Real-Time Sign Language Recognition Using Deep Learning

Abstract

Communication is one of the most powerful human abilities, yet people with hearing or speech impairments often face barriers when interacting with others.
This project presents an **AI-based sign language recognition system** that detects and interprets hand gestures in real time and converts them into **text and speech**.
Using a **MobileNetV2 deep learning model** trained on the **American Sign Language (ASL) Alphabet dataset**, the system achieves high accuracy and performs robustly with live webcam input.
This work demonstrates how **artificial intelligence can enhance accessibility** and promote inclusive human–computer interaction.

Objectives

To recognize ASL alphabet signs (A–Z, space, delete, nothing) from real-time video.

To convert detected gestures into readable **text** and audible **speech**.

To implement **transfer learning** for accurate, fast gesture recognition.

To create a **human-centric AI tool** for assisting hearing-impaired communication.

Methodology

Dataset

Dataset Source: Kaggle – *ASL Alphabet Dataset (Grassknoted)*

Contains ~87,000 training images and ~21,750 validation images.

Each class (A–Z + 3 special symbols) has 3,000 samples.

Uniform lighting and centered gestures enable high-quality training.

Data Pre-Processing

All images resized to **128×128 pixels.**

Normalized pixel values (rescale = 1/255).

Data augmentation: rotation, shift, zoom, and horizontal flip.

Dataset split into **80% training / 20% validation** using Keras ImageDataGenerator.

Model Architecture

Base Network: MobileNetV2 (pretrained on ImageNet).

Fine-Tuning: Last 30 layers unfrozen to learn ASL-specific features.

Custom Head

GlobalAveragePooling2D

Dense(128, ReLU)

Dropout(0.3)

Dense(29, Softmax)

Optimizer: Adam (lr = 1×10^{-4})

Loss Function: Categorical Cross-Entropy

Training

10 epochs with batch size = 32.

Hardware: Google Colab GPU.

Validation accuracy reached > 98 – 100 %, indicating strong performance.

Real-Time Testing

Implemented webcam capture in Colab using JavaScript and OpenCV.

Each captured frame is resized and passed to the trained model for prediction

The system outputs the predicted letter with confidence score.

Text-to-Speech Integration

Added **pyttsx3** engine to convert predicted text into speech.

The model can pronounce individual letters or full words.

Word Formation Logic

Allows multiple sign captures to form a **complete word**.

Supports “space” and “delete” gestures.

When finished, the system **speaks the entire word aloud**.

Results

Metric	**Value**
Training Accuracy	99.9 – 100 %
Validation Accuracy	98 – 100 %
Dataset Size	108 k images
Classes	29
Inference Time	< 100 ms per frame

Live Output:

Predicted: P (99.5 %)   Speaking: P

Word Formation Demo:

Predicted sequence: H, E, L, L, O
Final Predicted Word: P
Speaking: P

Benefits

Converts **hand gestures → text → voice**, bridging communication barriers.

Demonstrates **real-world AI for accessibility** and inclusion.

Operates in **real time**, enabling natural interaction.

Lightweight architecture suitable for **mobile or embedded deployment**.

Future Scope

Extend to **dynamic gestures** (continuous word signing).

Train on **Indian Sign Language (ISL)** or multilingual datasets.

Deploy as a **Streamlit web app** or **Android application**.

Integrate with **speech-to-text** for two-way communication.

Combine with **sensor fusion (camera + IMU)** for complex gesture detection.

Conclusion

This project successfully implements a **real-time deep learning-based sign language recognition system** using MobileNetV2.
The model interprets ASL hand gestures with exceptional accuracy and translates them into text and speech, enhancing interaction between hearing-impaired and non-signing individuals.
It demonstrates the potential of **AI for social good**, proving that technology can be inclusive, human-centered, and empowering.

Key Technologies Used

Category	**Tools / Libraries**
Deep Learning	TensorFlow / Keras
Model	MobileNetV2 (Transfer Learning)
Data Processing	ImageDataGenerator
Visualization	Matplotlib
Real-Time Input	OpenCV + JavaScript (Colab)
Text-to-Speech	pyttsx3
Platform	Google Colab / Kaggle