PROJECT REPORT

SignSpeak AI

Real-Time Sign Language Alphabet Gesture to Voice Translator Using Deep Neural Networks

Developer: Dr. Subramani

1. Abstract

Communication is a fundamental human right; however, for the deaf and mute community, effective interaction with non-signers remains a significant challenge. This project, **SignSpeak AI**, presents an Artificial Intelligence–based real-time **Alphabet Sign Language (A–Z) to Voice conversion system** designed to bridge this communication gap.

The system leverages **Computer Vision (MediaPipe)** to extract hand landmark keypoints from webcam input and **Deep Learning (TensorFlow/Keras)** to classify gestures. A **Feed-Forward Deep Neural Network**, commonly known as a **Multi-Layer Perceptron (MLP) classifier**, is employed as the core supervised learning model for **multi-class classification**. The MLP learns discriminative patterns from normalized hand joint coordinates and maps them to corresponding **American Sign Language (ASL) alphabet classes**.

The model architecture consists of multiple fully connected layers with nonlinear activation functions, enabling it to capture complex spatial relationships between hand landmarks while maintaining fast inference speed suitable for real-time applications. The system is trained on a large-scale **Kaggle ASL dataset containing 25,192 samples across 29 classes** (A–Z along with control tokens) and achieves a **validation accuracy of 99.54%** with low latency. These results demonstrate that **SignSpeak AI** is a reliable, efficient, and accessible solution for real-time daily communication.

2. Introduction

Sign Language is the primary mode of communication for millions of people worldwide. However, the majority of the general population does not understand it, creating barriers in education, healthcare, employment, and social interaction. While several systems focus on static image-based gesture recognition, real-time alphabet-level sign language translation remains a challenge.

This project focuses on **real-time recognition of ASL alphabet gestures [A–Z]**, forming the foundation for text and voice-based communication. The proposed solution is lightweight, browser-based, and optimized for real-time performance.

The system:

1. Captures video input from a standard webcam
2. Extracts **skeletal hand landmarks** instead of raw image pixels for speed and privacy
3. Classifies alphabet gestures using a trained neural network
4. Converts detected alphabets into speech output instantly

3. Objectives

- **Real-Time Processing:** Achieve smooth alphabet gesture recognition at approximately 30 FPS
- **Accessibility:** Provide a user-friendly interface requiring only a webcam
- **Accuracy:** Accurately detect ASL alphabet gestures under varying lighting and backgrounds
- **Dual-Hand Support:** Enable alphabet gesture recognition using either left or right hand
- **Foundational Translation:** Enable alphabet-level recognition as a base for future sentence formation

4. System Architecture

The system follows a modular pipeline architecture:

1. **Input Module:** Captures real-time video stream using OpenCV
2. **Feature Extraction:** MediaPipe Hands detects 21 three-dimensional hand landmarks
3. **Preprocessing:**
    - Conversion to relative coordinates (position invariance)
    - Normalization (scale and distance invariance)
    - **Mirror Logic:** X-axis flipping to support both left- and right-hand alphabet gestures
4. **Classification:** Feed-forward neural network predicts ASL alphabet classes
5. **Frontend Interface:** Flask-based web dashboard displays live video, confidence metrics, and recognized alphabets
6. **Output Module:** Recognized alphabets are converted into speech using the Web Speech API

5. Technology Stack

- **Programming Language:** Python 3.10+
- **Computer Vision:** OpenCV, MediaPipe
- **Deep Learning:** TensorFlow, Keras
- **Model Optimization:** TensorFlow Lite (TFLite)
- **Backend Framework:** Flask
- **Frontend:** HTML5, CSS3 (Glassmorphism & Cyber-Aesthetic UI), JavaScript
- **Data Handling:** NumPy, Pandas

6. Implementation Details

6.1 Dataset Description

The model was trained using the **Kaggle American Sign Language (ASL) Alphabet Dataset**, specifically processed for **alphabet-level gesture recognition**.

Feature	Description
Dataset Source	Kaggle – ASL Alphabet Dataset
Gesture Type	ASL Alphabets [A–Z]
Control Classes	Space, Del, Nothing
Total Classes	**29**
Samples per Class	~1,000
Total Samples	**25,192**
Input Features	42 (x, y coordinates of 21 hand landmarks)

The reduction from raw image pixels to skeletal key points significantly improves processing speed and robustness.

6.2 Model Architecture

A deep **Multi-Layer Perceptron (MLP)** neural network was implemented for alphabet gesture classification:

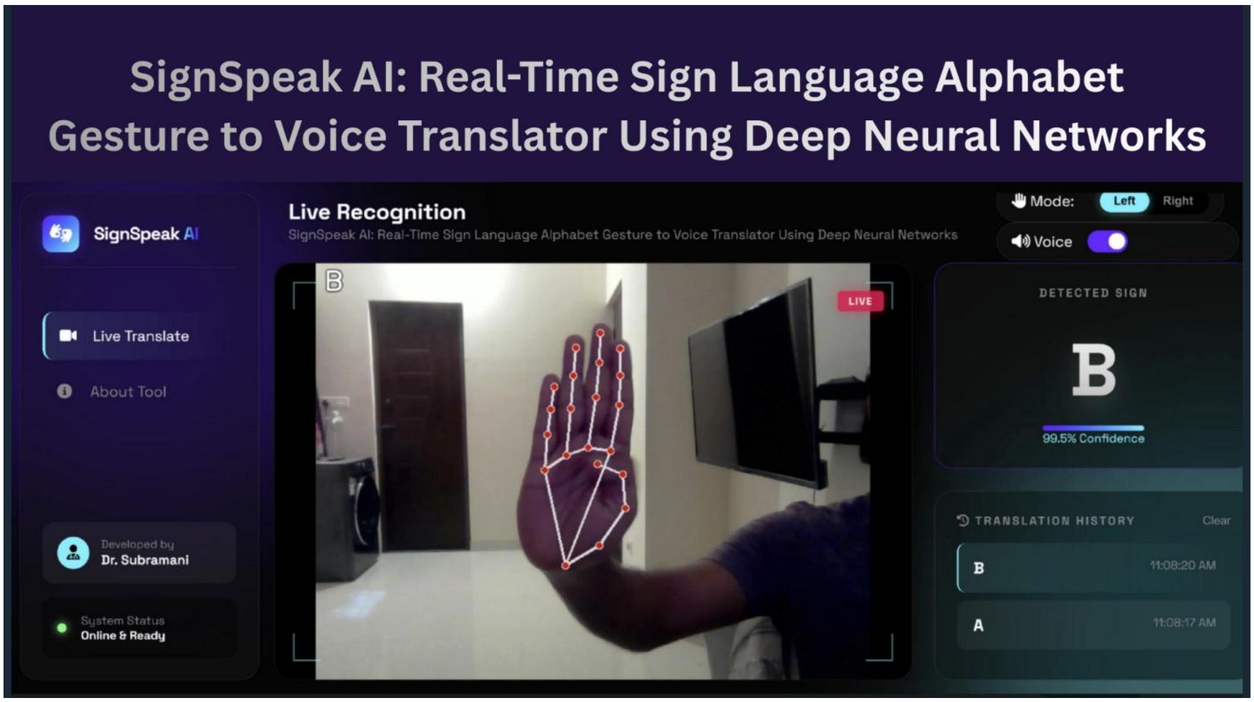
- **Input Layer:** 42 neurons (21 landmarks × 2 coordinates)
- **Hidden Layers:**
    - Dense Layer (128 neurons, ReLU)
    - Dropout Layer (20%)
    - Dense Layer (64 neurons, ReLU)
    - Dropout Layer (20%)
    - Dense Layer (32 neurons, ReLU)
- **Output Layer:** Dense Layer with **29 neurons** and **Softmax** activation
  (Represents A–Z alphabets + space, del, nothing)

Total Trainable Parameters: 16,797

6.3 Interface Design

The user interface reflects a modern **SignSpeak AI** identity:

- **Glassmorphism UI:** Semi-transparent cards with blur effects
- **Live Dashboard:** Real-time alphabet prediction with confidence scores
- **Hand Mode Toggle:** Enables left- or right-hand alphabet recognition via mirror logic



7. Results and Performance

7.1 Accuracy Metrics

The model was trained for **100 epochs** on the complete dataset.

Metric	Value
Training Accuracy	**99.40%**
Validation Accuracy	**99.54%**
Validation Loss	**0.0199**

7.2 Performance Evaluation

- **Latency:** < 50 ms per frame
- **Real-Time Capability:** Smooth alphabet recognition
- **Robustness:** Performs well under low lighting and dynamic backgrounds due to skeletal tracking

8. Conclusion

SignSpeak AI effectively demonstrates how AI and Computer Vision can be applied to build a highly accurate **real-time ASL alphabet recognition system**. By combining MediaPipe-based hand landmark extraction with a lightweight deep learning model, the system achieves near-human-level accuracy while remaining computationally efficient. This project provides a strong foundation for future expansion into full sign language sentence translation.

9. Future Scope

- Recognition of **dynamic alphabet gestures** (e.g., J and Z) using LSTM
- Alphabet-to-word and sentence construction
- Multilingual voice output
- Mobile application deployment (Android/iOS)
- Continuous sign language recognition