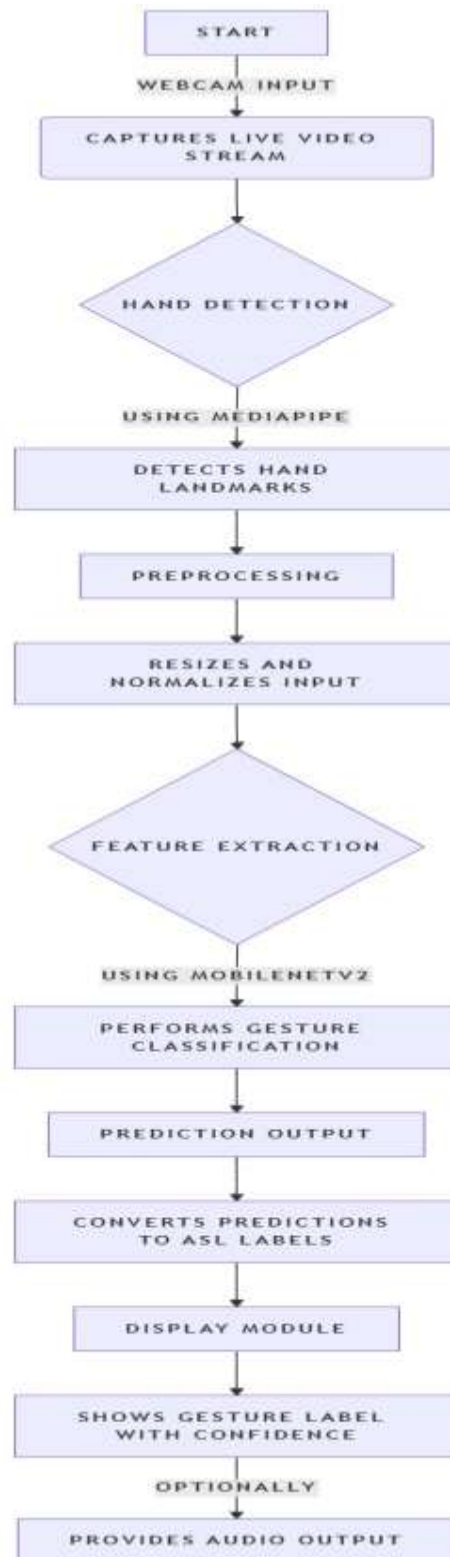


Figure 1: Flow Diagram of Silent Voice Amplified

Result and Conclusion:

By the end of the project, it is expected that a functional prototype will be developed that can recognize and classify 14 ASL gestures in real-time with high accuracy. The system will serve as a real-time assistive tool, capable of helping users with hearing impairments to communicate more effectively through technology. The project will also demonstrate the potential of lightweight deep learning models in accessible AI solutions.

Project Outcome & Industry Relevance:

This project aims to address a key societal issue bridging the communication gap for individuals with hearing disabilities. The proposed system will be highly relevant for use in:

- Assistive devices in healthcare and education.
- Learning platforms for teaching sign language.
- Real-time translation services in public service domains.

The project aligns with the growing interest in AI for accessibility and has the potential to be extended and adopted by industries focused on inclusive technology.

Working Model vs. Simulation/Study:

The project will result in a working real-time model capable of recognizing 14 ASL hand gestures via webcam input using deep learning techniques. It is not a theoretical simulation; rather, it is intended to produce a deployable real-time application for practical use.

Project Outcomes and Learnings:

Upon completion, the students will gain practical experience in:

- Deep learning model development and fine-tuning.
- Real-time video processing using OpenCV and MediaPipe.
- Handling challenges in gesture recognition such as background noise and lighting variability.
- Building assistive AI solutions that have real-world impact.

Future Scope:

- Expanding the system to recognize the full ASL alphabet, common words, and sentences.
- Adding Text-to-Speech (TTS) conversion for spoken output.
- Developing a mobile version using TensorFlow Lite.
- Integrating the system with video conferencing tools for live sign language interpretation.
- Supporting other sign languages for wider application.
- Using feedback loops to improve model accuracy through continual learning.